

A dual-weighted trust-region adaptive POD 4-D Var applied to a finite-volume shallow water equations model on the sphere

X. Chen¹, S. Akella² and I. M. Navon^{3,*},[†]

¹*Department of Mathematics, Florida State University, Tallahassee, FL 32306, U.S.A.*

²*Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, U.S.A.*

³*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, U.S.A.*

SUMMARY

In this paper we study solutions of an inverse problem for a global shallow water model controlling its initial conditions specified from the 40-yr ECMWF Re-analysis (ERA-40) data sets, in the presence of full or incomplete observations being assimilated in a time interval (window of assimilation) with or without background error covariance terms. As an extension of the work by Chen *et al.* (*Int. J. Numer. Meth. Fluids* 2009), we attempt to obtain a reduced order model of the above inverse problem, based on proper orthogonal decomposition (POD), referred to as POD 4D-Var for a finite volume global shallow water equation model based on the Lin–Rood flux-form semi-Lagrangian semi-implicit time integration scheme. Different approaches of POD implementation for the reduced inverse problem are compared, including a dual-weighted method for snapshot selection coupled with a trust-region POD adaptivity approach. Numerical results with various observational densities and background error covariance operator are also presented. The POD 4-D Var model results combined with the trust-region adaptivity exhibit similarity in terms of various error metrics to the full 4D Var results, but are obtained using a significantly lesser number of minimization iterations and require lesser CPU time. Based on our previous and current work, we conclude that POD 4-D Var certainly warrants further studies, with promising potential of its extension to operational 3-D numerical weather prediction models. Copyright © 2011 John Wiley & Sons, Ltd.

Received 4 August 2010; Revised 23 November 2010; Accepted 28 November 2010

KEY WORDS: proper orthogonal decomposition; 4-D Var; finite volume; shallow water equations; trust-region method; inverse problem

1. INTRODUCTION

In this article, we address a proper orthogonal decomposition (POD) model reduction along with inverse solution of a two-dimensional global shallow water equations (SWE) model. Solutions of SWE [1–3] exhibit some of the important properties of large-scale atmospheric flow and the equations have certain important features (such as horizontal dynamical aspects) in common with more complicated Numerical Weather Prediction (NWP) models. Therefore, derivation and testing of various algorithms for solving SWE have often been a first step towards developing new atmosphere and ocean general circulation models. The explicit flux-form semi-Lagrangian finite volume (FV) scheme has been used to solve the SWE henceforth referred to as FV-SWE [4–8] in the forward model integration.

Our intention here is to generalize the efficient state-of-the-art POD implementation from our previous work on finite element SWE on the limited area [9, 10] (FE-SWE) to a global FV-SWE

*Correspondence to: I. M. Navon, Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, U.S.A.

[†]E-mail: inavon@fsu.edu

model with realistic initial conditions, i.e. combining efficient snapshot selection in the presence of data assimilation system by merging dual weighting of snapshots with trust-region POD techniques.

The POD identifies basis functions or modes that optimally capture the average energy content from numerical or experimental data. POD was introduced in the context of analysis of turbulent flow by Holmes and Lumley [11]. Sirovich [12] introduced the idea of snapshots. See also the book of Holmes [11]. In other disciplines, the same procedure goes by the names of Karhunen–Loeve decomposition (KLD) (see [13, 14]) or principal component analysis (PCA). The method originated in the work of Pearson [15] 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.

The trust-region method goes back to [16–18]. See also the work of Winfield [19] followed by the important work of Powell [20, 21]. Finally, the terminology of trust region and Cauchy point was put forward by Dennis [22] and systematized by More and Sorensen [23]. The trust-region proper orthogonal decomposition (TRPOD) was recently proposed in [24–28] as a way to overcome difficulties related to POD reduced order modeling (ROM) used for solving partial differential equation (PDE)-constrained optimization problem. Combining the POD technique with the concept of trust-region (see Toint and Conn [29–31] for a comprehensive survey or Nocedal and Wright [32] 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, we have a global convergence result for TRPOD [24] proving that the iterates produced by the optimization algorithm, started at an arbitrary initial iterate, will converge to a local optimizer for the original model.

In the previous work [10], we studied the effect of combining TRPOD in conjunction with dual weighting Data Assimilation System (DAS) snapshot selection in the framework of Galerkin projection-based POD-ROM for FE-SWE on the limited area without a background error covariance term, in which the observations were available at all the time steps and distributed at all the grid points during the entire window of assimilation. One of the goals of this article is to show that dual-weighted TRPOD 4-D Var can also be applied to the global FV-SWE model with real initial conditions. Even in a case where a Galerkin projection scheme is unavailable from full space to the POD reduced space. We have considered cases of (a) complete observations distributed in space and time with an unbalanced background error covariance term being provided, and (b) incomplete observations in both time and space, with or without a balanced background error covariance term being included in the cost functional. In the framework of TRPOD 4-D Var with background error covariance term inclusion, an ideal preconditioning of the POD 4-D Var is derived so that the Hessian matrix of the POD reduced order background error covariance matrix becomes the identity matrix. In this paper we show that TRPOD 4-D Var performs satisfactorily in the presence of incomplete observations, just as in the case of full 4-D Var, if a geostrophically balanced background error covariance matrix is available during the implementation of data assimilation.

In the numerical experiments, we compared (I) the *ad hoc* update adaptivity of the POD 4-D VAR, (II) the trust-region update adaptivity with or without dual weighting and (III) full 4-D Var (high-fidelity model). We confirmed that the combination of TRPOD and dual-weighted snapshots yields the best results in all error metrics (see [10, 33–35]). The advantage of TRPOD adaptivity over *ad hoc* POD adaptivity is due to the fact that TRPOD can appropriately determine a trust region within which the step size is not too small, so that it is guaranteed to compute a sufficient decrease for the cost functional of the full model by projecting the Quasi–Newton direction of POD reduced order cost functional into the trust-region box as a substitute for the Cauchy point in the standard trust-region methods using quadratic approximation. Hence, TRPOD allows for successively refreshing the POD basis by following iterates produced by the optimization algorithm to converge to a local optimizer of the high-fidelity model [30, 31].

Also, we notice that there are almost twice as many outer projections (refreshing the snapshots) related to TRPOD adaptivity compared to the number of projections in *ad hoc* adaptivity in the framework of Galerkin projection in our previous work [10]. In this work there are almost thrice as many outer projections related to TRPOD adaptivity as compared to *ad hoc* adaptivity without

a Galerkin projection scheme in the presence of incomplete observations in time and space. The CPU time required by TRPOD 4-D Var is still a fraction of the CPU time required by the full 4-D Var, due to the fact that most of the functional evaluations are carried out in the lower dimensional POD 4-D Var while the full model will be evaluated only when an appropriate descent direction in the TRPOD reduced order space is obtained. Therefore, TRPOD avoids unnecessary full model evaluations and also reduces the cost of minimization inside the inner TRPOD loop.

The plan of the paper is as follows: in Section 2 the FVSWE model description is followed by a brief presentation of the POD model reduction method. For the sake of clarity and self-containedness, in Section 3 we provide the framework of POD for reduced order 4-D Var data assimilation. This section is comprised of subsections detailing dual weighting of snapshots and implementation of the reduced order 4-D VAR, specifically for the FV-SWE model. Section 4 addresses the TR-POD methodology. Section 5 details the numerical experiments carried out in order to validate the accuracy of the POD reduced order model and the POD 4-D VAR approach for the various numerical methods enumerated above. For the recent work on POD 4-D VAR, see [36–51]. In particular, we compare *ad hoc* adaptivity for POD 4-D VAR with trust-region adaptivity in combination with dual-weighted snapshots when full observations are available in our experiment. We also compared TRPOD 4-D Var with full 4-D Var for incomplete observations in space, addressing this issue in the discussion of numerical results. Finally, the paper concludes with a summary and conclusions.

2. GLOBAL SHALLOW WATER EQUATION MODELS

In spherical coordinates the vorticity divergence form of the SWE can be written as the mass conservation law for a shallow layer of water

$$\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{V}h) = 0 \quad (1)$$

and the vector-invariant form of momentum equations

$$\frac{\partial u}{\partial t} = \Omega v - \frac{1}{a \cos \theta} \frac{\partial(\kappa + \Phi)}{\partial \lambda} \quad (2)$$

$$\frac{\partial v}{\partial t} = -\Omega u - \frac{1}{a} \frac{\partial(\kappa + \Phi)}{\partial \theta} \quad (3)$$

$$(\lambda, \theta) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times [-\pi, \pi], \quad t \geq 0$$

where h represents the fluid height (above the surface height h_s), $\mathbf{V} = (u, v)$, u and v represents the zonal and meridional wind velocity components, respectively, θ and λ are the latitudinal and longitudinal directions, respectively, ω is the angular speed of rotation of the earth, a is the radius of earth. The free surface geopotential is given by $\Phi = \phi_s + gh$, where $\phi_s = gh_s$, $\kappa = \frac{1}{2} \mathbf{V} \cdot \mathbf{V}$ is the kinetic energy and $\Omega = 2\omega \sin \theta + \nabla \times \mathbf{V}$ is the absolute vorticity.

In this paper we have used a discretized (FV, semi-Lagrangian) version of the above SWE model, which serves as the dynamical core in the community atmosphere model (CAM), version 3.0, and its operational version implemented at NCAR and NASA is known as FV-general circulation model (FV-GCM). In brief, a two-grid combination based on C-grid and D-grids is used for advancing from time step t_n to $t_n + \Delta t$. In the first half of the time step, advective winds (time centered winds on the C-grid: (u^*, v^*)) are updated on the C-grid, and in the other half of the time step, the prognostic variables (h, u, v) are updated on the D-grid.

Using the FV method, within each cell of the discrete grid, if we consider a piecewise linear approximation to the solution, whose slope is *limited* in a certain way depending on the values of the solution at the neighboring grid cells, one can consistently derive a family of van Leer schemes. We will follow the suggestion in [7] and always use the unconstrained van Leer [52–54] scheme

to advect winds on the C-grid. The same advection scheme will be used on D-grid as well. This strategy provides solutions whose accuracy is comparable to those obtained by using more CPU demanding advection schemes, for e.g. constrained van Leer schemes.

3. DUAL-WEIGHTED POD METHOD

3.1. Method of snapshots

An ensemble of snapshots is chosen in the analysis time interval $[0, T]$ written as $\{y^1, y^2, \dots, y^n\}$, where $y^i = (h^i, u^i, v^i)^T \in \mathbb{R}^N$, $i = 1, \dots, n$, n is the number of snapshots and $N = 3N_x N_y$ is triple the dimension of discrete mesh, N_x and N_y are the mesh points of the latitudinal and longitudinal directions, respectively. Our choice of snapshots number was to take a snapshot at each time step ($\Delta t = 450$ s) of the window of assimilation whose length was taken in our case to be 15 h. We could have chosen another snapshot distribution; however, we selected to implement this choice as the most intuitive one (15 h = 120 time steps of 450 s, each). Define the dual-weighted ensemble average of the snapshots as $\bar{y} = \sum_{i=1}^n w_i y^i$, 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 $w_i = \frac{1}{n}$. Subtracting the mean from each snapshot, we obtain the following $N \times n$ -dimensional matrix $\mathbf{Y} = [y^1 - \bar{y}, y^2 - \bar{y}, \dots, y^n - \bar{y}]$.

The POD modes $\Psi = \{\psi^1, \psi^2, \dots, \psi^M\}$ of order $M \leq n$ provide an optimal representation of the ensemble data in an M -dimensional state subspace by minimizing the averaged projection error

$$\begin{aligned} \min_{\{\psi^1, \psi^2, \dots, \psi^M\}} & \sum_{i=1}^n w_i \|(y^i - \bar{y}) - \Pi_{\Psi, M}(y^i - \bar{y})\|^2 \\ \text{s.t.} & \langle \psi^i, \psi^j \rangle_{l_2} = \delta_{ij} \end{aligned} \quad (4)$$

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

$$\Pi_{\Psi, M} = \sum_{i=1}^M \langle y, \psi_i \rangle_{l_2} \psi_i$$

We define the dual-weighted spatial correlation matrix, $\mathbf{A} = \mathbf{Y}\mathbf{W}\mathbf{Y}^T$, where $\mathbf{W} = \text{diag}\{w_1, w_2, \dots, w_n\}$ is the diagonal matrix of weights.

To compute the dual-weighted POD modes $\psi^i \in \mathbb{R}^N$, one must solve an N -dimensional eigenvalue problem, $\mathbf{A}\psi_i = \lambda_i \psi_i$.

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

$$\mathbf{K}^{n \times n} = \mathbf{W}^{1/2} \mathbf{Y}^T \mathbf{Y} \mathbf{W}^{1/2} \quad (5)$$

and compute the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ of $\mathbf{K}^{n \times n}$ with its corresponding eigenvectors χ_1, \dots, χ_n

Hence, the corresponding POD modes are thus obtained by defining

$$\psi_i = \frac{1}{\sqrt{\lambda_i}} \mathbf{Y} \mathbf{W}^{1/2} \chi_i, \quad i = 1, \dots, M \quad (6)$$

where

$$\langle \psi^i, \psi^j \rangle_{l_2} = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (7)$$

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

$$I(m) = \frac{\sum_{i=1}^{i=m} \lambda_i}{\sum_{i=1}^{i=n} \lambda_i} \tag{8}$$

and choose M such that $M = \arg \min\{I(m) : I(m) > \gamma\}$, where $0 \leq \gamma \leq 1$ is the percentage of total information retained in the reduced space and the tolerance γ must be chosen to be close to unity in order to capture most of the energy of the snapshots basis.

3.2. POD reduced order model for the FV-SWE model

Define the following vectors:

$$\begin{aligned} \mathbf{h}^k &= (h_1^k \ h_2^k \ \dots \ h_N^k)^T, & \mathbf{u}^k &= (u_1^k \ u_2^k \ \dots \ u_N^k)^T, & \mathbf{v}^k &= (v_1^k \ v_2^k \ \dots \ v_N^k)^T \\ \mathbf{h}^* &= (h_1^* \ h_2^* \ \dots \ h_N^*)^T, & \mathbf{u}^* &= (u_1^* \ u_2^* \ \dots \ u_N^*)^T, & \mathbf{v}^* &= (v_1^* \ v_2^* \ \dots \ v_N^*)^T \end{aligned}$$

thus \mathbf{h}^* , \mathbf{u}^* and \mathbf{v}^* are obtained on the C-grid [6, 7], in the following way:

$$\begin{aligned} \mathbf{h}^* &= \mathbf{h}^k + \mathbf{F}_h^c(\mathbf{h}^k, \mathbf{u}^k, \mathbf{v}^k) \\ \mathbf{u}^* &= \mathbf{u}^k + \frac{\Delta t}{2} \mathbf{F}_u^c(\mathbf{h}^k, \mathbf{u}^k, \mathbf{v}^k) \\ \mathbf{v}^* &= \mathbf{v}^k + \frac{\Delta t}{2} \mathbf{F}_v^c(\mathbf{h}^k, \mathbf{u}^k, \mathbf{v}^k) \end{aligned} \tag{9}$$

Define the following vectors:

$$\Psi = (\Psi_h \ \Psi_u \ \Psi_v)^T, \quad \bar{y} = (\bar{\mathbf{h}} \ \bar{\mathbf{u}} \ \bar{\mathbf{v}})^T \tag{10}$$

and we obtain the POD reduced order model on the C-grid by projection as follows, where the α coefficients are the modal coefficients of the flow field with respect to the POD basis;

$$\begin{aligned} \alpha_h^k &= \alpha_h^k + \Psi_h^T (\mathbf{F}_h^c(\Psi_h \alpha_h^k + \bar{\mathbf{h}}, \Psi_u \alpha_u^k + \bar{\mathbf{u}}, \Psi_v \alpha_v^k + \bar{\mathbf{v}}) - \bar{\mathbf{h}}) \\ \alpha_u^k &= \alpha_u^k + \frac{\Delta t}{2} \Psi_u^T (\mathbf{F}_u^c(\Psi_h \alpha_h^k + \bar{\mathbf{h}}, \Psi_u \alpha_u^k + \bar{\mathbf{u}}, \Psi_v \alpha_v^k + \bar{\mathbf{v}}) - \bar{\mathbf{u}}) \\ \alpha_v^k &= \alpha_v^k + \frac{\Delta t}{2} \Psi_v^T (\mathbf{F}_v^c(\Psi_h \alpha_h^k + \bar{\mathbf{h}}, \Psi_u \alpha_u^k + \bar{\mathbf{u}}, \Psi_v \alpha_v^k + \bar{\mathbf{v}}) - \bar{\mathbf{v}}). \end{aligned} \tag{11}$$

Similarly, we can rewrite the D-grid [6, 7] time integration as the following vector formulation:

$$\begin{aligned} \mathbf{h}^{k+1} &= \mathbf{h}^k + \mathbf{F}_h^d(\mathbf{h}^*, \mathbf{u}^*, \mathbf{v}^*) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \frac{\Delta t}{2} \mathbf{F}_u^d(\mathbf{h}^*, \mathbf{u}^*, \mathbf{v}^*) \\ \mathbf{v}^{k+1} &= \mathbf{v}^k + \frac{\Delta t}{2} \mathbf{F}_v^d(\mathbf{h}^*, \mathbf{u}^*, \mathbf{v}^*) \end{aligned} \tag{12}$$

and the POD reduced order model on the D-grid by projection as follows:

$$\begin{aligned} \alpha_h^{k+1} &= \alpha_h^k + \Psi_h^T (\mathbf{F}_h^d(\Psi_h \alpha_h^* + \bar{\mathbf{h}}, \Psi_u \alpha_u^* + \bar{\mathbf{u}}, \Psi_v \alpha_v^* + \bar{\mathbf{v}}) - \bar{\mathbf{h}}) \\ \alpha_u^{k+1} &= \alpha_u^k + \frac{\Delta t}{2} \Psi_u^T (\mathbf{F}_u^d(\Psi_h \alpha_h^* + \bar{\mathbf{h}}, \Psi_u \alpha_u^* + \bar{\mathbf{u}}, \Psi_v \alpha_v^* + \bar{\mathbf{v}}) - \bar{\mathbf{u}}) \\ \alpha_v^{k+1} &= \alpha_v^k + \frac{\Delta t}{2} \Psi_v^T (\mathbf{F}_v^d(\Psi_h \alpha_h^* + \bar{\mathbf{h}}, \Psi_u \alpha_u^* + \bar{\mathbf{u}}, \Psi_v \alpha_v^* + \bar{\mathbf{v}}) - \bar{\mathbf{v}}) \end{aligned} \tag{13}$$

where $\alpha_h^k \in R^{M_h}$, $\alpha_u^k \in R^{M_u}$ and $\alpha_v^k \in R^{M_v}$, $k=0, 1, 2, \dots, n$ and initial values are

$$\alpha_h^0 = \Psi_h^T(\mathbf{h}^0 - \bar{\mathbf{h}}), \quad \alpha_u^0 = \Psi_u^T(\mathbf{u}^0 - \bar{\mathbf{u}}), \quad \alpha_v^0 = \Psi_v^T(\mathbf{v}^0 - \bar{\mathbf{v}}) \tag{14}$$

Formulas (11) and (13) are the POD reduced order model for the FV-SWE models (1) (2) and (3), and it only includes $(M_h + M_u + M_v) \times n$ degrees of freedom, where $M_h, M_u, M_v \ll N$ compared to the numerical FV-SWE model that contains $3N \times n$ degrees of freedom.

3.3. The generation of dual-weighted POD basis

One of the goals of 4-D Var data assimilation is to obtain an ‘optimal’ representation of the state of the atmosphere by fusing model predictions with observational data.

This is achieved by minimizing

$$J(y_0) = \frac{1}{2}(y_0 - y^b)^T \mathbf{B}^{-1}(y_0 - y^b) + \frac{1}{2} \sum_{k=0}^{k=n} (\mathbf{H}_k y_k - y_k^o)^T \mathbf{R}_k^{-1} (\mathbf{H}_k y_k - y_k^o) \tag{15}$$

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 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. By implementing a dual-weighted POD (DWPOD) method [10, 35], we can incorporate the information from the 4-D Var into the POD reduced order model. 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.

Assume that 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, $\forall \tau < t, y(t) = M_{\tau \rightarrow t}(y(\tau)) = M_{\tau, t}(y(\tau))$ 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 $t_i \leq t$ on J may be estimated using the tangent linear model $\mathbf{M}(t_i, t)$ and its adjoint model $\mathbf{M}^T(t, t_i)$, where the brackets stand for the l_2 product.

$$\begin{aligned} \delta J &\approx \langle \nabla J_{y(t)}(y(t)), \delta y(t) \rangle_{l_2} = \langle \nabla J_{y(t)}(y(t)), \mathbf{M}(t_i, t) \delta y(t_i) \rangle_{l_2} \\ &= \langle \mathbf{M}^T(t, t_i) \nabla J_{y(t)}(y(t)), \delta y(t_i) \rangle_{l_2} = \langle y_{t_i}^{**}, \delta y(t_i) \rangle_{l_2} \end{aligned} \tag{16}$$

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

In particular, the discrete model can be written as $\forall k, y_k = M_{k-1 \rightarrow k}(y_{k-1}) = M_k(y_{k-1})$, where $M_{k-1 \rightarrow k}$ is defined as the model forecast operator from time $k-1$ to k .

In order to derive the algorithm for the computation of dual weights 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^T :

$$\delta J_k \approx \langle \nabla J_k, \delta y_k \rangle_{l_2} = \langle \nabla J_k, \mathbf{M}_k \delta y_{k-1} \rangle_{l_2} = \langle \mathbf{M}_k^T \nabla J_k, \delta y_{k-1} \rangle_{l_2} = \langle y_{k-1}^{**}, \delta y_{k-1} \rangle_{l_2} \tag{17}$$

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

Therefore, $|\delta J_k| = |\langle y_{k-1}^{**}, \delta y_{k-1} \rangle_{l_2}| \leq \|y_{k-1}^{**}\|_{l_2} \|\delta y_{k-1}\|_{l_2}$.

Finally, the dual weights w_k associated with the snapshots selection are thus defined as normalized values, $c_k = \|y_{k-1}^{**}\|_{l_2} w_k = c_k / \sum_{j=1}^{j=n} c_j, k=1, \dots, n$, 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 practice, the evaluation of all dual weights requires only one adjoint model integration [10]; summarized below are the steps involved in the computation of dual weights.

- (1) Initialize the adjoint variables y^{**} at the final time to zero, $y_n^{**} = 0$.

- (2) At each time step, t_{k-1} , the adjoint variables y_{k-1}^{**} are obtained as, $y_{k-1}^{**} = \mathbf{M}_k^T y_k^{**} + \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{H}_k y_k - y_k^o)$, where \mathbf{M}_k is the tangent linear model and \mathbf{H} is the linearized observation operator at t_k .
- (3) For initial time, t_0 , we obtain, $y_0^{**} = y_0^{**} + \mathbf{B}^{-1}(y_0 - y^b)$.
- (4) The weights are obtained as

$$c_k = \|(\mathbf{A})^{-1} y_k^{**}\|_{\mathbf{A}} \quad \text{and} \quad w_k = \frac{c_k}{\sum_{j=1}^n c_j}, \quad k = 1, \dots, n.$$

Since the adjoint model is available during the implementation of the full 4-D Var, no additional cost is required for the development of DWPOD 4-D Var over the classic POD 4-D Var.

4. TRUST-REGION METHOD APPLIED TO POD

In the classical trust-region method [55] our goal is to define a region around the current iterate within which we trust the model to be an adequate representation of the objective function, f , and then choose a (direction and size of) step to be the approximate minimizer of the model in the trust region. 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

$$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \mathbf{B}_k p \tag{18}$$

where $f_k = f(x_k)$ and $\nabla f_k = \nabla f(x_k)$ and \mathbf{B}_k is an approximation to the Hessian. Therefore, the function and gradient values from the above model are same as the exact function and gradient values, respectively.

In order to obtain each step, we seek a solution of the following subproblem for which we only need an approximate solution to obtain convergence and good practical behavior [32]

$$\min \quad m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \mathbf{B}_k p \tag{19}$$

$$\text{s.t.} \quad \|p\| \leq \delta_k \tag{20}$$

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

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)} \tag{21}$$

where the numerator is called the actual reduction, and the denominator is called the predicted reduction. We measure agreement between 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 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.

Here, the POD-ROM is based on the solution of the original model for specified control variables: the model initial conditions. It is therefore necessary to reconstruct the POD-ROM when the resulting control variables from the latest optimization iteration are significantly different from the ones upon which the POD model is based. Hence, it is natural to improve the POD reduced order control model successively by updating the snapshots that are used to generate the POD basis in the process of reduced order 4-D Var.

The 4-D Var cost functional in (15),

$$J(y_0) = \underbrace{\frac{1}{2}(y_0 - y^b)^T \mathbf{B}^{-1}(y_0 - y^b)}_{J_b} + \underbrace{\frac{1}{2} \sum_{k=0}^{k=n} (\mathbf{H}_k y_k - y_k^o)^T \mathbf{R}_k^{-1} (\mathbf{H}_k y_k - y_k^o)}_{J_o}$$

can be separated into $J = J_b + J_o$, where $J_b = \frac{1}{2}(y_0 - y^b)^T \mathbf{B}^{-1}(y_0 - y^b)$ is the background cost functional and

$$J_o = \frac{1}{2} \sum_{k=0}^{k=n} (\mathbf{H}_k y_k - y_k^o)^T \mathbf{R}_k^{-1} (\mathbf{H}_k y_k - y_k^o) \tag{22}$$

is the observational cost functional. Let $\delta y = y_0 - y^b$, so that the background cost functional can be rewritten as, $J_b = \frac{1}{2}(\delta y)^T \mathbf{B}^{-1}(\delta y)$.

Define an approximation to the control variable, $y_0 \approx \Psi \alpha_0 + \bar{y}$, where the POD modes are given by $\Psi = \{\psi^1, \psi^2, \dots, \psi^M\}$ and the dual-weighted ensemble average of the snapshots is given as before, in which α_0 is the corresponding control variable in the M -dimensional POD reduced order space. Define the coefficient, $\alpha_b = \Psi^T(y^b - \bar{y})$, and we obtain the background term y^b in terms of POD modes, $y^b \approx \Psi \alpha_b + \bar{y}$.

From the above equations we obtain, $\delta y = y_0 - y^b \approx (\Psi \alpha_0 + \bar{y}) - (\Psi \alpha_b + \bar{y}) = \Psi(\alpha_0 - \alpha_b)$.

Let $\delta \alpha = \alpha_0 - \alpha_b$, so that $\delta y = \Psi \delta \alpha$.

Hence, the 4-D Var cost functional in (15) can be approximated by

$$J(y_0) \approx \hat{J}(\delta \alpha) = \hat{J}_b(\delta \alpha) + \hat{J}_o(\delta \alpha) \tag{23}$$

where

$$\hat{J}_b(\delta \alpha) = \frac{1}{2}(\delta \alpha)^T (\Psi^T \mathbf{B}^{-1} \Psi)(\delta \alpha) \tag{24}$$

$$\hat{J}_o(\delta \alpha) = \frac{1}{2} \sum_{k=0}^{k=n} (\mathbf{H}_k \mathbf{M}_k (y^b + \Psi \delta \alpha) - y_k^o)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{M}_k (y^b + \Psi \delta \alpha) - y_k^o) \tag{25}$$

Since the inverse of the background error covariance matrix \mathbf{B}^{-1} is a symmetric positive-definite matrix (SPD), it is easy to verify that $\Psi^T \mathbf{B}^{-1} \Psi$ is SPD from the fact that $\Psi^T \Psi = \mathbf{I}$.

Define

$$\hat{\mathbf{B}}^{-1} = \Psi^T \mathbf{B}^{-1} \Psi. \tag{26}$$

Therefore $\hat{\mathbf{B}}^{-1}$ is SPD and (24) can be written as,

$$\hat{J}_b(\delta \alpha) = \frac{1}{2}(\delta \alpha)^T \hat{\mathbf{B}}^{-1}(\delta \alpha) \tag{27}$$

Since $\hat{\mathbf{B}}^{-1}$ is SPD, we can find the square-root matrix

$$\hat{\mathbf{B}} = \hat{\mathbf{B}}^{\frac{1}{2}} \hat{\mathbf{B}}^{\frac{T}{2}} \tag{28}$$

using the inverse Cholesky decomposition methodology without finding $\hat{\mathbf{B}}$ itself. Define a transformation $\delta \alpha = \hat{\mathbf{B}}^{1/2} v^\alpha$.

Hence, we obtain that

$$\begin{aligned} \tilde{J}_b(v^\alpha) &= \hat{J}_b(\delta \alpha) = \hat{J}_b(\hat{\mathbf{B}}^{\frac{1}{2}} v^\alpha) = \frac{1}{2}(\delta \alpha)^T \hat{\mathbf{B}}^{-1}(\delta \alpha) = \frac{1}{2}(\hat{\mathbf{B}}^{\frac{1}{2}} v^\alpha)^T \hat{\mathbf{B}}^{-1}(\hat{\mathbf{B}}^{\frac{1}{2}} v^\alpha) \\ &= \frac{1}{2}(\hat{\mathbf{B}}^{\frac{1}{2}} v^\alpha)^T (\hat{\mathbf{B}}^{\frac{1}{2}} \hat{\mathbf{B}}^{\frac{T}{2}})^{-1} (\hat{\mathbf{B}}^{\frac{1}{2}} v^\alpha) = \frac{1}{2}(v^\alpha)^T \hat{\mathbf{B}}^{\frac{T}{2}} \hat{\mathbf{B}}^{-\frac{T}{2}} \hat{\mathbf{B}}^{\frac{1}{2}} \hat{\mathbf{B}}^{\frac{1}{2}} v^\alpha \\ &= \frac{1}{2}(v^\alpha)^T v^\alpha. \end{aligned} \tag{29}$$

The methodology of construction of $\mathbf{B}^{1/2}$ and $\mathbf{B}^{T/2}$ using univariate correlation and multivariate geostrophic balancing operators is detailed in an Appendix A (see also [56]). Therefore, the gradient of the background cost functional, $\tilde{J}_b(v^\alpha)$ with respect to v^α is given by

$$\nabla_{v^\alpha} \tilde{J}_b = v^\alpha \tag{30}$$

and the Hessian of the background cost functional, $\tilde{J}_b(v^\alpha)$, with respect to v^α is given by,

$$\nabla_{v^\alpha}^2 \tilde{J}_b = \mathbf{I}_M. \tag{31}$$

To summarize, we obtain that the cost functional can be approximated by,

$$J(y_0) \approx \tilde{J}(v^\alpha) = \tilde{J}_b(v^\alpha) + \tilde{J}_o(v^\alpha) = \frac{1}{2}(v^\alpha)^T v^\alpha + \tilde{J}_o(v^\alpha) \tag{32}$$

and the gradient of the cost functional with respect to v^α is given by the chain rule,

$$\begin{aligned} \nabla_{v^\alpha} \tilde{J} &= v^\alpha + (\nabla_{v^\alpha} \alpha^0)^T ((\nabla_{\alpha^0} y^0)^T \nabla_{y^0} J_o) \\ &= v^\alpha + \hat{\mathbf{B}}^{\frac{T}{2}} \Psi^T \nabla_{y^0} J_o. \end{aligned} \tag{33}$$

4.1. Trust-region POD optimal control

From (24) and (25), we obtain the POD reduced order cost functional [28, 57] in terms of $\delta\alpha$,

$$\begin{aligned} \hat{J}(\delta\alpha) &= \frac{1}{2}(\delta\alpha)^T (\Psi^T \mathbf{B}^{-1} \Psi)(\delta\alpha) \\ &\quad + \frac{1}{2} \sum_{k=0}^{k=n} (\mathbf{H}_k \mathbf{M}_k(y^b + \Psi \delta\alpha) - y_k^o)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{M}_k(y^b + \Psi \delta\alpha) - y_k^o). \end{aligned} \tag{34}$$

Or from (32), we obtain POD reduced order cost functional in terms of v^α ,

$$\begin{aligned} \tilde{J}(v^\alpha) &= \frac{1}{2}(v^\alpha)^T v^\alpha \\ &\quad + \frac{1}{2} \sum_{k=0}^{k=n} (\mathbf{H}_k \mathbf{M}_k(y^b + \Psi(\hat{\mathbf{B}}^{1/2} v^\alpha)) - y_k^o)^T \mathbf{R}_k^{-1} (\mathbf{H}_k \mathbf{M}_k(y^b + \Psi(\hat{\mathbf{B}}^{1/2} v^\alpha)) - y_k^o) \end{aligned}$$

From an implementation point of view, we first start with a random perturbation of given initial condition $y_0^{(0)} = y^b$ and compute the snapshots $\mathbf{Y}^{(0)}$ and dual-weighted ensemble average $\bar{y}^{(0)}$ that correspond to the flow behavior forced by $y_0^{(0)}$. We then use these snapshots and dual-weighted ensemble average to compute the first POD basis $\Psi^{(0)}$ and build up the corresponding POD-based background term forced by applying inner projection $\alpha_b^{(0)} = (\Psi^{(0)})^T (y^b - \bar{y})$. Also, from (26) and (28), we can find $(\hat{\mathbf{B}}^{(0)})^{1/2}$.

Since initially we have $y_0^{(0)} = y^b$, we obtain the initial guess of the initial condition for the POD reduced order model

$$\alpha_0^{(0)} = (\Psi^{(0)})^T (y_0^{(0)} - \bar{y}^{(0)}) = \alpha_b^{(0)} \tag{35}$$

and $\delta\alpha^{(0)} = \alpha_0^{(0)} - \alpha_b^{(0)} = 0$, trivially; also, we obtain $v^{\alpha(0)} = (\hat{\mathbf{B}}^{(0)})^{-\frac{1}{2}} \alpha^{(0)} = 0$. Therefore, at the beginning of minimization of the POD reduced order cost functional $\tilde{J}(v^\alpha)$, we initialize POD reduced order control variable $v^\alpha = 0$.

We now implement the inner minimization iteration based on $\Psi^{(0)}$ to obtain the new control variable $v^{\alpha(1)}$ in the reduced order space. Thus, we can compute $\delta\alpha^{(1)} = (\hat{\mathbf{B}}^{(0)})^{1/2} v^{\alpha(1)}$ and update the initial condition using $\alpha_0^{(1)} = \alpha_b^{(0)} + \delta\alpha^{(1)}$ for the POD reduced order model.

Finally, when we carry out an outer iteration, we obtain $y_0^{(1)} = \bar{y}^{(0)} + \Psi^{(0)}\alpha_0^{(1)}$. If we use $y_0^{(1)}$ for the computation of new snapshots and a new POD basis $\Psi^{(1)}$, we can improve the initial condition of the PDE and thus improve the POD-based model.

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

$$\min \quad m_k(v^{\alpha(k)} + s) := \tilde{J}(v^{\alpha(k)} + s) \tag{36}$$

$$\text{s.t.} \quad \|s\| \leq \delta_k \tag{37}$$

Recall $m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T \mathbf{B}_k p$, where $f_k = f(x_k)$ and $\nabla f_k = \nabla f(x_k)$ and \mathbf{B}_k is an approximation to the Hessian, from (18).

Furthermore, from (33), we have the gradient of subproblem above,

$$\nabla_{v^{\alpha(k)}} m_k = v^{\alpha(k)} + (\hat{\mathbf{B}}^{(k)})^{\frac{1}{2}} (\Psi^{(k)})^T \nabla_{y_0} J_o$$

Using $y_0^{(k)} = \bar{y}^{(k)} + \Psi^{(k)}(\alpha_b^{(k)} + (\hat{\mathbf{B}}^{(k)})^{1/2} v^{\alpha(k)})$, we can define

$$y_0^* = \bar{y}^{(k)} + \Psi^{(k)}(\alpha_b^{(k)} + (\hat{\mathbf{B}}^{(k)})^{1/2} (v^{\alpha(k)} + s_k)) \tag{38}$$

Based on the trust-region strategy from optimization [25, 57], we can decide to increase or decrease the trust-region radius by comparing the actual (for the full-order model), $J(y_0^{(k)}) - J(y_0^*)$, with the predicted decrease (for the reduced order model), $m_k(v^{\alpha(k)}) - m_k(v^{\alpha(k)} + s_k)$. Algorithmic details of the steps involved are summarized below. Let $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1 \leq \gamma_3$ and $y_0^{(0)} = y^b$ and δ_0 be given, set $k = 0$.

- (1) Compute snapshot set $\mathcal{Y}_k^{\text{SNAP}}$ based on initial condition $y_0^{(k)}$.
- (2) Compute the POD basis $\Psi^{(k)}$ and dual-weighted ensemble average $\bar{y}^{(k)}$.
- (3) Compute POD-based background term $\alpha_b^{(k)} = (\Psi^{(k)})^T (y_0^{(k)} - \bar{y}^{(k)})$ and build the POD-based background covariance matrix $(\hat{\mathbf{B}}^{(k)})^{\frac{1}{2}}$.
- (4) Initialize $v^{\alpha(k)} = 0$ and compute the minimizer s^k of

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

- (5) Compute the new $J(y_0^*)$ of the full model so that we obtain

$$\rho_k = \frac{J(y_0^{(k)}) - J(y_0^*)}{m_k(v^{\alpha(k)}) - m_k(v^{\alpha(k)} + s_k)} \tag{39}$$

where $y_0^* = \bar{y}^{(k)} + \Psi^{(k)}(\alpha_b^{(k)} + (\hat{\mathbf{B}}^{(k)})^{1/2} (v^{\alpha(k)} + s_k))$.

- (6) Update the trust-region radius:

- If $\rho_k \geq \eta_2$: implement outer projection $y_0^{(k+1)} = \bar{y}^{(k)} + y_0^*$ 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 $y_0^{(k+1)} = \bar{y}^{(k)} + y_0^*$ and decrease trust-region radius $\delta_{k+1} = \gamma_2 \delta_k$ and GOTO 1
- If $\rho_k \leq \eta_1$: set $y_0^{(k+1)} = y_0^{(k)}$ and decrease trust-region radius $\delta_{k+1} = \gamma_1 \delta_k$ and GOTO 3

See Figure 1 for a flow chart of the methodology combining dual-weighted snapshots and TRPOD adaptivity.

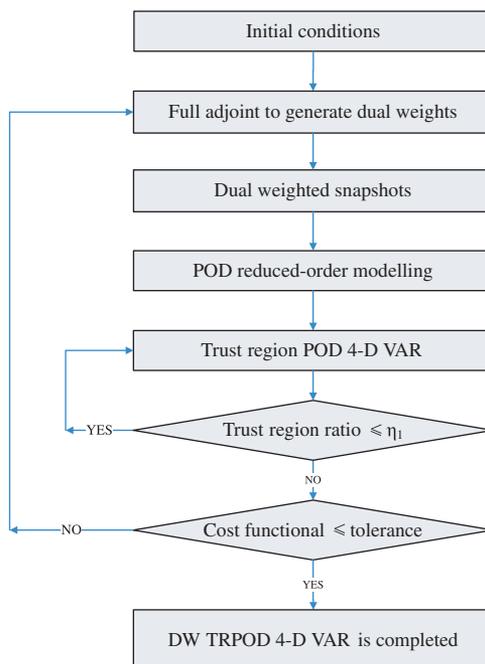


Figure 1. Flowchart of the methodology combining dual-weighted snapshots and TRPOD adaptivity.

5. NUMERICAL EXPERIMENTS

5.1. ERA-40 observations and POD reduced order model

Reanalyzed data on a $2.5^\circ \times 2.5^\circ$ grid (500 hPa pressure level-geopotential height and velocity fields) from the ERA-40, 40-yr re-analysis system (<http://www.ecmwf.int/research/era/>), valid at 0000 UTC 2 February 2001 was used to specify the initial conditions for forward model integration. These initial conditions were unchanged in all the following test cases. As for boundary conditions, since the domain being considered is spherical, it is obvious that the boundary conditions remain unchanged. The *unconstrained van Leer scheme* with a $2.5^\circ \times 2.5^\circ$ (144×72 cells) grid resolution and time step of $\Delta t = 450$ s, has been used in this article, to generate a *reference trajectory*. Synthetic observations are obtained by randomly perturbing the reference trajectory, in which the observational error covariance matrix has been taken to be a block diagonal matrix $R = [10^4 I \ 10^2 I \ 10^2 I]$. For the entries in R , the values of the variances are specified based on typical values of the variables. The zonal and meridional winds vary on a scale of 10–100 m/s. Hence, a value of 100 was specified for their variances. For the geopotential height field, $\Phi = gh$ varies on a scale of $10^4 \text{ m}^2/\text{s}^2$.

In the numerical experiment, we carried out a 1% normally distributed random perturbation on the true initial conditions over the entire vector $X = [u, v, h]$ field in Figure 2(a) specified from ERA-40 in order to provide twin-experiment ‘observations’. Also, the 18-h forecast of the FV-SWE model was taken to be forecast verification time displayed in Figure 2(b). The 4-D Var optimization loop was stopped when the l_2 norm of the gradient was less than a tolerance of 10^{-3} . Since we did not change the tolerance, the results we obtained are not affected. It is obvious that if we were to make the tolerance more stringent, the optimization would have required more iterations. The reduction of the cost functional is measured by the value of the current cost functional normalized by the initial one with or without the logarithmic scale. We computed the errors between the true initial conditions and the retrieved initial conditions related to a 1% normally distributed random perturbations of the true initial conditions as the initial guess of the reduced order 4-D Var. The data assimilation was carried on a 15 h window using the $\Delta t = 450$ s in time and a mesh of 144×72

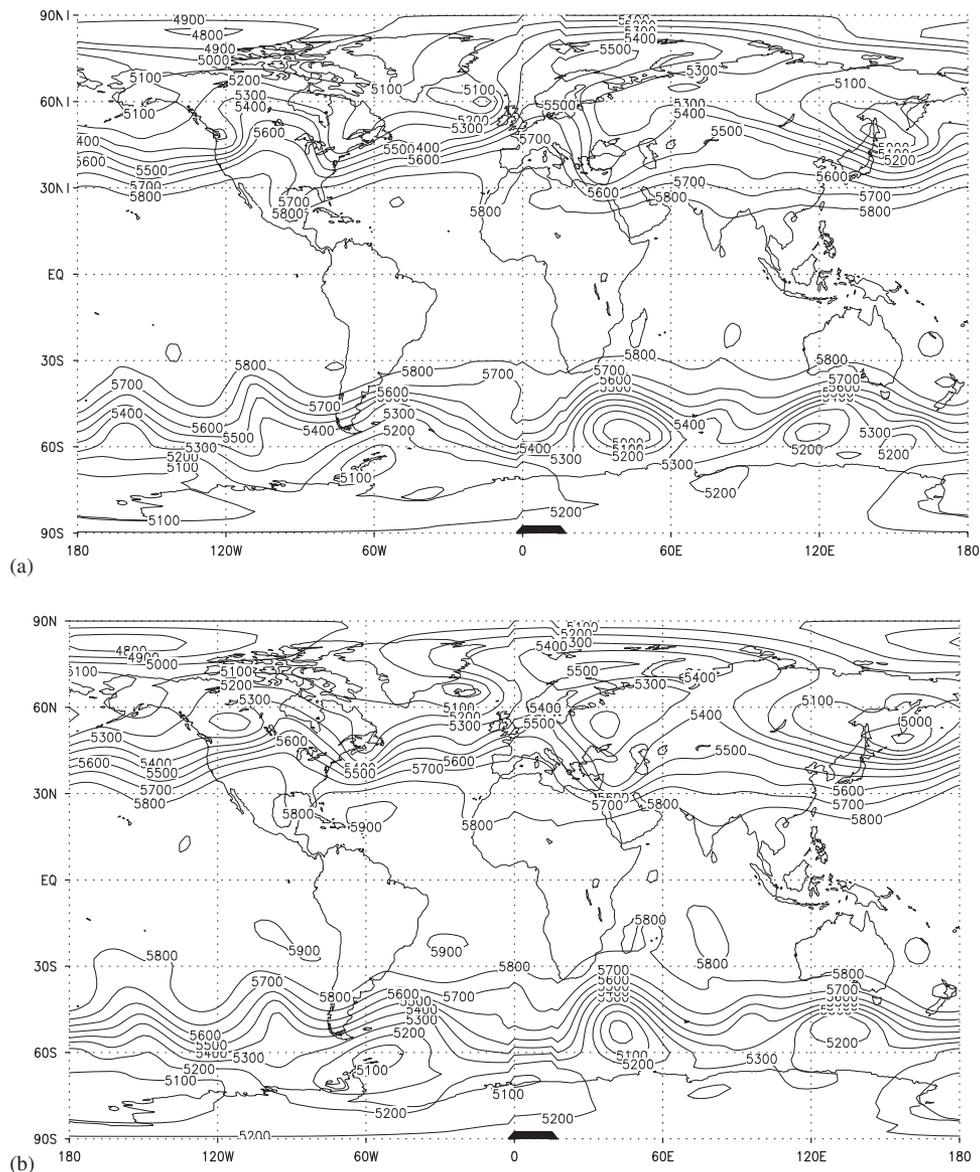


Figure 2. Isopleths of the geopotential height for the reference trajectory. (a) The configuration at the initial time specified from ERA-40 data sets and (b) the 18-h forecast of the FV-SWE model using the unconstrained Van-Leer advection scheme.

grid points in space and the observations are available every 3 h in time including the initial time. Thus, we have $144 \times 72 \times 3 \times 6$ observations distributed in time and space.

Now, we generated 120 snapshots by integrating the full FV-SWE model forward in time, from which we choose 15 POD modes or 15 DWPOD modes for each of the $(u(x, y), v(x, y), \phi(x, y))$ to capture over 99.9% of the energy. The singular value decomposition for both POD modes and DWPOD modes from the snapshots is displayed in Figure 3(a). The energy captured by the leading POD modes or DWPOD modes from the snapshots as a function of the dimension of the POD reduced space is displayed in Figure 3(b). Also, the isopleths of the POD modes of dimensions 1, 5 and 10 are displayed in Figure 4. The other POD modes, though not plotted show a gradual shift in where most energy is localized; that is, the leading POD modes display most energy uniformly distributed almost on the entire globe, whereas the latter POD modes show a shift toward the north and south poles, we attribute this observation to our particular FV-SWE model. Similar observation was made by Akella and Navon [58] in terms of where the largest

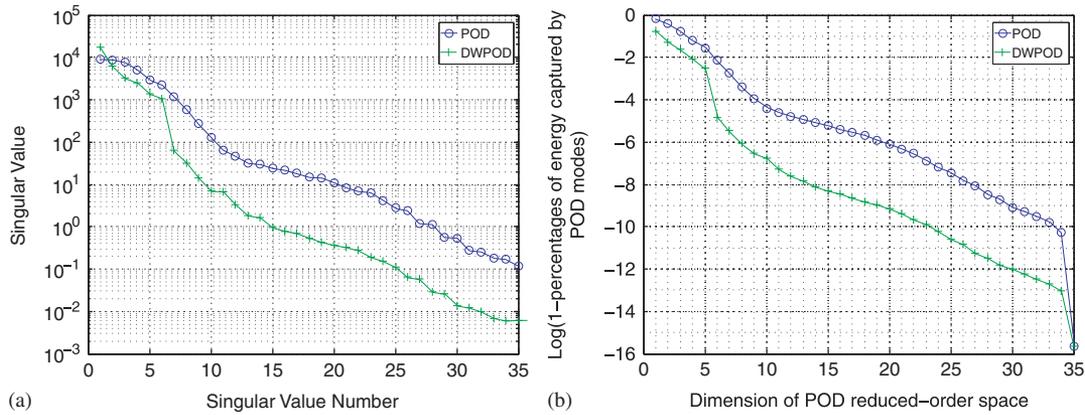


Figure 3. Singular value decomposition. (a) Unweighted SVD and dual-weighted SVD and (b) the percentage of energy captured by POD modes.

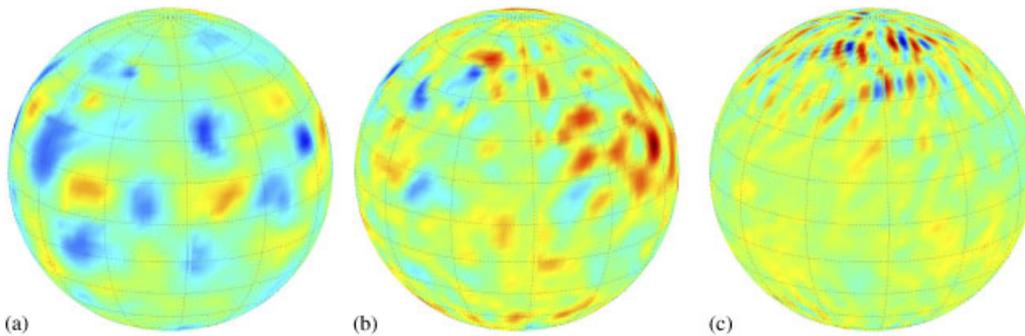


Figure 4. Isopleths of the POD modes of dimension 1, 5 and 10 respectively. (a) 1st POD mode; (b) 5th POD mode; and (c) 10th POD mode.

errors in the retrieved initial conditions were obtained in their 4-D Var twin experiments using the FV-SWE model. Furthermore, the dimension of control variables vector for the POD reduced order 4-D Var thus is $15 \times 3 = 45$ compared to $144 \times 72 \times 3 = 31\,104$ for the full 4-D Var.

5.2. POD reduced order model 4-D Var using full observations

5.2.1. POD reduced order 4-D Var experiments. Two POD reduced order 4-D Var experiments are set up, in which the first experiment, hereafter referred as DAS-I, had no background term included in the POD reduced order cost functional and the second, hereafter referred as DAS-II, had the background error covariance term included in the POD reduced order cost functional. The background state was generated using a 1% normal random perturbations on the initial conditions, in which the background error covariance matrix has been taken to be a block diagonal matrix $B = [2 \times 10^4 I \ 10^2 I \ 10^2 I]$. In practice, by applying random number generator using CPU clock cycle, we made sure that the seeds used to generate pseudonormal random perturbations for twin-experiment ‘observations’ are nearly uncorrelated with the seeds used to generate normal random perturbations for background terms in the reduced order cost functional.

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 the root mean-square error (RMSE) metrics between the full-model solutions and reduced order solutions were consistently improved after each outer projection was carried out.

The limited memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) update algorithm for quasi-Newton minimization [59] was employed for high-fidelity full-model 4-D Var and all variants of

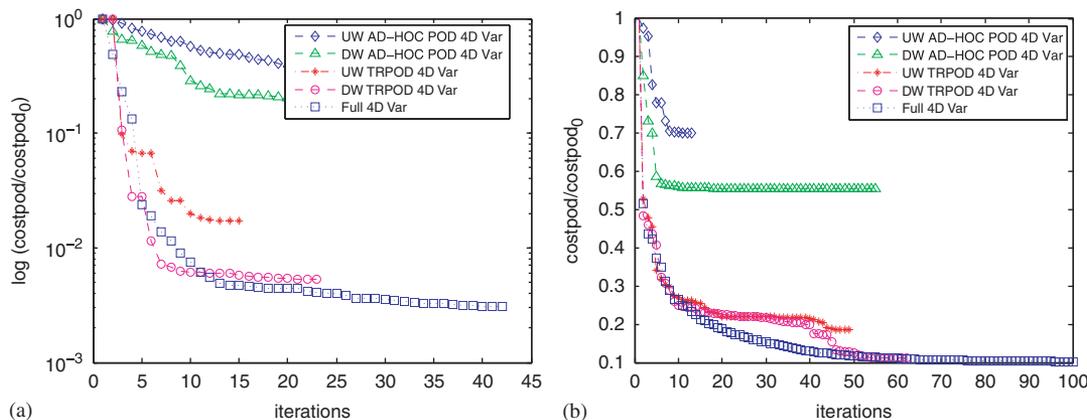


Figure 5. Comparison of the performance of the iterative minimization process of the scaled cost functional for unweighted *ad hoc* POD 4-D Var, dual-weighted *ad hoc* POD 4-D Var, unweighted trust-region POD 4-D Var, dual-weighted trust-region 4-D Var, and full-model 4-D Var, respectively. (a) DAS-I and (b) DAS-II.

ad hoc POD 4-D Var, while a variant of the LBFGS, called LBFGS-B [60, 61] which can handle box-constraints on the variables was employed for the trust-region POD 4-D Var within the trust-region radius and provides a sufficient reduction of the high-fidelity model quantified in terms of the Cauchy point [32]. In the *ad hoc* POD 4-D Var [41, 43], the POD bases are re-calculated when the value of the cost function cannot be decreased by more than a factor of 0.5 for *ad hoc* POD 4-D Var and 0.1 for *ad hoc* DWPOD 4-D Var between consecutive minimization iterations. The reason for the particular choice of these values is based on numerical experience and relative rate of convergence of the *ad hoc* and dual-weighted POD methods, respectively. 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 computational 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 of 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 to 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 in four orders of magnitude compared to the minimization of the cost functional obtained by applying the unweighted *ad hoc* POD 4-D Var (see Figures 5(a) and (b)), showing that the combination of the dual-weighted approach and trust-region method to model reduction is significantly beneficial in attaining a local minimum of the cost functional almost identical to one obtained by the high-fidelity full 4-D Var, while the computation of effort for dual-weighted trust-region POD 4-D Var is much less than the one required for full 4-D Var (see Tables I(a) and (b)).

In Figure 6(a) and (b), we found that the minimization of the cost functional using full 4-D Var will be terminated if the scaled norm of the gradient of the cost functional can decrease by 2 orders of magnitude, while the one using DWTRPOD 4-D Var will be terminated if the corresponding scaled norm of the gradient can decrease by 3 orders of magnitude, which can be explained by the fact that the POD reduced order space is dimensionally lower than the full space.

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 used the metric namely the RMSE of the difference between the POD reduced order simulation and high-fidelity model.

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

	UWAHPOD	DWAHPOD	UWTRPOD	DWTRPOD	Full
(a) DAS-I					
Iterations	23	24	16	23	42
Outer projections	2	2	14	14	NA
$\log\left(\frac{J_f}{J_0}\right)$	$10^{-0.37}$	$10^{-0.69}$	$10^{-1.78}$	$10^{-2.32}$	$10^{-2.50}$
CPU time (s)	117.1	149.2	143.2	181.7	601.7
(b) DAS-II					
Iterations	14	59	50	62	100
Outer projections	2	2	15	16	NA
$\frac{J_f}{J_0}$	0.72	0.54	0.17	0.13	0.10
CPU time (s)	100.3	207.7	280.1	352.5	966.7

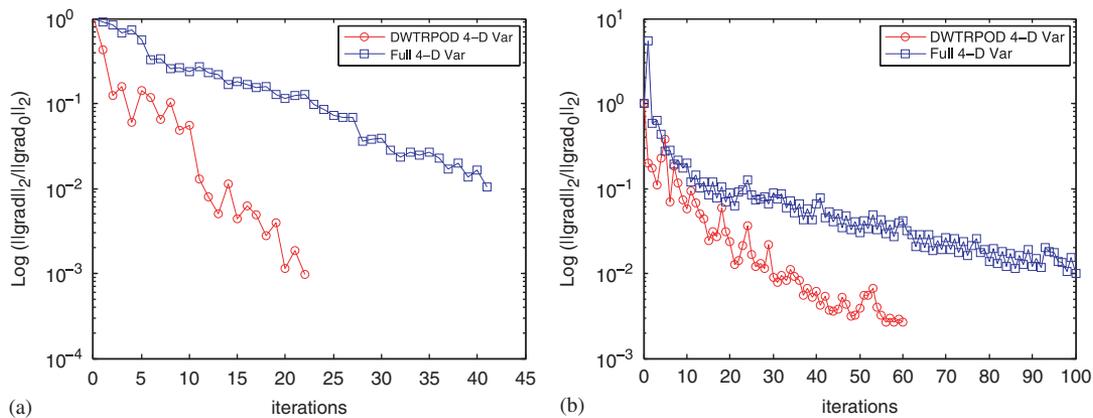


Figure 6. Comparison of the performance of the iterative minimization process of the scaled norm of the gradient of the cost functional for dual-weighted trust-region 4-D Var and full model 4-D Var. (a) DAS-I and (b) DAS-II.

In particular, the RMSE 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.

$$RMSE^i = \sqrt{\frac{\sum_{j=1}^{j=N} (U_{i,j} - U_{i,j}^{POD})^2}{N}}, \quad i = 1, \dots, n \tag{40}$$

where $U_{i,j}$ and $U_{i,j}^{POD}$ 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 those corresponding to the POD reduced order model, respectively.

Although it turned out to be advantageous to combine the dual-weighted 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 15 to 25. This remark is based on RMSE and also the difference between the 18-h forecast using true initial conditions and the one using retrieved initial condition after data assimilation. However, increasing the dimension of the POD reduced order space from 45 to 75 can increase the computational cost of POD reduced order 4-D Var. This agrees with results obtained in [35] that 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 equation models, see Daescu and Navon [34].

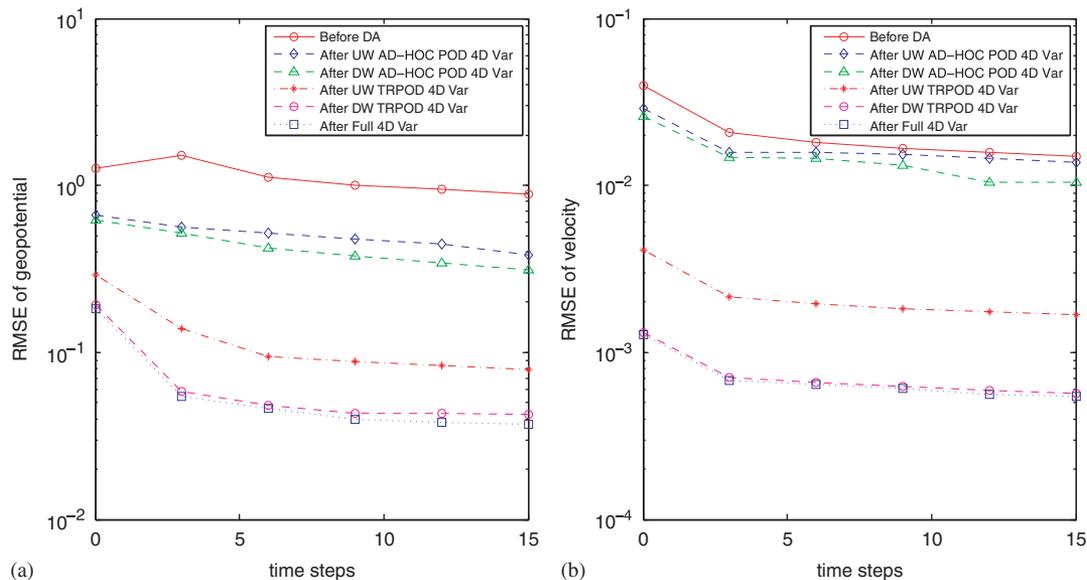


Figure 7. Comparison of the RMSE in DAS-II experiments among unweighted *ad hoc* POD 4-D Var, dual-weighted *ad hoc* POD 4-D Var, unweighted trust-region POD 4-D Var, dual-weighted trust-region 4-D Var, and full-model 4-D Var, respectively. (a) RMSE of Geopotential in DAS-II and (b) RMSE of wind velocity in DAS-II.

Finally, in Figures 8(a) and (b) we compared the errors in retrieved initial conditions without and with background error covariance terms (i.e. DAS-I and DAS-II experiments). Notice that in both cases the largest errors occur in the polar regions (see, note in Section 5.1). With the background term, we obtained an improved estimation of the true initial condition in DAS-II, compared to DAS-I, as evident through the RMSE plots (Figures 7(a) and (b)) as well. Such advantages of the background term in ‘full’ 4-D Var are well documented in [62].

5.2.2. Nonlinearity in the projection. Owing to the complexity of the Lin-Rood FV code, the numerical fluxes had to be computed at the element boundaries. This required us to go back to the full model in order to evaluate the numerical fluxes, in order to deal with the nonlinearity in the projection. The numerical problem of reducing the complexity of evaluating the nonlinear terms of the POD reduced model in the context of FV requires for this quadratic nonlinearity a pre-computing of a special POD-Galerkin projection. However, the pre-computing technique proved to be very difficult to implement due to the algorithmic features of the Lin-Rood FV scheme. This explains why we obtained only a speed up of a factor of order 3 as shown in Table I(a) and (b).

An elegant solution to this problem was put forward by Chaturantabut [63], Chaturantabut and Sorensen [64, 65] where they proposed a method referred as a Discrete Empirical Interpolation Method (DEIM). DEIM achieves a complexity reduction of the nonlinearities which is proportional to the number of reduced variables while POD retains a complexity proportional to the original number of variables. The DEIM approach approximates a nonlinear function by combining projection with interpolation. DEIM constructs specially selected interpolation indices that specify an interpolation-based projection so as to provide a nearly l_2 optimal subspace approximation to the nonlinear term, without the expense of orthogonal projection.

5.3. Results with incomplete observations

5.3.1. The observations of height field only. In DAS-II, meteorological observations are temporarily available every 3 h but spatially distributed at all the grid points. So the question arises as to what will happen if we decrease the number of observations in space [66], i.e. observational operator in the cost functional becomes a sparse matrix.

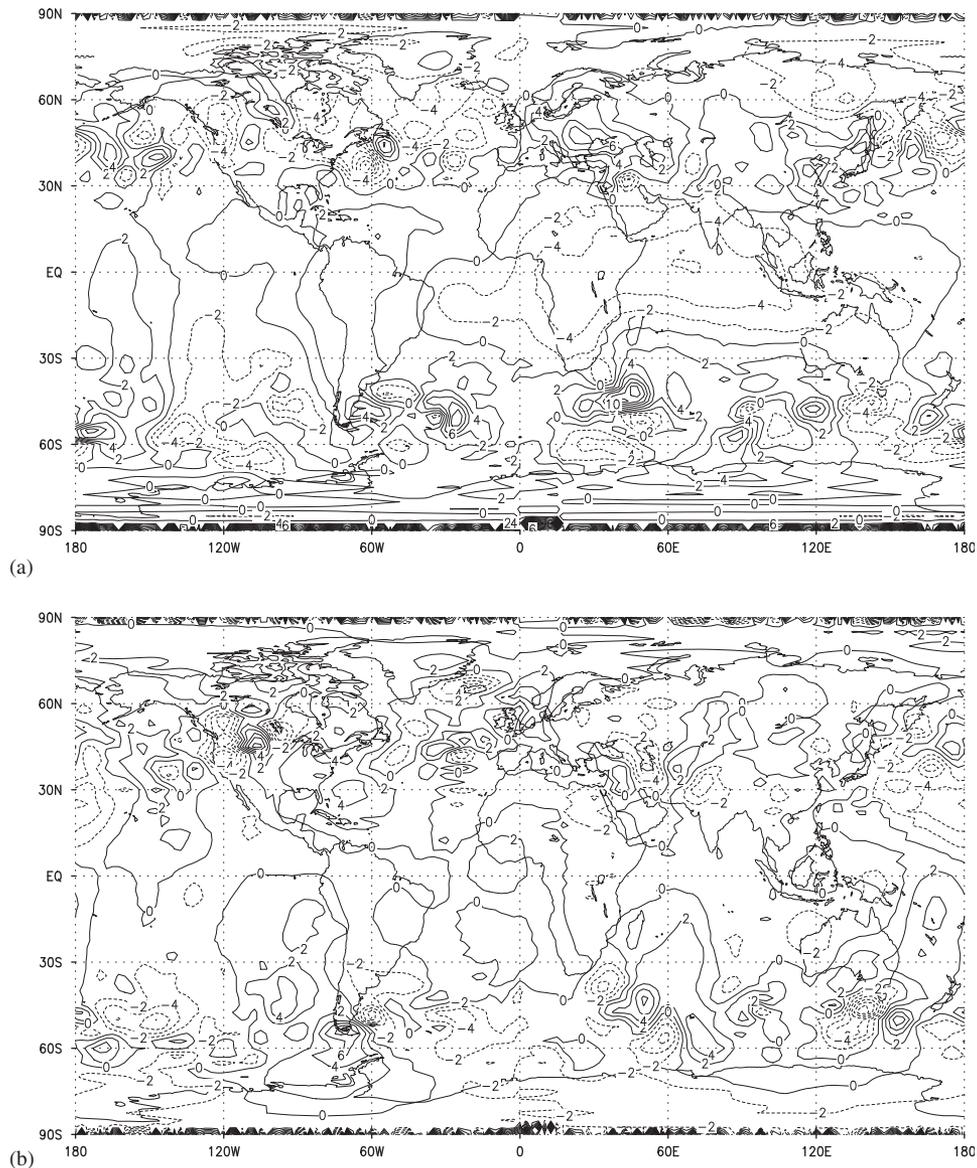


Figure 8. Isopleths (scaled by multiplying 1000) of the geopotential height for the difference between the 18-h forecast using true initial conditions and the one using retrieved initial condition after DWTRPOD 4-D Var. (a) DAS-I and (b) DAS-II.

Suppose that only the geopotential field is observed but the observations for the wind field are unavailable (i.e. the number of observations is decreased from $144 \times 72 \times 3 \times 6$ to $144 \times 72 \times 6$). We refer to this case by DAS-III(a), in which the initial perturbed field is the same as the one used to start DAS-I. In DAS-III(a), the numerical results in Figure 9(a) show that it takes more iterations for the cost functional of full 4-D Var with only incomplete observations to converge than the one with full observations. Furthermore, the POD reduced cost functional in DAS-III(a) using the UWTRPOD 4-D Var can be reduced to almost the same degree of magnitude as full 4-D Var in DAS-III(a) displayed in Figure 9(a). Also, in DAS-III(a) the norm of the gradient of the POD reduced cost functional using UWTRPOD 4-D Var and the cost functional using full 4-D Var both decrease by only 2 orders of magnitude, displayed in Figure 9(b). In Figure 10(a), an additional experiment was carried out comparing results for UWTRPOD 4-D Var, DWTRPOD 4-D Var as well as full 4-D Var in the case of observations being available only for the geopotential field. It was also found out that the results for DWTRPOD 4-D Var produced similar results as those obtained in the

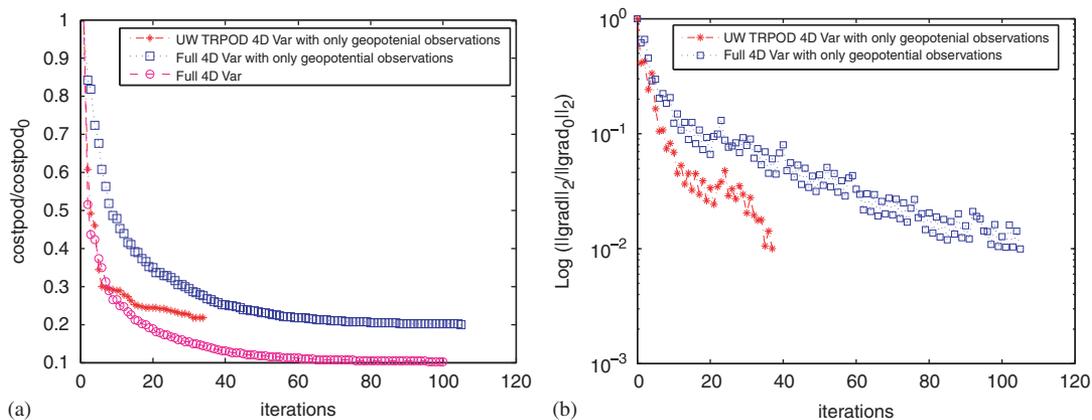


Figure 9. DAS-III(a) (Observations of height field only): comparison of the performance of the iterative minimization process of the scaled cost functional and the scaled norm of the gradient of the cost functional for unweighted trust-region POD 4-D Var and full 4-D Var. (a) Scaled cost functional and (b) scaled norm of the gradient.

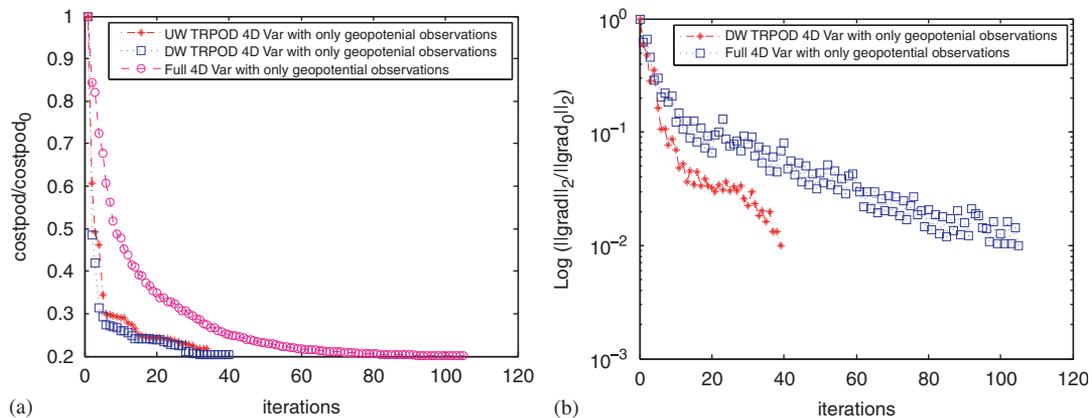


Figure 10. DAS-III(a) (Observations of height field only): comparison of the performance of the iterative minimization process of the scaled cost functional and the scaled norm of the gradient of the cost functional for unweighted trust-region POD 4-D Var, dual-weighted trust-region POD 4-D Var and full 4-D Var. (a) Scaled cost functional and (b) scaled norm of the gradient.

case of DAS III (b) (c) (d) (not shown) experiments with incomplete observations. In Figure 10(b), corresponding results were displayed for the scaled norm of the gradient for DWTRPOD 4-D Var and full 4-D Var. Again, the other experiments (not shown) exhibited similar results of incomplete observations. In Figure 11, we obtained the errors in retrieved initial conditions using UWTRPOD 4-D Var with incomplete observations (i.e. only the geopotential observations are available). Notice that in this case the largest errors are still dominant in the polar regions, while the overall RMSE becomes larger than the results obtained in DAS-II.

5.3.2. *Incomplete observations in space.* Next, we consider fewer observations along the longitudinal direction. From the earlier number of 144 observations, we specified only 72. Hence, the observational resolution is 72×72 . But we have observations for $[h, u, v]$ at every 3 h as in DAS-II. The reduction in cost functional and scaled gradient norm are plotted in Figures 12(a) and (b), respectively. Notice that the performance of both the full 4-D Var and UWTRPOD is affected due the alternating observations in one direction.

We follow on the above approach and test what happens when instead of having fewer observations along the longitudinal direction, we have lesser observations along the latitudinal direction,

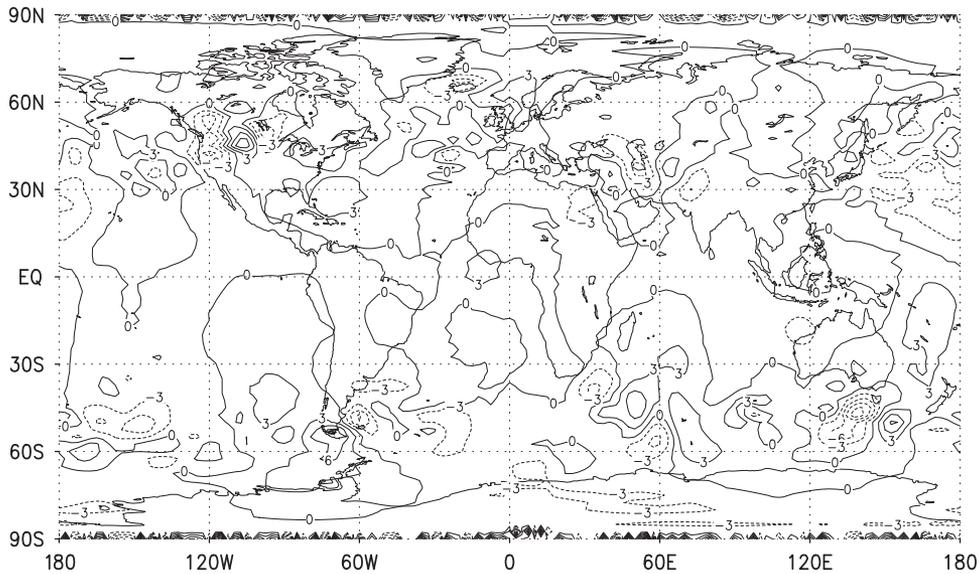


Figure 11. DAS-III(a): Isopleths (scaled by multiplying 1000) of the geopotential height for the difference between the 18-h forecast using true initial conditions and the one using retrieved initial condition after UWTRPOD 4-D Var.

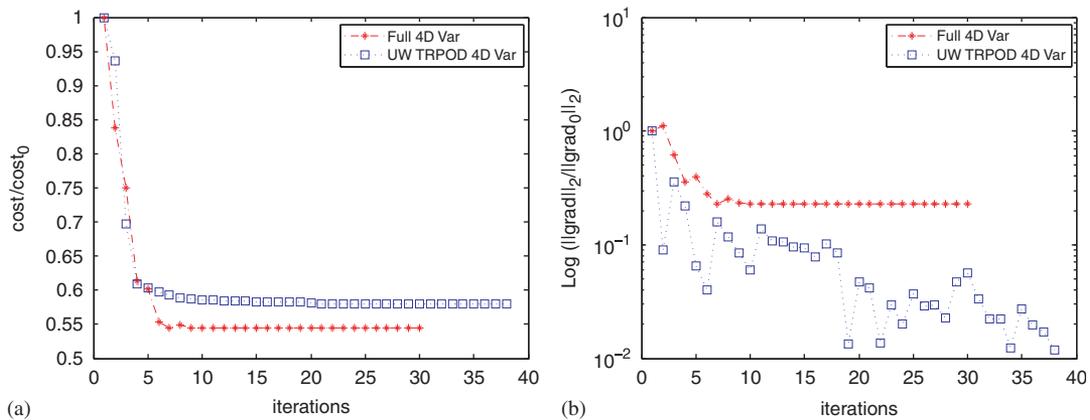


Figure 12. DAS-III(b) (5×2.5 Resolution): comparison of the performance of the iterative minimization process of the scaled cost functional and the scaled norm of the gradient of the cost functional for unweighted trust-region POD 4-D Var and full 4-D Var. (a) Scaled cost functional and (b) scaled norm of the gradient.

i.e. instead of 72, have only 36 observations, which implies an observational resolution of 144×36 . Notice that the performance is not as severely impacted (see Figures 13(a) and (b)) as in earlier results with 5×2.5 observational resolution. Based on the above two experiments, with observations at 5×2.5 and 2.5×5 grid resolutions, though the cost functional and gradient norm could be minimized, as remarked for e.g. [66], such alternating sparsity of the observations affects the condition number of the Hessian of cost functional, resulting in a poorly conditioned minimization problem. Based on our results, we remark that the POD 4-D Var also suffers from the ill-conditioning as the full 4-D Var for such an observational grid resolution.

In addition, we conducted another experiment where we retained observations of height field at all grid points, whereas the wind components, u and v were observed as follows. The observations for the winds fields were not available from 20° North/South to the North/South poles, that is, we masked the observations for u and v fields near the poles. The decrease in scaled cost and gradient norm are plotted in Figures 14(a) and (b), respectively. We note a comparable performance of the TRPOD 4D-Var and the full 4D-Var. This example illustrates that the background error covariance

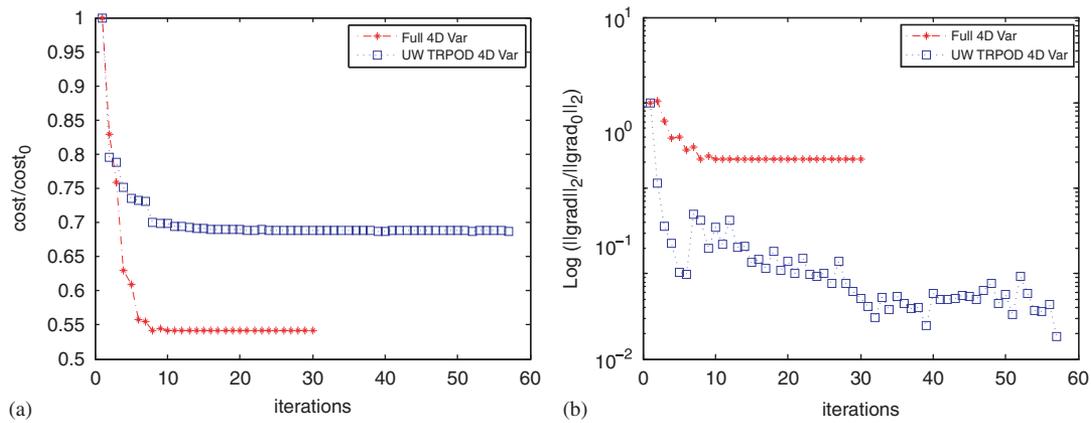


Figure 13. DAS-III(c)(2.5×5 Resolution): comparison of the performance of the iterative minimization process of the scaled cost functional and the scaled norm of the gradient of the cost functional for unweighted trust-region POD 4-D Var and full 4-D Var: (a) Scaled cost functional and (b) scaled norm of the gradient.

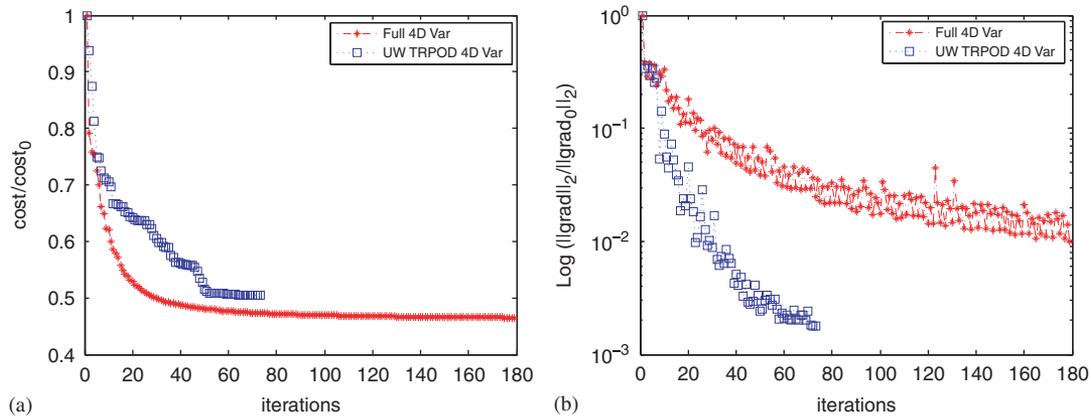


Figure 14. DAS-III(d): 2.5×2.5 Resolution with incomplete observations for u and v wind fields from 20° north to north pole and 20° south to south pole and complete observations for geopotential field, over entire globe. Comparison of the performance of the iterative minimization process of the scaled cost functional and the scaled norm of the gradient of the cost functional for unweighted trust-region POD 4-D Var and full 4-D Var. (a) Scaled cost functional and (b) scaled norm of the gradient.

(Appendix A, Section A), which was implemented using geostrophic balance assumptions is beneficial in POD 4-D Var case, just like it is for the full 4-D Var.

6. CONCLUSIONS

In this paper, we solved an inverse problem for the POD reduced order global shallow water equations model using an FV formulation, controlling its initial conditions in the presence of observations being assimilated in a time window. In this POD 4-D Var, we developed the full adjoint of the FV-SWE and by projection we obtained 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 projected the gradient of the full cost functional onto the gradient of the POD reduced order cost functional. Furthermore, after the projection of full background error covariance matrix to low-dimensional reduced space,

an ideal preconditioning of the POD 4-D Var was obtained so that the Hessian matrix of the POD reduced order background error covariance matrix became the identity matrix.

In the numerical experiments, we set up two types of 4-D Var experiments, namely, DAS-I without background terms and DAS-II with background terms. For both DAS-I and DAS-II, we compared several variants of POD 4-D Var, namely unweighted *ad hoc* POD 4-D Var, dual-weighted *ad hoc* POD 4-D Var, unweighted trust-region POD 4-D Var and dual-weighted trust-region POD 4-D Var, respectively. We found that the *ad hoc* POD 4-D Var version yielded the least reduction of 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 TRPOD allowing local minimizer of the full problem to be attained by minimizing the TRPOD subproblem. Thus, trust-region 4-D Var resulted in global convergence to the high-fidelity local minimum starting from any initial iterates. The experiments carried out in DAS-III with incomplete observational data indicate that in the case of insufficient data, the minimization is slower. Nevertheless many experiments with incomplete observations show satisfactory performance of the POD reduced 4-D Var, indicating its robustness to lack of observations.

The TRPOD approach for the optimal flow control problem can be viewed as a modification of classical trust-region method with a non-quadratic POD model function. In our context, TRPOD was thus implemented for FV-SWE model in order to obtain the robust global convergence based on only a small number of POD basis function. The dual-weighted POD selection of snapshots allows propagation of information from the data assimilation system onto the reduced order model, possibly capturing lower energy modes that may play a 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 equation models in a two-dimensional spatial domain. Therefore, the advantage of the dual-weighted TRPOD can be viewed as either the economization of the full 4-D Var without sacrificing the global convergence or as the feasibility of implementation of optimal control of a large dynamical models based on a relatively lower dimensional POD control space.

In particular, we observed that a similar reduction in cost functional and RMSE could be obtained using the POD 4-D Var method, such as the dual-weighted TRPOD compared to the full 4-D Var, but at a significantly less computational effort and reduced storage requirements (about 1/3 CPU-time less compared to full 4-D Var). These results indicate a potential for huge benefits within operational 4-D Var data assimilation systems with state-of-the-art NWP models. In order to obtain a drastic speed up of CPU time by at least an order of magnitude, we plan to explore implementation of DEIM to exploit the full potential of the POD reduced order model in the framework of dual-weighted TRPOD in our future research work.

APPENDIX A: FORMULATION OF BACKGROUND ERROR COVARIANCE TERMS

From Section 4.1, the cost functional is given by, $J = J_b + J_o$. A static-in-time \mathbf{B} is constructed in the grid point space as an operator, which is based on the formulation provided in [67] and [68].

Let $\delta\mathbf{x} = \mathbf{x}(t_0) - \mathbf{x}^b$, and define a transformation, $\mathbf{v} = \mathbf{B}^{-1/2} \delta\mathbf{x}$, which implies that $\delta\mathbf{x} = \mathbf{B}^{1/2} \mathbf{v}$. Where the $\mathbf{B}^{1/2}$ is taken to be any square-root matrix, such that $\mathbf{B} = \mathbf{B}^{1/2} \mathbf{B}^{T/2}$; $\mathbf{B}^{T/2}$ denotes the transpose of $\mathbf{B}^{1/2}$. Therefore the background cost functional can be rewritten as,

$$J_b = \frac{1}{2} \delta\mathbf{x}^T \mathbf{B}^{-1} \delta\mathbf{x} = \frac{1}{2} \delta\mathbf{x}^T (\mathbf{B}^{1/2} \mathbf{B}^{T/2})^{-1} \delta\mathbf{x} = \frac{1}{2} \mathbf{v}^T \mathbf{v}.$$

Hence, the contribution to the gradient of the cost functional, J from the background cost functional is equal to $\nabla_{\mathbf{v}} J_b = \mathbf{v}$, and to the Hessian of the cost functional, $\nabla_{\mathbf{v}}^2 J_b = I$. At the beginning of the minimization, $\mathbf{v} = \delta \mathbf{x} = 0$, such that the initial guess for $\mathbf{x}(t_0)$ is \mathbf{x}^b . This transformation of variables preconditions the minimization problem for faster convergence of the minimization algorithm. An ideal preconditioning is obtained if the Hessian matrix is an identity matrix. A good approximation to this is to ensure that the Hessian of J_b is equal to I , which is indeed the case here, since the minimization is performed in the \mathbf{v} space. To summarize,

$$J = J_b + J_o = \frac{1}{2} [\mathbf{x}(t_0) - \mathbf{x}^b]^T \mathbf{B}^{-1} [\mathbf{x}(t_0) - \mathbf{x}^b] + J_o = \frac{1}{2} \mathbf{v}^T \mathbf{v} + J_o.$$

Therefore, the gradient of the cost functional with respect to \mathbf{v} is given by,

$$\nabla_{\mathbf{v}} J = \mathbf{v} + \nabla_{\mathbf{v}} J_o = \mathbf{v} + \mathbf{B}^{T/2} \nabla_{\mathbf{x}_0} J_o.$$

Thus every minimization iteration requires application of $\mathbf{B}^{1/2}$ to obtain the analysis increment $\delta \mathbf{x}$ from \mathbf{v} and $\mathbf{B}^{T/2}$ to get the gradient $\nabla_{\mathbf{v}} J_o$ from $\nabla_{\mathbf{x}_0} J_o$ (which is computed by a single integration of the adjoint model backward in time). As evident, we do not require inverse of \mathbf{B} in the above formulation.

The model variables $(\mathbf{h}, \mathbf{u}, \mathbf{v})$ are partitioned into balanced and unbalanced components. The so-called balancing operator, \mathbf{K}_b acts on the unbalanced components of the model variables and in turn, $\mathbf{K}_b = \mathbf{K}'_b + I$. Following [68], \mathbf{K}'_b is formulated using the linear balance equations, based on geostrophic balance (written in spherical coordinates) and hydrostatic hypothesis.

Geostrophic balance:

$$u = -\frac{1}{\rho f} \left[\frac{1}{a} \frac{\partial p}{\partial \theta} \right],$$

$$v = \frac{1}{\rho f} \left[\frac{1}{a \cos \theta} \frac{\partial p}{\partial \lambda} \right].$$

Hydrostatic hypothesis: $p = \rho gh$.

Which implies,

$$u = -\frac{g}{f} \left[\frac{1}{a} \frac{\partial h}{\partial \theta} \right],$$

$$v = \frac{g}{f} \left[\frac{1}{a \cos \theta} \frac{\partial h}{\partial \lambda} \right].$$

Therefore,

$$\mathbf{K}_b = \mathbf{K}'_b + I = \begin{pmatrix} I & 0 & 0 \\ -\frac{g}{af} \frac{\partial}{\partial \theta} & I & 0 \\ \frac{g}{af \cos \theta} \frac{\partial}{\partial \lambda} & 0 & I \end{pmatrix}$$

which is a lower triangular matrix, since our control vector is of the form $(\mathbf{h}, \mathbf{u}, \mathbf{v})^T$.

Remark

At the north and south poles, one-sided differences have been used for computing the above derivative with respect to the latitude and at the equator, where $\theta = \pi/2$, we have used the average values of the derivative (with respect to the longitude) from the two neighboring latitude circles, above and below the equator.

Using the balance operator, we can write $\mathbf{B} = \mathbf{K}_b \mathbf{B}_u \mathbf{K}_b^T$, where \mathbf{B}_u is a block diagonal error covariance matrix for the unbalanced component of the variables (see, [67]), which implies that the cross-covariances between the unbalanced variables is taken to be negligible. Thus, $\mathbf{B}_u = \Sigma_{\mathbf{b}} \mathbf{C} \Sigma_{\mathbf{b}}$,

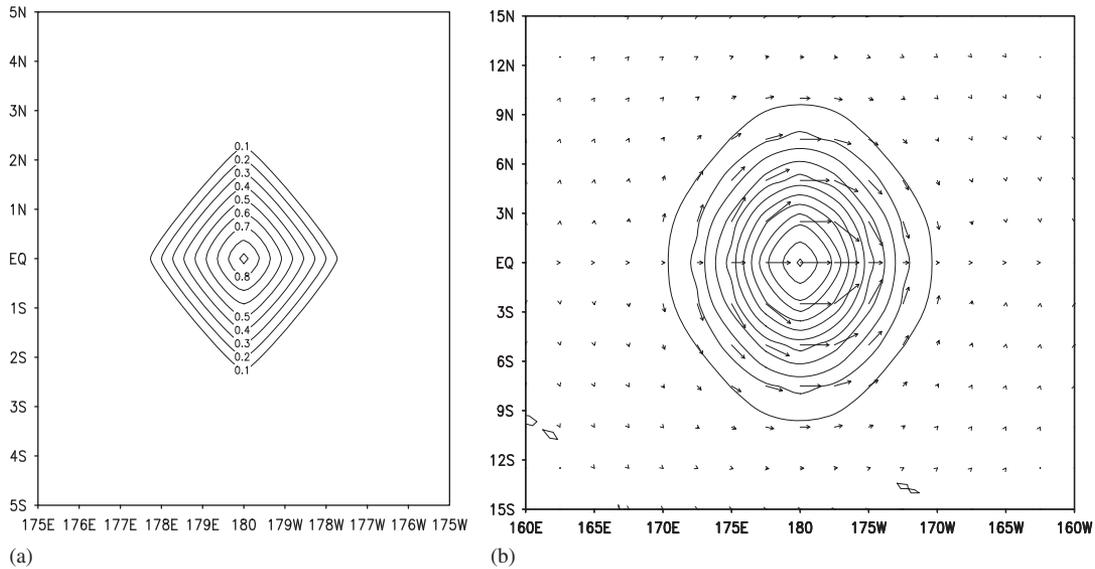


Figure A1. Result obtained by operating with \mathbf{B} on a single Dirac delta pulse in the height field (a), isolines of the height field (b), geostrophic wind plotted along with the isolines of the height field.

where Σ_b is a block-diagonal matrix of the background error variances in the grid point space, such that the diagonal entries represent error variances at every grid point (in this work, we prescribed $\Sigma_b = [2000 I, 100 I, 100 I]$).

\mathbf{C} is a symmetric matrix of background error correlations for the unbalanced component of the variables. Assuming that \mathbf{C} is block-diagonal, which is a valid assumption, since \mathbf{B}_u has already been assumed to be block-diagonal, we obtain the square-root factorization $\mathbf{C} = \mathbf{C}^{1/2} \mathbf{C}^{T/2}$.

Thus, the square-root factorization of the background error covariance can be written as,

$$\begin{aligned} \mathbf{B} &= \mathbf{K}_b \mathbf{B}_u \mathbf{K}_b^T = \mathbf{K}_b (\Sigma_b \mathbf{C} \Sigma_b) \mathbf{K}_b^T = \mathbf{K}_b (\Sigma_b \mathbf{C}^{1/2} \mathbf{C}^{T/2} \Sigma_b) \mathbf{K}_b^T \\ &= (\mathbf{K}_b \Sigma_b \mathbf{C}^{1/2}) (\mathbf{C}^{T/2} \Sigma_b \mathbf{K}_b^T) \\ &= \mathbf{B}^{1/2} \mathbf{B}^{T/2}. \end{aligned} \tag{A1}$$

Notice that the above formulation ensures that \mathbf{B} is symmetric and positive definite, both of these properties are usually required to be satisfied by any preconditioning matrix. The analysis increment is given by $\delta \mathbf{x} = \mathbf{B}^{1/2} \mathbf{v} = \mathbf{K}_b \Sigma_b \mathbf{C}^{1/2} \mathbf{v}$. Since \mathbf{C} is block-diagonal, the operation $\mathbf{C}^{1/2} \mathbf{v}$ can be split into individual operators $\mathbf{C}_\alpha^{1/2} \mathbf{v}_\alpha$, that act independently on different components of the variable \mathbf{v} , such as \mathbf{v}_α . For each variable, the univariate operator can be factorized into $\mathbf{C}_\alpha = \mathbf{C}_\alpha^{1/2} \mathbf{C}_\alpha^{T/2}$. The procedure suggested by Weaver and Courtier [67] and Derber and Bouttier [68] has been implemented to model the univariate correlation operator, \mathbf{C}_α as an isotropic diffusion operator, assuming Gaussianity with a decorrelation length equal to 500 km.

We considered height field which was comprised a single Dirac delta pulse located at equator and longitude 180° , and prescribed no wind field, the action of \mathbf{B} on such a field is shown in Figure A1(a). We see the effect of the correlation operator on the Dirac pulse and also on the wind field obtained under geostrophic balance assumption (Figure A1(b)), which is parallel to the isobars of the pressure. Since there is a *high pressure* at the center, the direction of the wind is clockwise in the Northern hemisphere and anti-clockwise in the Southern hemisphere; at the equator due to the balancing of the pressure gradient and Coriolis forces, the wind blows straight.

ACKNOWLEDGEMENTS

The authors thank the ECMWF for providing us the ERA-40 data, which was obtained from the ECMWF data server. The research of Prof. I. M. Navon is supported by the National Science Foundation (NSF), grant ATM-03727818. The authors thank two anonymous reviewers that helped to clarify the presentation of this paper.

REFERENCES

1. Tan WY. *Shallow-water Hydrodynamics: Mathematical Theory and Numerical Solution for a Two-dimensional System of Shallow-water Equations*. Elsevier Oceanography Series: Beijing, 1992.
2. Vreugdenhil CB. *Numerical Methods for Shallow-Water Flow*. Kluwer Academic Publishers: Boston, 1994.
3. 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.
4. Lin SJ, Chao WC, Sud YC, Walker GK. A class of the van Leer transport schemes and its applications to the moisture transport in a general circulation model. *Monthly Weather Review* 1994; **122**:1575–1593.
5. Lin SJ, Rood RB. Multidimensional flux-form semi-Lagrangian transport schemes. *Monthly Weather Review* 1996; **124**:2046–2070.
6. Lin SJ, Rood RB. An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Quarterly Journal of the Royal Meteorological Society* 1997; **123**:2477–2498.
7. Lin SJ, Rood RB. A finite-volume integration method for computing pressure gradient force in general vertical coordinates. *Quarterly Journal of the Royal Meteorological Society* 1997; **123**:1749–1762.
8. Lin SJ. A ‘vertically Lagrangian’ finite-volume dynamical core for global models. *Monthly Weather Review* 2004; **132**:2293–2307.
9. 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.
10. Chen X, Navon IM, Fang F. A dual-weighted trust-region adaptive POD 4D-Var applied to a finite-element shallow-water equations model. *International Journal for Numerical Methods in Fluids* 2009; DOI: 10.1002/flid.2198.
11. Holmes P, Lumley JL, Berkooz G. *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge Monographs on Mechanics: London, 1996.
12. 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.
13. Karhunen K. *Zur Spektral Theorie Stochastischer Prozesse*. Annals of the New York Academy of Sciences: Feniceae, 1946.
14. Loeve M. *Fonctions Aleatoires De Second Ordre*. Comptes rendus de l’Académie des sciences: Paris, 1945.
15. Pearson K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 1901; **2**(1):559–572.
16. Levenberg K. A method for the solution of certain problems in least squares. *Quarterly Journal on Applied Mathematics* 1944; **2**:164–168.
17. Morrison DD. *Proceedings of the Seminar on Tracking Programs and Orbit Determination*. Jet Propulsion Laboratory, U.S.A., 1960.
18. Marquardt D. An Algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 1963; **11**:431–441.
19. Winfield D. *Function and Functional Optimization by Interpolation in Data Tables*. Cambridge, U.S.A., 1969.
20. Powell MJD. A fortran subroutine for solving systems of nonlinear algebraic equations. 1970; 115–161.
21. Powell MJD. A fortran subroutine for unconstrained minimization requiring first derivatives of the objective function. 1970; R-6469.
22. Dennis Jr JE. A brief introduction to quasi-Newton methods. In *Numerical Analysis*, Golub G, Olinger J (eds). Proceedings of Symposia in Applied Mathematics Series, vol. 22. AMS: Providence, RI, 1978.
23. More JJ, Sorensen DC. Computing a trust region step. *SISSC* 1983; **4**(3):553–572.
24. Fahl M. Trust-region methods for flow control based on reduced order modeling. *Ph.D. Thesis*, Trier University, 2000.
25. Arian E, Fahl M, Sachs EW. Trust-region proper orthogonal decomposition for flow control. Institute for Computer Applications in Science and Engineering. *Technical Report: TR-2000-25*, 2000.
26. Bergmann M, Cordier L, Brancher JP. Control of the cylinder wake in the laminar regime by trust-region methods and POD reduced order models. *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, Seville, Spain, 12–15 December 2005.
27. Bergmann M, Cordier L, Brancher JP. Drag minimization of the cylinder wake by trust-region proper orthogonal decomposition. *Notes on Numerical Fluid Mechanics and Multidisciplinary Design* 2007; **95**:309–324.
28. 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.
29. Conn AR, Gould NIM, Toint PL. *Trust-region Methods*. SIAM: Philadelphia, 2000.
30. 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.

31. Gratton S, Sartenaer, Toint PL. Recursive trust-region methods for multiscale nonlinear optimization. *SIAM Journal on Optimization* 2008; **19**(1):414–444.
32. Nocedal J, Wright SJ. *Numerical Optimization* (2nd edn). Springer Series in Operations Research and Financial Engineering, 2006.
33. 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.
34. 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.
35. Daescu DN, Navon IM. A dual-weighted approach to order reduction in 4DVAR data assimilation. *Monthly Weather Review* 2008; **136**(3):1026–1041.
36. Afanasiev K, Hinze M. Adaptive control of a wake flow using proper orthogonal decomposition. Preprint, 1999.
37. Hinze M, Kunish K. Three control methods for time-dependent fluid flow. *Flow, Turbulence and Combustion* 2000; **65**:273–298.
38. Kunisch K, Volkwein S. Galerkin proper orthogonal decomposition methods for parabolic problems. *Numerische Mathematik* 2001; **90**(1):117–148.
39. 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.
40. Kunish K, Volkwein S. Proper orthogonal decomposition for optimality systems. *Mathematical Modelling and Numerical Analysis* 2008; **42**(1):1–23.
41. 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.
42. 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.
43. 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.
44. Vermeulen PTM, Heemink AW. Model-reduced variational data assimilation. *Monthly Weather Review* 2006; **134**(10):2888–2899.
45. Luo Z, Zhou Y, Yang XZ. A reduced finite element formulation based on proper orthogonal decomposition for Burgers equation. *Applied Numerical Mathematics* 2009; **59**(8):1933–1946.
46. Sachs EW, Schu M. Reduced order models(POD) for calibration problems in finance. *Numerical Mathematics and Advanced Applications*, ENUMATH, 2007; 735–742.
47. Altaf MU, Heemink AW, Verlaan M. Inverse shallow water flow modeling using model reduction. *International Journal for Multiscale Computational Engineering* 2009; Submitted.
48. 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.
49. Rowley CW, Colonius T, Murray RM. Model reduction for compressible flows using POD and Galerkin projection. *Physica D* 2004; **189**(1–2):115–129.
50. Noack BR, Michael S, Morzynski M, Tadmor G. System reduction strategy for Galerkin models of fluid flows. *International Journal for Numerical Methods in Engineering* 2010; **63**(2):231–248.
51. 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; **22**(3):Article no. 034102. DOI: 10.1063/1.3298960.
52. Van Leer B. Towards the ultimate conservative difference scheme I. The quest for monotonicity. *Proceedings of the Third International Conference on Numerical Methods in Fluid Mechanics*, Lecture Notes in Physics, vol. 18, 1973; 163–198. DOI: 10.1007/BFb0118673.
53. Van Leer B. Towards the ultimate conservative difference scheme: a new approach to numerical convection. *Journal of Computational Physics* 1977; **23**:276–299.
54. Van Leer B. Towards the ultimate conservative difference scheme: a second order sequel to Godunov’s method. *Journal of Computational Physics* 1977; **32**:101–136.
55. Celis M, Dennis JE, Tapia RA. A trust region strategy for nonlinear equality constrained optimization. In *Numerical Optimization 1984*, Boggs P, Byrd R, Schnabel R (eds). SIAM: Philadelphia, 1985; 71–82.
56. Akella S. Deterministic and stochastic aspects of data assimilation. *Ph.D. Thesis*, 2006. Available from: <http://etd.lib.fsu.edu/theses/available/etd-04052006-192927/>.
57. Bergmann M, Bruneau C, Iollo A. Enablers for robust pod models. *Journal of Computational Physics* 2009; **228**(2):516–538.
58. Akella S, Navon IM. Different approaches to model error formulation in 4D-Var: a study with high resolution advection schemes. *Tellus* 2009; **61**(A):112–128.
59. Liu DC. Nocedal, a limited memory algorithm for bound constrained optimization. *Mathematical Programming* 1989; **45**:503–528.
60. Byrd RH, Lu P, Nocedal J. On the limited memory BFGS method for large scale minimization. *SIAM Journal on Scientific and Statistical Computing* 1995; **16**(5):1190–1208.
61. Zhu C, Byrd RH, Nocedal J. L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. *ACM Transactions on Mathematical Software* 1997; **23**(4):550–560.

62. Bannister RN. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society* 2008; **134**:1951–1970.
63. Chaturantabut S. Dimension reduction for unsteady nonlinear partial differential equations via empirical interpolation methods. *Master of Arts Thesis*, 2008; Rice University.
64. Chaturantabut S, Sorensen DC, Steven JC. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing* 2010; **32**(5):2737–2764.
65. Anthony RK, Chaturantabut S, Sorensen DC. Morphologically accurate reduced order modeling of spiking neurons. *Journal of Computational Neuroscience* 2010; **28**:477–494.
66. Zou X, Navon IM, Le Dimet FX. Incomplete observations and control of gravity waves in variational data assimilation. *Tellus* 1992; **44**(A):273–296.
67. Weaver A, Courtier P. Correlation modeling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society* 2001; **121**:1815–1846.
68. Derber JC, Bouttier F. A reformulation of the background error covariance in the ECMWF global data assimilation. *Tellus* 1999; **51A**:195–221.