Optimal Control of a Finite-Element Limited-Area Shallow-Water Equations Model

Xiao Chen

Department of Mathematics, Florida State University, Tallahassee, Florida

Email: xc06@fsu.edu

I. M. Navon

Department of Scientific Computing, Florida State University, Tallahassee, Florida

Email: inavon@fsu.edu

Dedicated to Professor Neculai Andrei on the occasion of his 60th birthday.

Abstract: Optimal control of a finite element limited-area shallow water equations model is explored with a view to applying variational data assimilation(VDA) by obtaining the minimum of a functional estimating the discrepancy between the model solutions and distributed observations. In our application, some simplified hypotheses are used, namely the error of the model is neglected, only the initial conditions are considered as the control variables, lateral boundary conditions are periodic and finally the observations are assumed to be distributed in space and time. Derivation of the optimality system including the adjoint state, permits computing the gradient of the cost functional with respect to the initial conditions which are used as control variables in the optimization. Different numerical aspects related to the construction of the adjoint model and verification of its correctness are addressed. The data assimilation set-up is tested for various mesh resolutions scenarios and different time steps using a modular computer code. Finally, impact of large-scale minimization solvers L-BFGS is assessed for various lengths of the time windows.

Keywords: Variational data assimilation; Shallow-Water equations model; Galerkin Finite-Element; Adjoint model; Limited-area boundary condition.

Xiao Chen holds a M.Sc. degree in Applied Mathematics(2006) from Zhejiang University, Hangzhou, China. Currently, he is enrolled as a doctoral student in the department of mathematics at Florida State University, USA. His research interests include optimal control in Fluid Dynamics, four dimensional variational data assimilation method, probability and stochastic process, scientific computing and automatic differentiation.

Ionel Michael Navon graduated in Mathematics and Physics from Hebrew University,Jerusalem and holds a M.Sc. in Atmospheric Sciences from the same University and a Ph.D. degree in Applied Mathematics from University of Witwatersrand, Johannesburg (1979). He is presently Program Director and Professor in the Department of Scientific Computing at Florida State University , Tallahassee, FL where he joined in 1985. He served previously as Chief Research Officer at National Research institute for Mathematical Sciences in Pretoria, South-Africa.(1975-1984) He is the author of more than 160 peer reviewed high impact journal papers in areas of optimal control, data assimilation, parameter estimation, finite element modeling, large-scale numerical optimization and model reduction applied to the geosciences that are highly cited on ISI Web of Science, along with more than 100 scientific and technical reports.. He is also co-author of one book on adjoint sensitivity as well as contributor of chapters in a dozen books including recently the Handbook for Numerical Analysis Series (Elsevier) and a chapter on Data Assimilation in a recent Springer book on data assimilation.(2009). He is a Fellow of the American Meteorological Society, was senior NRC fellow, former Editor of major journals in applied mathematics and atmospheric sciences and is presently Editor of the International Journal for Numerical Methods in Fluids (Wiley) and is included since 1992 in Who's Who in America. Prof. Navon educated 7 doctoral students in Applied Mathematics serving as their major Professor. His current research interests focus on all aspects of POD model reduction, data assimilation in the geosciences, large-scale optimization, optimal control and ensemble filters. His research is funded by the National Science Foundation and NASA. He collaborates since 2002 with a major research group at Imperial College as well as with a research group on ocean data assimilation at IAP, Academia Sinica, Beijing, China.

1. Introduction

This paper explores the feasibility of carrying out a modular structured variational data assimilation (VDA) using a finite-element method of the nonlinear shallow water equations model on a limited area domain, in which we improve the methodology(Courtier and Talagrand 1987; Zhu et al. 1994) and addresses issues in the development of the adjoint of a basic finite-element model. Specific numerical difficulties in the adjoint derivation, for example, the treatment of the adjoint of the iterative process required for solving the systems of linear algebraic equations resulting from the finite-element discretizations using Crank-Nicholson time

differencing scheme (see Wang et al. 1972; Douglas and Dupont 1970) are explicitly addressed. The systems of algebraic linear equations resulting from the finite-element discretizations of the shallow-water equations model were solved by a Gauss-Seidel iterative method. To save computer memory, a compact storage scheme for the banded and sparse global matrices was used (see Hinsman, 1975). We emphasize the development of the tangent linear (TLM) and the adjoint models of the finite-element shallow-water equations model and illustrate its use on various retrieval cases when the initial conditions are served as control variables.

The plan of this paper is as follows. The finite-element Galerkin method for the shallow-water equations model on an f plane, the derivation of its tangent linear model and its adjoint are briefly described in section 2. The full finite element discretizations of the model of the nonlinear shallow-water equations model is described in section 3. Section 4 introduces the optimal control methodology including the development of the tangent linear model and its adjoint as well as formulation of the cost functional aimed at allowing the derivation of optimal initial conditions reconciling model forecast and observations in a window of data assimilation by minimizing the cost functional measuring lack of fit between model forecast and observations. Particular attention is paid to the development of adjoint of iterative Gauss-Seidel solver. Verification of the correctness of the adjoint is carried out in a detailed manner for all stages of the calculations (i.e. TLM, adjoint and gradient test).

Set-up of numerical experiments and the experimental design are detailed in Section 5. Basic assimilation experiments using a random perturbation of the initial conditions as observations and their results are presented. Particular attention is paid to the effectiveness of limited memory Quasi-Newton method L-BFGS for minimizing the cost functional in retrieving optimal initial conditions.

Various scenarios involving mesh resolution, different time steps as well as various lengths of the assimilation windows are tested and numerical conclusions are drawn. Finally Section 6 presents Summary and Conclusions. A detailed description of the entire optimal control set-up code organization is provided and illustrated in Appendix A.

2. Description of Problems

2.1 Shallow-Water equations model on an f plane

The shallow-water equations model is one of the simplest forms of the equations of motion for incompressible fluid for which the depth is relatively small compared to the horizontal dimensions, which can be applied to describe the horizontal structure of an atmosphere. They describe the evolution of an incompressible fluid in response to gravitational and rotational accelerations (See Tan 1992 and Vreugdenhil 1994 Galewsky 2004).

The shallow-water equations can be written as:

$$\frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \nabla \vec{v} + \nabla \phi + f \vec{k} \times \vec{v} = 0 \qquad (1)$$

$$\frac{\partial \phi}{\partial t} + \nabla \cdot (\phi \vec{v}) = 0 \qquad (2)$$

$$(x, y) \in [0, L] \times [0, D], \quad t \geq 0$$

where L and D are the dimensions of a rectangular domain of integration, $\vec{v}$ is a vector function:

$$\vec{v} = (u(x, y, t), v(x, y, t)) \qquad (3)$$

where 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 vector $\vec{k}$ is the vertical unit vector pointing away from the center of the planet. The scalar function f is the Coriolis parameter defined by the β-plane approximation:

$$f = \hat{f} + b\left(y - \frac{D}{2}\right) \qquad (4)$$

The Coriolis parameter

$$\hat{f} = 2\Omega \sin\theta \qquad (5)$$

is defined at a mean latitude θ_0, where Ω is the angular velocity of the earth's rotation and θ is latitude.

2.2 Initial and boundary conditions

The shallow-water equations require specifying appropriate initial and boundary

conditions. An initial condition is imposed as:

$$w(x,y,0) = \varphi(x,y), \quad (6)$$

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

$$w(0,L,t) = w(0,D,t), \quad (7)$$

while solid wall boundary condition in y-direction is:

$$\vec{v}(x,0,t) = \vec{v}(x,D,t) = 0. \quad (8)$$

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

2.3 Linearization of the Shallow-Water equations model

The linearization of the shallow-water equations model (1) - (2) can be written as:

$$\frac{\partial \vec{v}'}{\partial t} + \vec{v}'\cdot\nabla\vec{v} + \vec{v}\cdot\nabla\vec{v}' + \nabla\phi' + f\vec{k}\times\vec{v}' = 0 \quad (9)$$

$$\frac{\partial \phi'}{\partial t} + \nabla\cdot(\phi'\vec{v}) + \nabla(\phi\vec{v}') = 0 \quad (10)$$

where the prime denotes a perturbation around the basic state variables.

The form above can also be written explicitly (Jacques Blum, François-Xavier Le Dimet, I. Michael Navon 2008) as continuous tangent linear model (TLM):

$$\frac{\partial u'}{\partial t} + u'\frac{\partial u}{\partial x} + v'\frac{\partial u}{\partial y} + \frac{\partial \phi'}{\partial x} + u\frac{\partial u'}{\partial x} + v\frac{\partial u'}{\partial y} - fv' = 0$$

$$\frac{\partial v'}{\partial t} + u'\frac{\partial v}{\partial x} + v'\frac{\partial v}{\partial y} + \frac{\partial \phi'}{\partial y} + u\frac{\partial v'}{\partial x} + v\frac{\partial v'}{\partial y} + fu' = 0$$

$$\frac{\partial \phi'}{\partial t} + \frac{\partial(\phi' u)}{\partial x} + \frac{\partial(\phi' v)}{\partial y} + \frac{\partial(\phi u')}{\partial x} + \frac{\partial(\phi v')}{\partial y} = 0$$

and its first order continuous adjoint model with weighting forcing terms may be written as:

$$\frac{\partial u^*}{\partial t} = -\left(-u\frac{\partial u^*}{\partial x} - \frac{\partial(vu^*)}{\partial y} + v^*\frac{\partial v}{\partial x} + fv^* - \phi\frac{\partial \phi^*}{\partial x}\right) + W_u(u-u')$$

$$\frac{\partial v^*}{\partial t} = -\left(u^*\frac{\partial u}{\partial y} - fu^* - v\frac{\partial v^*}{\partial y} - \frac{\partial(uv^*)}{\partial x} - \phi\frac{\partial \phi^*}{\partial y}\right) + W_v(v-v')$$

$$\frac{\partial \phi^*}{\partial t} = -\left(\frac{\partial u^*}{\partial x} - \frac{\partial v^*}{\partial y} - u\frac{\partial \phi^*}{\partial x} - v\frac{\partial \phi^*}{\partial y}\right) + W_\phi(\phi-\phi')$$

with final conditions equal to zeros:

$$u(T) = v(T) = \phi(T) = 0$$

By integrating the first order continuous adjoint model reversely in time, the gradient of a given cost functional J is obtained by the adjoint model solutions as follows:

$$\nabla J(w_0) = \nabla J(u_0, v_0, \phi_0) = w^*(0) = \begin{pmatrix} u^*(0) \\ v^*(0) \\ \phi^*(0) \end{pmatrix}$$

where $w^* = (u^*, v^*, \phi^*)$ is the first order adjoint variable vector, W_u, W_v, W_ϕ are weighting factors which are chosen to be the inverse of estimates of the statistical root-mean-square observational errors on geopotential and wind components respectively. In our test problem, values of $W_f = 10^{-4} m^{-4} s^4$ and $W_u = W_v = 10^{-2} m^{-2} s^2$ are used.

The operator form of the discretized (9)-(10) can be written as (see Navon et al. 1992)

$$w'(x,y,t) = \mathbf{P}(w(x,y,t))w'(x,y,0) \quad (11)$$

where the control variable $w'(x,y,0)$ is the random perturbation variable of the initial state variable $w(x, y, 0)$, while $\mathbf{P}(w(x,y,t))$ represents the tangent linear operator, so that we can obtain the control variable $w'(x,y,t)$ that contains the values of wind fields and geopotential field at the final time step.

Generally speaking, there are two approaches which could be employed for calculating the gradient of the cost functional with respect to the initial conditions of shallow water equations. The first approach is called continuous adjoint, in which we need to differentiate the nonlinear shallow water equations model with respect to its initial conditions first and then discretize its adjoint PDE to compute the approximate gradient of the given cost functional. Another approach is called discrete approach, in which we need to approximate the nonlinear PDE by a discretized nonlinear system of equations first and then

differentiate the discretized nonlinear system with respect to the parameters. The discrete adjoint approach is easy to implement with the help of automatic differentiation tools, such as ADIFOR and TAMC. In the following sections, we demonstrate the methodology of discrete adjoint to carry on the VDA.

3. Discretization of the Shallow-Water Equations Model

3.1 Formulation of Galerkin Finite-Element model

We employ linear piecewise polynomials on triangular elements in the formulation of Galerkin Finite-Element model (1) - (2) for the sake of of simplicity. Over each given element, a variable ξ can be written as (see Zienkiewicz 2005)

$$\xi_{el} = \sum_{j=1}^{3} \xi_j(t) V_j(x,y)$$

where $\xi_j(t)$ represents the scalar node value of variable ξ_j at the node of the triangular element, and V_j represents a basis function(interpolation function) defined by the coordinates of the nodes.

The advection terms in the continuity equation (2) are usually integrated by parts using Green's theorem to shift the derivative from the variable to the basis function, which yields:

$$\left\langle \frac{d\phi}{dt}, V_i \right\rangle + \left\langle \nabla \cdot (\phi \vec{v}), V_i \right\rangle = 0 \qquad (12)$$

$$\Rightarrow \left\langle \frac{d\phi}{dt}, V_i \right\rangle + \int \nabla \cdot (\phi V_i \vec{v}) - \left\langle \phi \vec{v}, \nabla V_i \right\rangle = 0, \quad (13)$$

where the notation:

$$\left\langle \vec{f}, V_i \right\rangle = \sum_{elements}^{M} \int \int \vec{f}(x,y) \cdot V_i dxdy \quad (14)$$

defines the inner product when a function is multiplied by the trial function V_i. where $\cdot$ represents the inner product between two real vectors. In Galerkin FEM method, we choose the trial function to coincide with the test function. Taking into account the boundary conditions (see Navon 1979), the second term of equation (13) vanishes so that we obtain the final expression for the continuity equation:

$$\left\langle \frac{d\phi}{dt}, V_i \right\rangle - \left\langle \phi \vec{v}, \nabla V_i \right\rangle = 0. \qquad (15)$$

Following the Galerkin FEM, the momentum equation (1) becomes:

$$\left\langle \frac{d\vec{v}}{dt}, V_i \right\rangle + \left\langle \vec{v} \cdot \nabla \vec{v}, V_i \right\rangle + \left\langle \nabla \phi, V_i \right\rangle + \left\langle f\vec{k} \times \vec{v}, V_i \right\rangle = 0. \qquad (16)$$

Over each element, we denote wind fields and geopotential fields

$$\vec{v} = \sum_{j=1}^{3} \vec{v}_j(t) V_j(x,y), \phi = \sum_{j=1}^{3} \phi_j(t) V_j(x,y) \quad (17)$$

where $\vec{v}_j(t)$ and $\phi_j(t)$ are the time-dependent nodal values of wind fields and geopotential fields respectively.

Upon substituting (17) into (15)-(16), one obtains:

$$\left\langle \frac{d\phi_j}{dt} V_j, V_i \right\rangle - \left\langle \phi_j \vec{v}_k V_j V_k, \nabla V_i \right\rangle = 0, \quad (18)$$

$$\left\langle \frac{d\vec{v}}{dt} V_j, V_i \right\rangle + \left\langle \vec{v}_k \cdot \nabla \vec{v}_k, V_i \right\rangle + \left\langle \nabla \phi_k, V_i \right\rangle + \left\langle f\vec{k} \times \vec{v}_k, V_i \right\rangle = 0. \qquad (19)$$

According to the definition (14), we may write (18) explicitly as:

$$\left\langle \frac{\partial \phi_j}{\partial t} V_j, V_i \right\rangle - \left\langle \phi_j u_k V_j V_k, \frac{\partial V_i}{\partial x} \right\rangle - \left\langle \phi_j v_k V_j V_k, \frac{\partial V_i}{\partial y} \right\rangle = 0. \qquad (20)$$

We may also write (19) explicitly as:

$$\left\langle \begin{pmatrix} \frac{\partial u_j}{\partial t} V_j \\ \frac{\partial v_j}{\partial t} V_j \end{pmatrix}, V_i \right\rangle +$$

$$\left\langle V_k(u_k, v_k) \begin{pmatrix} u_j \frac{\partial V_j}{\partial x} & u_j \frac{\partial V_j}{\partial y} \\ v_j \frac{\partial V_j}{\partial x} & v_j \frac{\partial V_j}{\partial y} \end{pmatrix}, V_i \right\rangle$$

$$+\left\langle \begin{pmatrix} \phi_k \frac{\partial V_i}{\partial x} \\ \phi_k \frac{\partial V_i}{\partial y} \end{pmatrix}, V_i \right\rangle + \left\langle \begin{pmatrix} -fv_k V_k \\ fu_k V_k \end{pmatrix}, V_i \right\rangle = 0$$

$$\Rightarrow \left\langle \frac{\partial u_j}{\partial t} V_j, V_i \right\rangle + \left\langle u_k V_k u_j \frac{\partial V_j}{\partial x}, V_i \right\rangle$$

$$+\left\langle v_k V_k u_j \frac{\partial V_j}{\partial y}, V_i \right\rangle + \left\langle \phi_k \frac{\partial V_i}{\partial x}, V_i \right\rangle \quad (21)$$

$$+\left\langle -fv_k V_k, V_i \right\rangle = 0$$

and

$$\Rightarrow \left\langle \frac{\partial v_j}{\partial t} V_j, V_i \right\rangle + \left\langle u_k V_k v_j \frac{\partial V_j}{\partial x}, V_i \right\rangle$$

$$+\left\langle v_k V_k v_j \frac{\partial V_j}{\partial y}, V_i \right\rangle + \left\langle \phi_k \frac{\partial V_i}{\partial y}, V_i \right\rangle \quad (22)$$

$$+\left\langle fu_k V_k, V_i \right\rangle = 0.$$

3.2 Time integration

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 FEM (see Navon 1979, 1987). The shallow-water equations system was then coupled at every time step so that the equations become quasi-linearized (see Wang et al. 1972; Douglas and Dupont 1970), since an average is taken at time level $n1$ and time level n of expressions, while the nonlinear advective terms are linearized by estimating them at time level $n+\frac{1}{2}$ using the following second-order approximation in time:

$$w^* = \frac{3}{2} w^n - \frac{1}{2} w^{n-1} + o(\Delta t^2) \quad (23)$$

where the state variables

$w = w(x, y, t) = (\vec{v}(x, y, t), \phi(x, y, t))$.

At each time step the shallow-water equations system was coupled, i.e. the solution of each equation after one iteration at a given time step was used to solve the other two equations for the same iteration for the same time step.

Upon introducing a finite difference discretization in time into the continuity equation (20), which is the first to be solved at a given time step, one obtains

$$\mathbf{M}\left(\phi_j^{n+1} - \phi_j^n\right) - \frac{\Delta t}{2} \mathbf{K}_1 \left(\phi_j^{n+1} + \phi_j^n\right) = 0 \quad (24)$$

where

$$\mathbf{M} = \iint_{ele} V_i V_j \, dA \quad (25)$$

and

$$\mathbf{K}_1 = \iint_{ele} V_j V_k u_k^* \frac{\partial V_i}{\partial x} dA + \iint_{ele} V_j V_k v_k^* \frac{\partial V_i}{\partial y} dA. \quad (26)$$

In this continuity equation, we need to use Crank-Nicholson to extrapolate u^* and v^* at the current time step so that we can proceed to solve ϕ^{n+1} at the next time step from $\left(u^*, v^*, \phi^n\right)$.

By introducing the same finite difference scheme into the u-momentum equations (21), one obtains:

$$\mathbf{M}\left(u_j^{n+1} - u_j^n\right) + \frac{\Delta t}{2} \mathbf{K}_2 \left(u_j^{n+1} + u_j^n\right)$$

$$+\frac{\Delta t}{2}\left(\mathbf{K}_{21}^{n+1} + \mathbf{K}_{21}^n\right) + \Delta t \mathbf{P}_2 = 0 \quad (27)$$

where

$$\mathbf{K}_2 = \iint_{ele} u_k^n V_i V_k \frac{\partial V_j}{\partial x} dA + \iint_{ele} v_k^* V_i V_k \frac{\partial V_j}{\partial y} dA, \quad (28)$$

$$\mathbf{K}_{21} = \iint_{ele} \phi_k^{n+1} V_i \frac{\partial V_k}{\partial x} dA, \quad (29)$$

$$\mathbf{P}_2 = -\iint_{ele} fv_k^* V_k V_i dA. \tag{30}$$

In this u-momentum equation, since we already know the most recent solution ϕ^{n+1} from solving the continuity equation above, we only need to extrapolate v^* at the current time step so that we can proceed to solve u^{n+1} at the next time step from $\left(u^n, v^*, \phi^{n+1}\right)$.

Finally, from the v-momentum equation (22), one obtains:

$$\mathbf{M}\left(v_j^{n+1} - v_j^n\right) + \frac{\Delta t}{2}\mathbf{K}_3\left(v_j^{n+1} + v_j^n\right) + \frac{\Delta t}{2}\left(\mathbf{K}_{31}^{n+1} + \mathbf{K}_{31}^n\right) + \Delta t\mathbf{P}_3 = 0, \tag{31}$$

where

$$\mathbf{K}_3 = \iint_{ele} u_k^{n+1} V_i V_k \frac{\partial V_j}{\partial x} dA + \iint_{ele} v_k^n V_i V_k \frac{\partial V_j}{\partial y} dA, \tag{32}$$

$$\mathbf{K}_{31} = \iint_{ele} \phi_k^{n+1} V_i \frac{\partial V_k}{\partial y} dA, \tag{33}$$

$$\mathbf{P}_3 = \iint_{ele} fu_k^{n+1} V_k V_i dA. \tag{34}$$

In this v-momentum equation, since we already know the most recent solution for both f^{n+1} and u^{n+1} at the current time step, we don't need any extrapolations at the current time step and we can proceed to solve v^{n+1} at the next time step from $\left(u^{n+1}, v^n, \phi^{n+1}\right)$.

3.3 Gauss-Seidel iterative method for the compact matrix of the Galerkin finite-element model

In this Galerkin finite-element model, a compact matrix form was adopted due to the local support property over the triangle mesh. In particular, the N x N global matrix, assembled from each small element matrix, has at most seven nonzero elements at each row of the matrix. Hence, we can store the global matrix into a compact matrix of size N x 7. (see Zhu, Navon and Zou 1994).

In order to implement boundary conditions in the Galerkin finite-element model, we have adopted the approach suggested by Payne and Irons (see Payne 1963) and mentioned by Huebner (see Huebner1975). 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} (chosen with a view to the significant number of digits possible with the given computer and the size of the field variables), 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 Navon 1979). variables have been treated (see Navon 1979).

4. Optimal Control of Galerkin Finite-Element Model

4.1 Brief descriptions of Discrete TLM and Adjoint Techniques

The S-W equations model can be written as:

$$\frac{\partial \mathbf{X}(t)}{\partial t} = \mathbf{F}(\mathbf{X}(t)) \tag{35}$$

and the discretized form of the numerical S-W equations model can be written as:

$$\mathbf{X}(t_r) = \mathbf{M}_{0\to r}\mathbf{X}_0 \tag{36}$$

where initial condition $\mathbf{X}_0$ is the control variable for the given numerical S-W equations model, $\mathbf{M}_{0\to r}$ is the predefined discretized nonlinear S-W equations model forecast operator, mapping the initial condition $\mathbf{X}_0$ into the model solution $\mathbf{X}_r$ at time t_r

In its general form, the *4D-Var data assimilation*, is defined as the minimization with respect to the initial condition $\mathbf{X}_0$ of the following discrete cost functional:

$$J(\mathbf{X}_0) = \frac{1}{2}\left(\mathbf{X} - \mathbf{X}_b\right)\mathbf{B}^{-1}\left(\mathbf{X} - \mathbf{X}_b\right) + \frac{1}{2}\sum_{r=0}^{n}\left(\mathbf{H}_r(\mathbf{X}_r) - \mathbf{Y}_r\right)^{\mathrm{T}}\mathbf{O}_r^{-1}\left(\mathbf{H}_r(\mathbf{X}_r) - \mathbf{Y}_r\right) \tag{37}$$

subject to the strong constraint, assuming

that the model is perfect, so that the sequence of model states $\mathbf{X}_r$ at time t_r must be a solution for the given model equations, where $\mathbf{B}$ is the background covariance matrix, $\mathbf{X}_r$ is the S-W equations model solution at time t_r, $\mathbf{O}_r$ is the observation error covariance matrix at time t_r, H_r is the observation operator at time t_r, representing projection of model variables into the observational variables. Since the $\mathbf{M}_{0\to r}(\mathbf{X}_0)$ is a nonlinear operator, the *4D-Var data assimilation method* becomes a nonlinear constrained optimization problem, with respected to the control variable $\mathbf{X}_0$ and it is very difficult to solve. Fortunately, it can be greatly simplified with two hypotheses.

The first hypothesis is the causality, in which the forecast model can be expressed as the product of intermediate forecast steps, so that the nonlinear S-W equations model forecast operator $\mathbf{M}_{0\to r}$ can be factorized into $\mathbf{M}_{0\to r} = M_r M_{r-1}\cdots M_1$, where each operator M_r denotes the discretized nonlinear forecast operator step from time r-1 to r and we have $\mathbf{X}_r = M_r \mathbf{X}_{r-1}$. Hence, by recurrence, we have $\mathbf{X}_r = M_r M_{r-1}\cdots M_1 \mathbf{X}_0$.

Another hypothesis is that, at each time step from both from r-1 to r, we obtain that the linearization of observation operator H_r can be written as $\mathbf{H}_r$, and that forecast operator M_r can also be linearized so that the the predefined discretized nonlinear S-W equations model forecast operator can be differentiated(perturbed) to obtain a so-called tangent linear model(TLM) :

$$\mathbf{X}'(\mathbf{t}_r) = \mathbf{M}_r \mathbf{X}_0' \tag{38}$$

where $\mathbf{M}_r$ represents the linearization of the discretized nonlinear S-W equations model forecast operator.

4.2 Adjoint of Galerkin Finite-Element model

Under those hypotheses, the quadratic cost functional above can be written as a summation as:

$$J = J^b + J^o = J^b + \sum_{r=0}^{n} \left(J^o\right)_r \tag{39}$$

where J^b and J^o are the background and observation terms respectively.

In order to obtain the optimal initial conditions of shallow water equations model that minimizes J above, the gradient of J needs to be calculated with respect to the control variable $\mathbf{X}_0$ as:

$$\nabla J = \nabla J^b + \nabla J^o \tag{40}$$

where the first term ∇J^b can be easily obtained as:

$$\nabla J^b = \mathbf{B}^{-1}\left(\mathbf{X} - \mathbf{X}_b\right) \tag{41}$$

and the second term ∇J^o requires the adjoint model integration which shall be briefly derived as follows:

On the one hand, consider the change in the cost functional J resulting from a small perturbation $\mathbf{X}_0'$ in the initial condition , which can be written as:

$$\begin{aligned}\left(J^o(\mathbf{X}_0)\right)' &= J^0\left(\mathbf{X}_0 + \mathbf{X}_0'\right) - J^0(\mathbf{X}_0) \\ &= \sum_{r=0}^{n} \mathbf{H}_r^{\mathrm{T}}\left(\mathbf{O}_r^{-1}\left(\mathbf{H}_r(\mathbf{X}_r) - \mathbf{y}_r\right)\right)^T \mathbf{X}_r'.\end{aligned} \tag{42}$$

On the other hand, to first order we can write the Taylor expansion of J as:

$$\left(J^o(\mathbf{X}_0')\right)' = \left(\nabla J^o(\mathbf{X}_0)\right)^T \mathbf{X}_0' + o\left(\left\|\mathbf{X}_0\right\|_2\right) \tag{43}$$

Furthermore, we are capable to find the gradient of the cost function by using the adjoint of the Tangent Linear Model of the given nonlinear time-dependent discrete Galerkin FEM model (see Navon et al. 1992).

By comparing (41) (42) (43) together, we obtain

$$\begin{aligned}\nabla J(\mathbf{X}_0) &= \mathbf{B}^{-1}\left(\mathbf{X} - \mathbf{X}_b\right) \\ &+ \sum_{r=0}^{n} \mathbf{M}_r^{\mathrm{T}} \mathbf{H}_r^{\mathrm{T}} \mathbf{O}_r^{-1}\left(\mathbf{H}_r(\mathbf{X}_r) - \mathbf{Y}_r\right)'\end{aligned} \tag{44}$$

where $\mathbf{M}_r^{\mathrm{T}}$ represents the adjoint of model at the r^{th} time step while the weighted differences $\mathbf{H}_r^{\mathrm{T}}\mathbf{O}_r^{-1}\left(\mathbf{H}_r(\mathbf{X}_r) - \mathbf{Y}_r\right)$ are forcing terms which can either be added to the adjoint variables whenever an

observational time is reached or can be initialized at the initial time stage if they are available at that time.

The basic techniques in coding the adjoint model above involves:

- Reset some temporary variables to zeros when using them in different statements;
- Saving and loading the state variables calculated in the forward model;
- Identifying the reused adjoint control variables in all the subroutines;
  - Reset the accumulations of reused adjoint variables to zeros when one period of accumulation is finished;
  - Finish the accumulations of reused adjoint variables only when calculating backwards into its first use;
- Handle the adjoint of iterative solver such as Gauss-Seidel;
- Handle the adjoint of boundary conditions;
- Identifying the inputs and outputs of each subroutine and the whole program;
- Make adjoint subroutines and parameters generic so that they can be reused for different adjoint variables without rewriting them over and over again.

4.2.1. Adjoint of iterative solver

The challenging part in the development of adjoint for nonlinear time-dependent discrete Galerkin Finite-Element model consists in the treatment of the Gauss-Seidel iterative procedure to solve the continuity equation systems and u-momentum equation systems as well as v-momentum linear systems, because some of the control variables to be solved at the current iteration level are reused while some are not (see Zhu, Navon and Zou 1994).

The key issues related to developing the adjoint of Gauss-Seidel iterative procedure are as follows:

We need to record the maximum number of the iterations when we integrate the nonlinear model forward in time, then, in order to obtain the adjoint of the Gauss-Seidel iterative procedure, the relationship of being reused among all the control variables must be analyzed. Finally, since the piecewise linear triangular Galerkin Finite-Element model has a local support of at most six nodes, while the minimum number of nodes is four when the node is on the boundary. Hence, the variable value at any given node inner or boundary is related to no more than six neighboring nodes surrounding it, and sometimes they are input variables and sometimes they are output variables. We are only concerned with the input variables when we speak about the reused variables, in other words, some of input variables in the iterative procedure are reused while other input variables are not, depending on the position in the grid as well as level of the iterations itself.

In addition, some control variables are firstly used in the setup of the *continuity system* and it will be used later twice in the setup of the *u-momentum system*. When dealing with situation to reuse adjoint variables in the adjoint code, we need to save the accumulated reused adjoint variables when calculating backwards into its first use. In other words, when we write the adjoint code, we need restore all the following accumulations into its first use when we finish the accumulation of reused adjoint variables.

4.3 Verification of correctness of the TLM and adjoint

The space increments used in this section are $\Delta x = \Delta y = 200km$, while in the section 5, we will adopt $\Delta x = \Delta y = 400km$ for convenience.

4.3.1. TLM test

Prior to checking the correctness of the adjoint model, we need to check the correctness of the discrete TLM (Figure 1). One idea is to consider a state vector **X** and a perturbation **X'** so that we can use Taylor expansion to verify the correlation between nonlinear Galerkin FEM and its corresponding TLM:

$$\psi(a) = \frac{\mathbf{G}(\mathbf{X} + \alpha\mathbf{X}') - \mathbf{G}(\mathbf{X})}{\alpha\mathbf{P}(\mathbf{X}')} = 1 + O(\alpha), \quad (45)$$

where **G** denotes the nonlinear Galerkin FEM and **P** represents its TLM operator, α defines the perturbation factor. Both the nonlinear Galerkin FEM and its TLM are integrated for a 5-hours period with various α values decreasing, and the results show that the correlation between Nonlinear Galerkin FEM model and its TLM is almost equal to one as α tends to zero.

Therefore, if the TLM test can be correct, we only need to code the adjoint model directly from the discrete TLM by rewriting the code of TLM statement by statement in the opposite direction. This simplifies not only the complexity of constructing the adjoint model but also avoids the inconsistency generally arising from the derivation of the adjoint equations in analytic form followed by the discrete approximation(due to non-commutativity of discretization and adjoint operators).

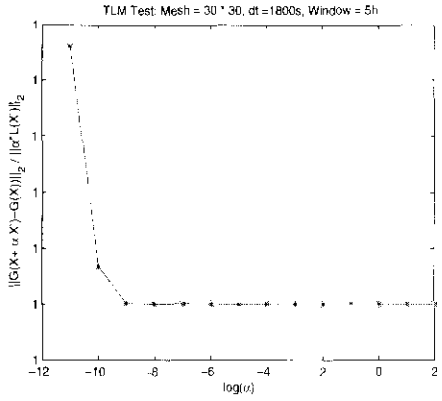


Figure 1. Correlation between Nonlinear Galerkin FEM model and its TLM, where α defines the perturbation factor

In addition, we also use an alternative idea to test the TLM (and thus the adjoint). It's called the *complex-step derivative approximation*. It is reasonably straightforward to implement, and it requires only slight modifications in the forward model code. The feature of this method is that it can avoid some cancellations in the finite difference calculation that will result in the loss of digit accuracy (see Martins 2003).

4.3.2. Transpose test

The correctness of the adjoint model checked by following the algebraic expression:

$$(\mathbf{PX})^{\mathrm{T}}(\mathbf{PX}) = \mathbf{X}^{\mathrm{T}}\left(\mathbf{P}^{\mathrm{T}}(\mathbf{PX})\right), \qquad (46)$$

where **X** represents the perturbation of input of the Galerkin FEM model, while the TLM denoted by **P** represents either a single *DO* loop or a *subroutine*. Each of them has its adjoint image *DO* loop or a *subroutine*, respectively. The left hand side involves only the tangent linear code, while the right hand side involves also the adjoint code. When we implement it, we first run the TLM code and use the output vector as the input vector of the adjoint calculation. There are some issues where we need to be careful, when running the test. First, we need to make sure all the state variables have been saved when we integrate TLM forward and restored or loaded when we integrate its adjoint backward. Second, we may need to run the different inputs to make sure we go through a rigorous check of the adjoint code into each single part of it. Finally, the results obtained illustrated that a 13 digits accuracy can be achieved in the input/output tests by using DOUBLE PRECISION.

4.3.3. Gradient test

We also tested the accuracy of the gradient of the cost function by using the so-called α test as follows (Figure 2):

$$F(\alpha) = \frac{J(\mathbf{X}+\alpha\mathbf{X}')-J(\mathbf{X})}{\alpha(\nabla J)^{T}(\nabla J)} = 1 + O(\alpha) \qquad (47)$$

and the results show that the vector we obtained from the adjoint model is almost equal to the gradient as α decreasingly tends to zero, if α is not too close to the machine accuracy(see Navon 1992).

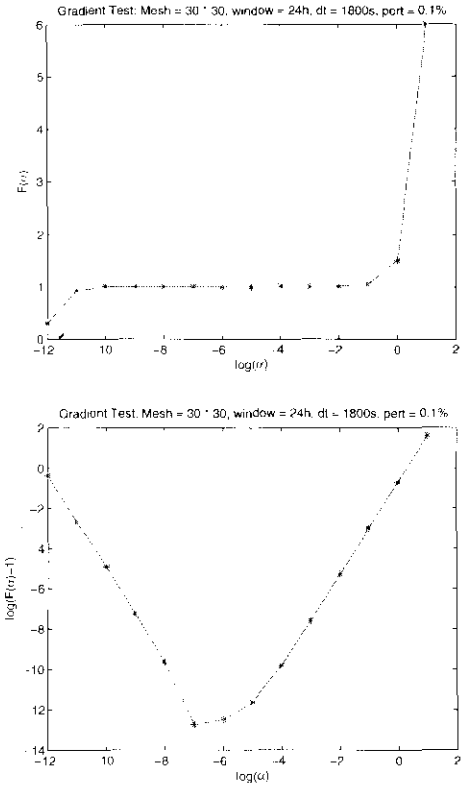


Figure 2. Gradient Test: (a) variation of $F(\alpha)$ with respect to log α and (b) variation of $\log(F(\alpha)-1)$ with respect to log α, where α defines the perturbation factor

5. Numerical Experiments

5.1 Description of problem

The test problem used here adopts the initial conditions (Figure 3) from the initial height field condition No.1 of Grammeltvedt (see Grammeltvedt 1969):

$$h(x,y) = H_0 + H_1 \tanh\left(\frac{9(D/2-y)}{2D}\right) + H_2\left(1/\cosh^2\left(\frac{9(D/2-y)}{D}\right)\right)\sin\left(\frac{2\pi x}{L}\right) \quad (48)$$

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

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

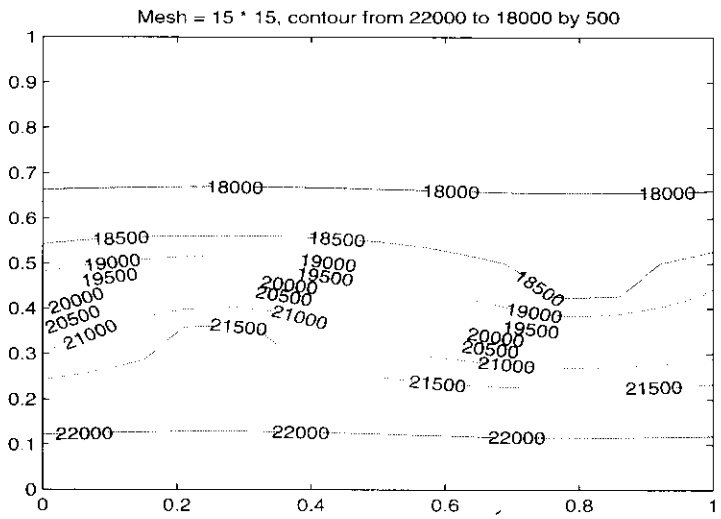
$$u = -\left(\frac{g}{f}\right)\frac{\partial h}{\partial y}, \quad v = \left(\frac{g}{f}\right)\frac{\partial h}{\partial x} \quad (49)$$

The dimensional constants used here are:

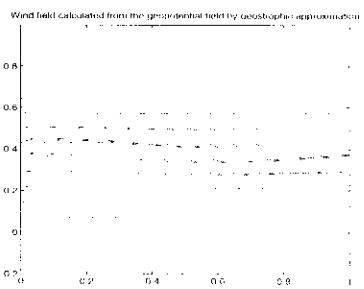
$$\begin{aligned} &L = 4400\,km; D = 6000\,km; \bar{f} = 10^{-4}\,s^{-1}; \\ &\beta = 1.5\times 10^{-11}\,s^{-1}m^{-1}; \\ &g = 10\,ms^{-1}; H_0 = 2000\,m; \\ &H_1 = 220\,m; H_2 = 133\,m. \end{aligned} \quad (50)$$

and the space increments used here are

$$\Delta x = \Delta y = 400km \quad (51)$$

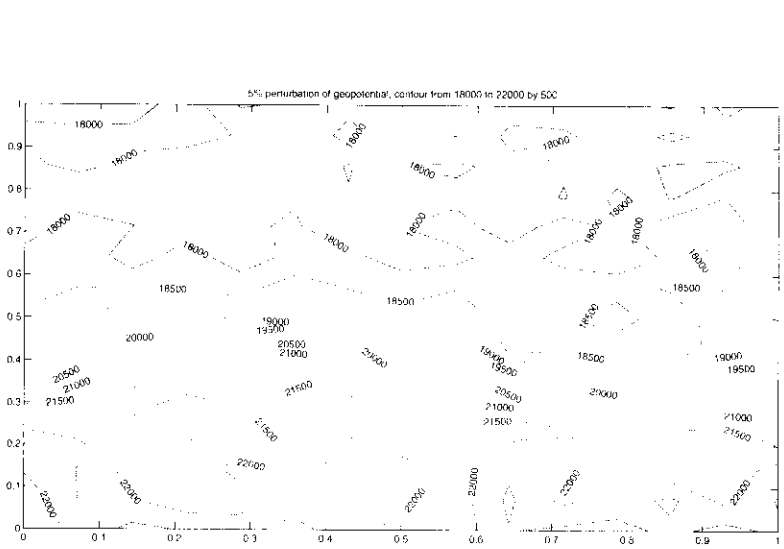


(a) initial geopotential

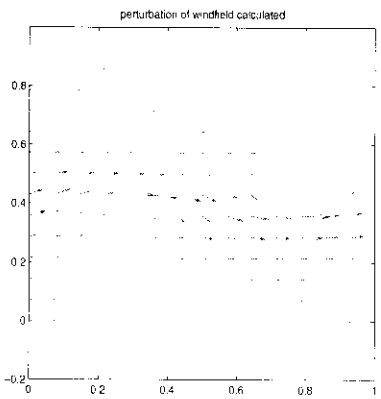


(b) initial wind-field

Figure 3. Initial condition:(a) Geopotential field for the Grammeltvedt initial condition. (b) Wind field calculated from the geopotential field by the geostrophic approximation.



(a) perturbation of initial geopotential



(b) perturbation of initial wind-field

Figure 4. 5% random perturbation of geopotential and as well as the wind fields of the shallow-water equations model

5.2. Perturbation of initial conditions

We applied a 5% uniform random perturbations (Figure 4) on the initial conditions in order to provide twin-experiment "observations" and we also computed the errors between the retrieved initial conditions related to the perturbed data and the reference state variables.

5.3 Retrieving the optimal initial conditions by applying

L-BFGS

The accuracy of a short-range numerical weather prediction greatly depends on the initial and boundary conditions. The following experiments illustrate the technology to retrieve the optimal initial condition from a noisy initial conditions. First, we randomly perturb the initial conditions to generate the so-called observations at each time step. Second, we generate another rondom perturbations of the initial conditions to obtain a initial guess of the initial conditions in the optimization. In this paper, we tried limited quasi-Newton method of Liu and Nocedal (1980,1989) and Richard and Nocedal (1995) to minimize the misfit between model solutions and artificial observations. The code is written in FORTRAN90 modularized with the control variables allocatable, so that any different mesh size can be tested in this code with a high accuracy. We also tested the different time steps as well as different data assimilation windows. The control variables are all defined as DOUBLE PRECISION so that a very high accuracy of approximation of the gradient of the cost functional with respect to the initial conditions can be achieved. In L-BFGS, we setup the number seven as the number of corrections (M=7) (See Liu and Nocedal 1989).

5.3.1. Testing different observations

The first experiment (Figure 5 and Figure 6) is performed on a short assimilation window for 12 hours with a small mesh size consisting of 15 x 15 grid points and we use a unconstrained minimization algorithm L-BFGS to minimize the cost functional. The adjoint model is integrated backward in time, with a forcing term being added, consisting of the difference between forecast and observation, interpolated at the same time and space location every time when an observation is encountered. We found out (Table 1) if we use 5% perturbation for both observations and initial guess, the L-BFGS converges in 31 iterations with 108 function evaluations to converge to prescribed tolerance $\varepsilon = 10^{-11}$ (Figure 7), but if we use 1% random perturbations, it will only take 28 iterations with 99 function evaluations to converge, which means both good observations and good initial guess will reduce the assimilation time required.

Table 1. L-BFGS: Data assimilation window = 12h, $\Delta x = \Delta y = 400km$, Δt = 1800s, and minimization convergence tolerance.

Random perturbations	Iterations	Function evaluations
5%	31	108
1%	28	99

Furthermore, if we extend the assimilation window from 12 hours to 48 hours, the L-BFGS minimization fails to achieve the prescribed tolerance no matter how accurate the observations and initial guess we choose for the optimization algorithms. If the mesh size is too coarse, say 5 x 5 grid points, even if we use 12 hours assimilation window, we will still fail to converge by using L-BFGS, which means either a too large assimilation window or a too small mesh size will affect the ability of the L-BFGS algorithm to converge to achieve the prescribed tolerance.

5.3.2. Testing different mesh resolutions

By increasing the mesh resolution from 15 x 15 to 30 x 30 (Figure 8) and still using L-BFGS, we found out that we can achieve a stricter tolerance $\varepsilon = 10^{-16}$, although it requires more iterations and function evaluations to converge (Table 2). Hence, it can be observed that the rate of the convergence of the cost functional associated with the coarse mesh is faster than the rate of convergence corresponding to the fine-resolution models, however, the value of the cost functional associated with the fine mesh can be reduced to achieve a higher level of tolerance that is by five orders of magnitude better than minimization of the cost functional achieved for the coarse mesh.

Table 2. Results of using L-BFGS: data assimilation window = 12h, $\Delta x = \Delta y = 200km$, mesh resolution= 30 x 30, Δt = 1800s, and minimization convergence tolerance $\varepsilon = 10^{-16}$

Random perturbations	Iterations	Function evaluations
5%	42	162
1%	38	149

Table 3. Results of using L-BFGS: data assimilation window = 12h, $\Delta x = \Delta y = 400km$, random perturbations = 5%, Δt = 900s, and tolerance of convergence of minimization is $\varepsilon = 10^{-15}$.

mesh size	Iterations	Function evaluations
15 x 15	28	97
30 x 30	35	140

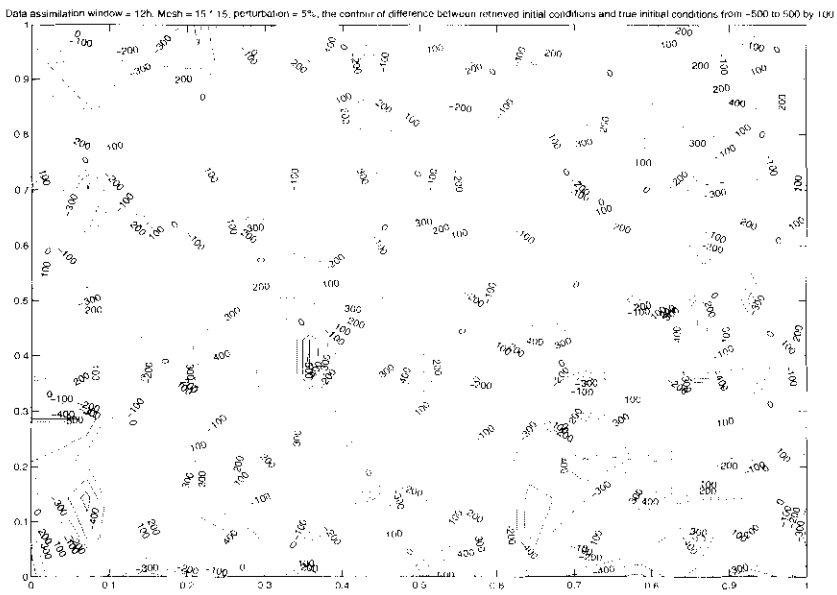
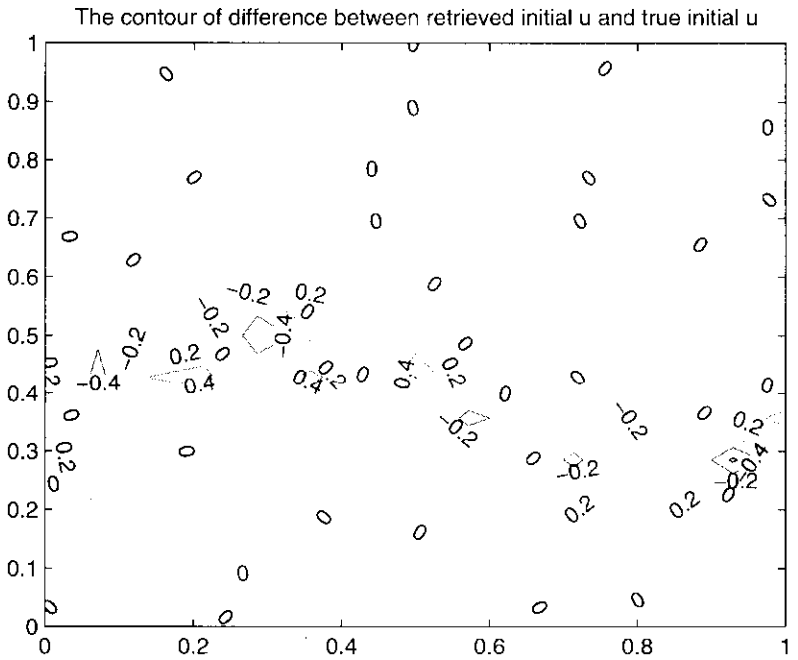


Figure 5. Data assimilation window = 12h, $\Delta x = \Delta y = 400km$, random perturbation = 5%. The contours of difference between retrieved initial geopotential and true initial geopotential are plotted.

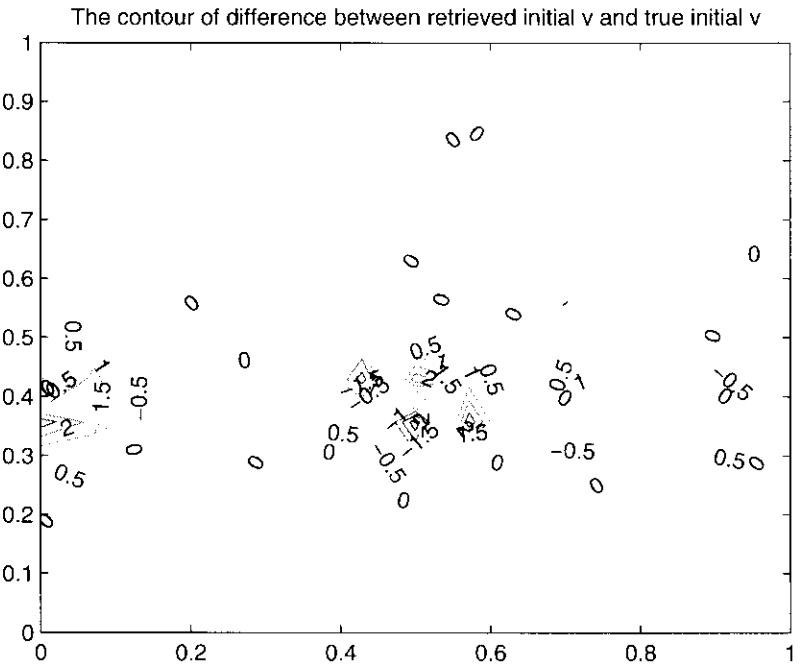
5.3.3. Testing different time steps

By decreasing the time length from 1800s to 900s while keeping an identical data assimilation window of 12 hours, which requires more time steps, we can achieve a convergence of minimization with tolerance $\varepsilon = 10^{-15}$ by using a coarse mesh size=15x15, which is beneficial especially when there are not enough observations of a fine mesh in space available everywhere but we could have the ability to measure them for every short time step length, we may still retrieve a very high accuracy of optimal initial conditions by shrinking each time step length and expanding number of data assimilation steps (Table 3).

This can also be explained by noting that the results from the fine mesh integrated contain more small-scale features than the corresponding ones from the coarse mesh integrated, and the dimension of the control variables also impacts upon the convergence rate so that the retrieval with fine-mesh model data becomes more difficult. The presence of small-scale features in an increase in the condition number of the Hessian of the cost function of the fine-mesh resolution model due to the introduction of small eigenvalues in the spectrum of the Hessian(see Axellson and Barker 1984). This situation becomes more apparent when the data assimilation is carried after a long time window of assimilation allowing reflections from limited boundaries thus causing short wave number noisy contaminations.

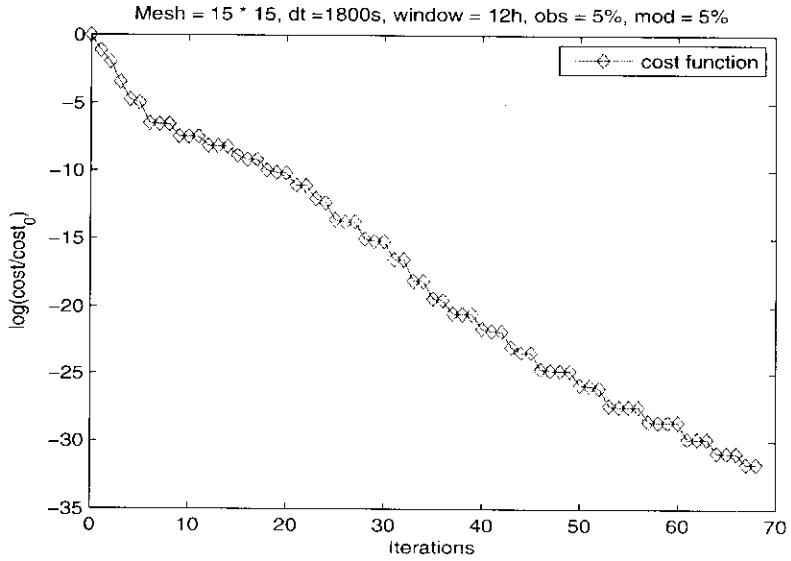


(a) The contour of difference between retrieved initial u-momentum and true Initial u-momentum

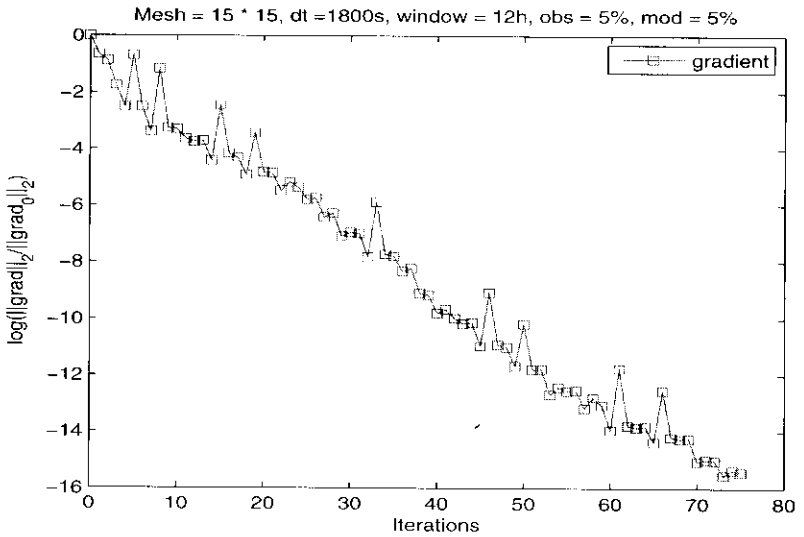


(b) The contour of difference between retrieved initial v-momentum and true initial v-momentum

Figure 6. Data assimilation window = 12h, $\Delta x = \Delta y = 400km$, random perturbation = 5%, (a) The contours of difference between retrieved initial u-momentum and true initial u-momentum from -0.5 to 0.5 by 0.2 are displayed. (b) The contours of difference between retrieved initial v-momentum and true initial v-momentum from -0.3 to 0.3 by 0.05 are also displayed.

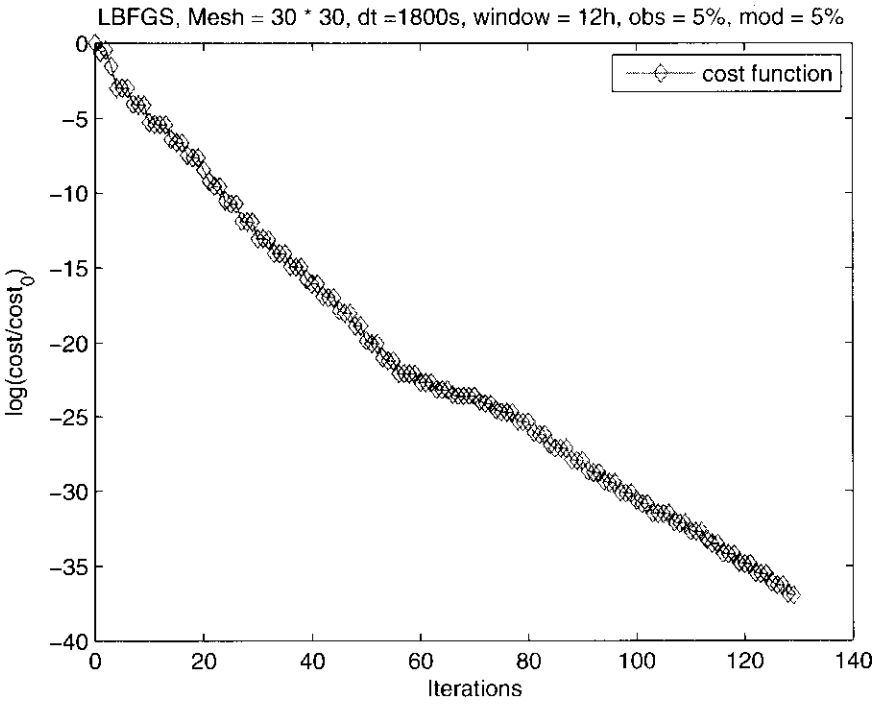


(a) Cost function

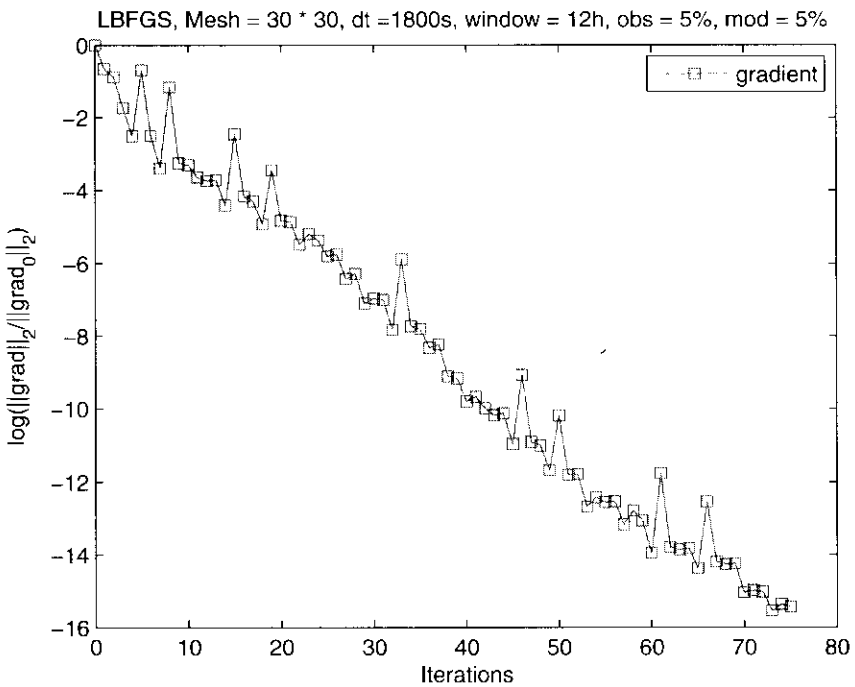


(b) L2 norm of gradient

Figure 7. L-BFGS minimization: Data assimilation window = 12h, $\Delta x = \Delta y = 400km$, mesh resolution= 15 x 15, random perturbation = 5%, (a) Normalized cost function scaled by initial cost function versus the number of minimization iterations (b) The norm of gradient scaled by initial norm of the gradient versus the number of minimization iterations.



(a) L2 norm of gradient



(b) The L2 norm of gradient

Figure 8. L-BFGS minimization: Data assimilation window = 12h, Mesh resolution = 30 x 30, random perturbations = 5%, (a) Normalized cost function scaled by initial cost function versus the number of minimization iterations.(b) The norm of gradient scaled by initial norm of the gradient versus the number of minimization iterations

6. Summary and Conclusions

In this paper, we developed a modularized code written in FORTRAN90 to present a VDA scheme using Galerkin FEM and its adjoint to generate minimization algorithms used to minimize cost functional so as to yield optimal initial conditions using model forecast with observations. The challenging part in this paper is how to handle the reused variables especially in constructing the adjoint of Gauss-Seidel iterative procedure for the Finite-Element Shallow-Water equations model over a limited area domain.

The large-scale unconstrained minimization limited-memory quasi-Newton method written by Liu and Nocedal (1989) was used to minimize the cost functional consisting of difference between model solutions and observations over the large assimilation window. We used the full random perturbation of the No. 1 of Grammeltvedt initial conditions (1969) to generate the observations and initial guess of the true initial conditions. We then carried the VDA numerical experiments using the adjoint model to assimilate the noisy observations.

The minimization of the cost functional was able to retrieve the true initial conditions when a coarse mesh size was employed. We also found out that the more accurate the observations as well as the initial guess of the initial conditions, the faster the rate of convergence of the minimization of the cost functional and the more accurate was the retrieval of the true initial conditions.

However, when carrying the L-BFGS to implement the VDA, it took a very long time to converge when applied to a very fine mesh and it failed to converge when a coarse mesh was employed. When we employed a coarse mesh in the model while using L-BFGS minimization and when observations were inserted frequently while shorter time steps were employed, we obtained similar accuracy results as in the case of fine mesh retrieval of the optimal initial conditions.

As we extended the length of the time window of the data assimilation of the forecast model, we impacted on the validity of the TLM model assumption and it became more and more difficult to employ the VDA scheme, since both effects of nonlinearity as well as limited area boundary conditions reflections impacted on the data assimilation procedure. To retrieve a high accuracy of optimal initial conditions, a fine mesh size is therefore required.

7. Appendix

7.1 Code organization

The nonlinear Galerkin FEM Model, TLM test (Figure 9), transpose test (Input/Output test), Gradient Test, and L-BFGS optimization were all written by a modularized FORTRAN90 language. In the graphs as follows, we only show the modularized Galerkin FEM code as well as the modularized L-BFGS optimization code flowchart.

In nonlinear Galerkin FEM model (Figure 10), four different modules are written as *Mesh*, *Assemble Matrix*, *Nonlinear Forward Model*, and *solver*. For example, in *Module Mesh*, we encapsulated a large amount of information such as the mesh size, the local and global element, compact local support, the area of each element, the coordinate and derivative of each node, and special geometries of the boundary structure.

In the graph of modularized L-BFGS optimization flowchart (Figure 11), we encapsulated the nonlinear Galerkin FEM model as well as its corresponding adjoint model. In the calls graph of L-BFGS implementation (Figure 12), we briefly list the function calls and subroutine calls to each other within each of the relevant modules.

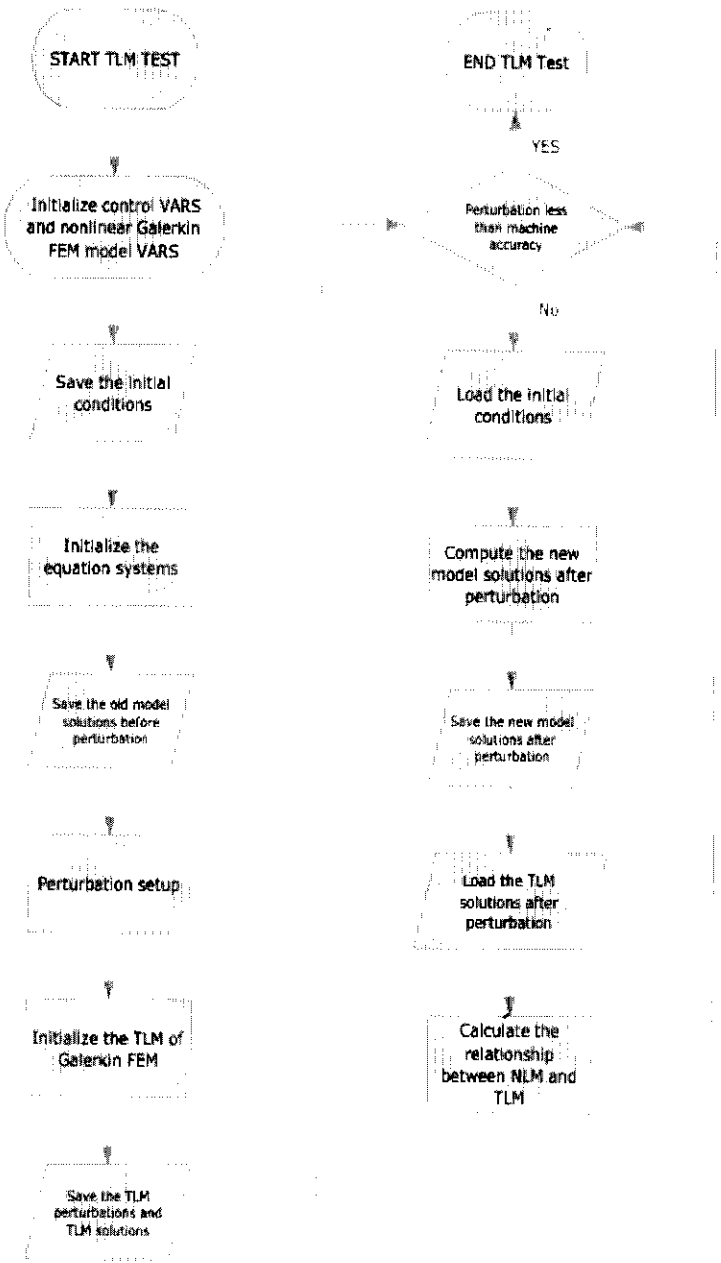


Figure 9. Flowchart of the Test of Tangent Linear Galerkin Finite-Element Model

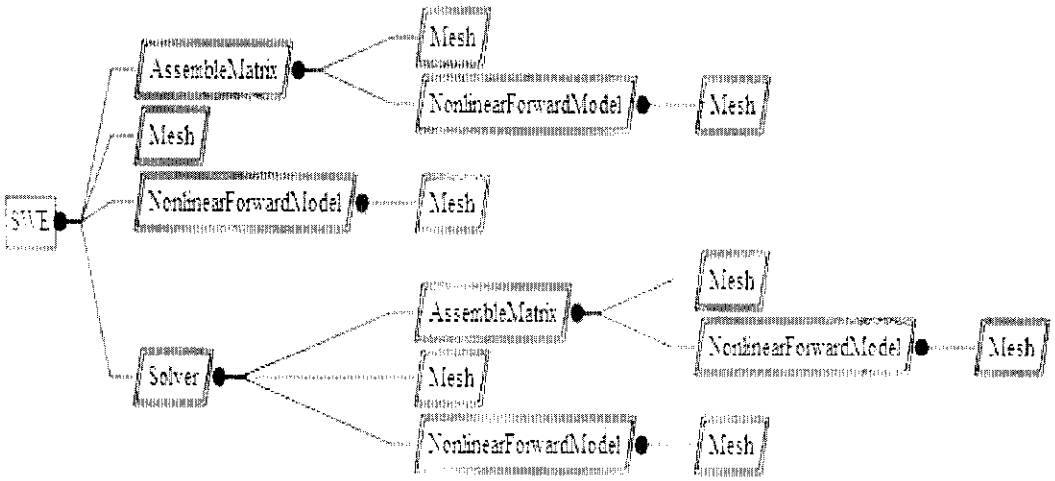


Figure 10. Modularized Galerkin FEM code organization

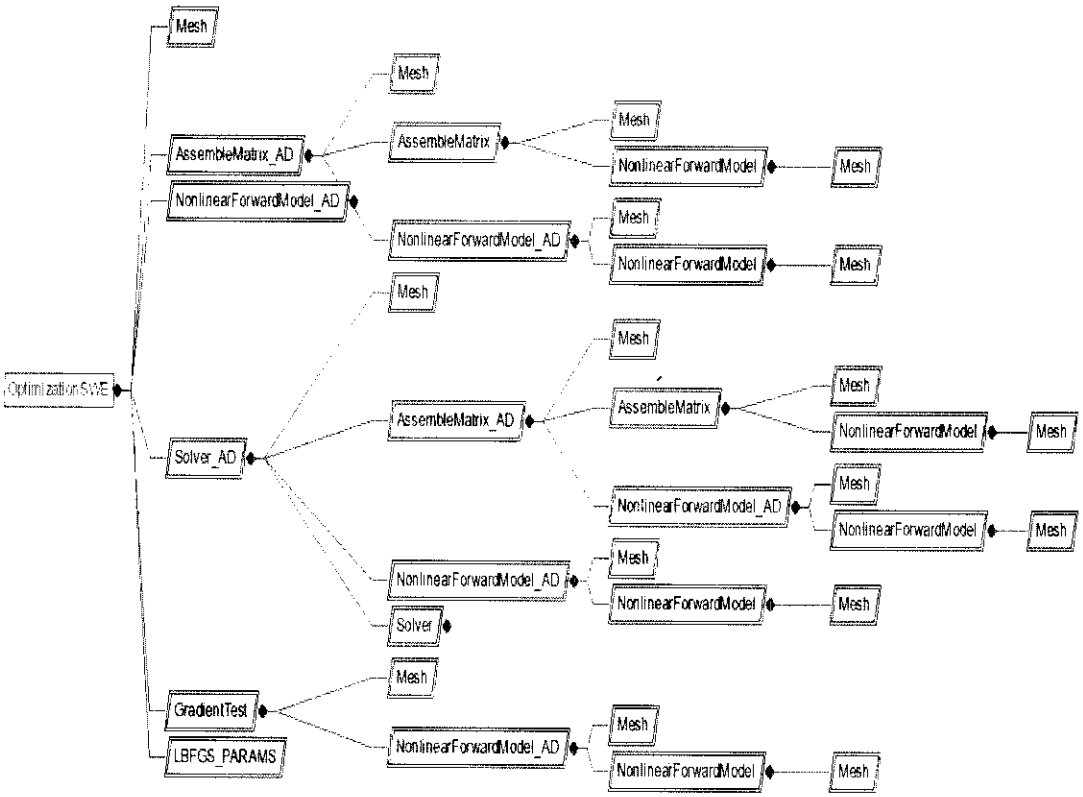


Figure 11. Modularized L-BFGS VDA code organization

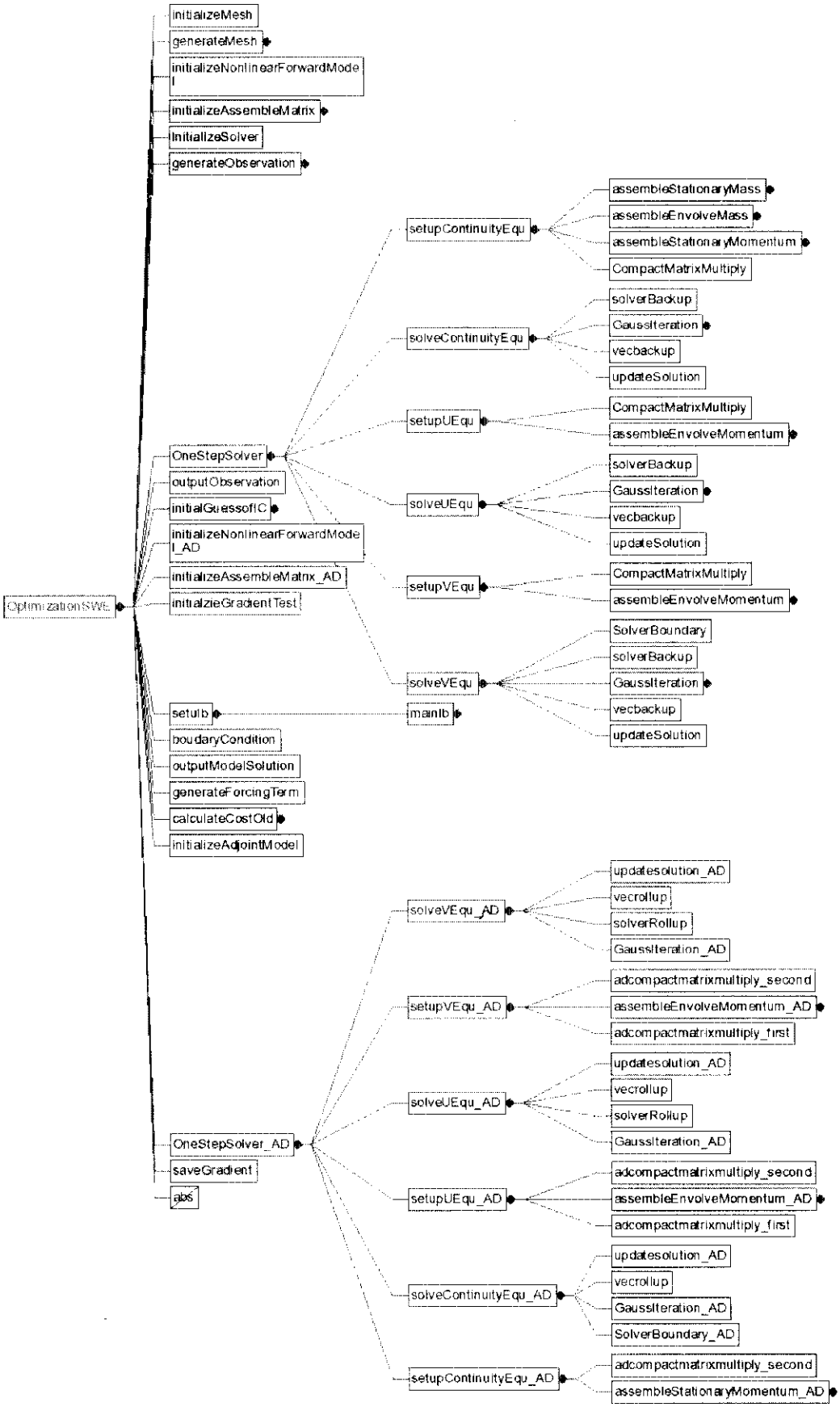


Figure 12. Calls graph of L-BFGS implementation

REFERENCES

1. AXELLSON, O., V. A. BARKER, **Finite element solution of boundary value problems: Theory and computation,** Computer Science and Applied Mathematics Series, Academic Press, 1984, p. 437.

2. BLUM, J., F. LE DIMET, I. MICHAEL NAVON, **Data Assimilation for Geophysical Fluids**, in press with Computational Methods for the Atmosphere and the Oceans, Volume 14: Special Volume (Handbook of Numerical Analysis). R. Temam and J. Tribbia, eds. Elsevier Science Ltd, New York (Philippe G. Ciarlet, Editor), 2008.

3. COURTIER, P., O. TALAGRAND, **Variational assimilation of meteorological observations with the the adjoint vorticity equation**, Part II, Numerical results, Quart. J. Roy. Meteor. Soc, 113, 1987, pp. 1329 - 1347.

4. DOUGLAS, J., T. DUPONT, **Galerkin method for parabolic problems**, S.I.A.M., J. Numer. Anal., 7, 1970, pp. 575 - 626.

5. GALEWSKY, J., R.K. SCOTT and L.M. POLVANI, **An initial-value problem for testing numerical models of the global shallow water equations**, Tellus, A 56 (5), 2004, pp. 429-440.

6. GRAMMELTVEDT, A., **A survey of finite-difference schemes for the primitive equations for a barotropic fluid**, Mon. Wea. Rev. 97, 1969, pp. 384-404.

7. HINSMAN, D.E., **Application of a finite-element method to the barotropic primitive equations**, M. Sc. Thesis, Naval Postgraduate School, Department of Meteorology, Monterrey, CA, 1975.

8. HUEBNER, K. H., **The finite-element method for engineers**, John Wiley & Sons, Chichester, 1975.

9. LIU, D. C., J. NOCEDAL, **On the limited memory BFGS method for large scale optimization, Mathematical Programming**, 45, 1989, pp. 503-528.

10. MARTINS, J. R. R. A., P. STURDZA, and J. J. ALONSO, **The complex-step derivative approximation**, ACM Transactions on Mathematical Software, Vol. 29 No. 3, 2003, pp. 245-262.

11. NAVON, I. M., **Finite-element simulation of the shallow-water equations model on a limited area domain**, Appl. Math. Model, 3, 1979, pp. 337-348.

12. NAVON, I. M., **FEUDX: a two-stage, high-accuracy, finite-element FORTRAN program for solving shallow-water equations**, Computers and Geosciences, 13, 1987, pp. 255-285.

13. NAVON, I. M., X. ZOU, J. DERBER and J. SELA, **Variational data assimilation with an adiabatic version of the NMC spectral mode**, Mon. Wea. Rev., Vol. 120, No.7, 1992, pp. 1435-143.

14. NOCEDAL, J., **Updating quasi-Newton matrices with limited storage**, Math. Comp., 24, 1980, pp. 773 - 782.

15. PAYNE, N. A., B. M. IRONS, **Private communication to O. Zienkiewicz**, 1963.

16. TAN, W. Y., **Shallow water hydrodynamics: mathematical theory and numerical solution for a two dimensional system of shallow water equations**, Beijing, China, Water & Power Press, Elsevier oceanography series, Elsevier Science Ltd, August 1992.

17. VREUGDENHIL, C.B., **Numerical methods for shallow-water flow**, Dordrecht, Boston, Kluwer Academic Publishers, 1994.

18. WANG, H. H., P. HALPERN, J. DOUGLAS, and I. DUPONT, **Numerical solutions of the one-dimensional primitive equations using Galerkin approximation with localized basic functions**, Mon. Wea. Rev., 100, 1972, pp. 738 -746.

19. ZHU, K., I. M. NAVON and X. ZOU, **Variational data assimilation with a variable resolution finite-element shallow-water equations model**, Mon. Wea. Rev., Vol. 122, No. 5, 1994, pp. 946-965.

20. ZHU, J., Z. R. L. TAYLOR, O. C. ZIENKIEWICZ, **The Finite Element Method**: Its Basis And Fundamentals, Butterworth - Heinemann, 2005, pp. 54-102.

21. ZIENKIEWICZ, O.C., R.L. TAYLOR, J. Z. ZHU, P. NITHIARASU, **The Finite Element Method**, Hardcover, Butterworth- Heinemann, 2005.

22. ZOU, X., I. M. NAVON, and J. SELA: **Control of gravitational Oscillations in Variational Data Assimilation**, Mon. Wea. Rev., 121, 1993, pp. 272 - 289,1993