Variational Data Assimilation with a Semi-Lagrangian Semi-implicit Global Shallow-Water Equation Model and Its Adjoint

Y. LI*

Supercomputer Computations Research Institute, The Florida State University, Tallahassee, Florida

I. M. NAVON

Department of Mathematics and Supercomputer Computations Research Institute, The Florida State University, Tallahassee, Florida

P. COURTIER

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire, United Kingdom

P. GAUTHIER

Aerospace Meteorology Division, Atmospheric Environment Service, Dorval, Quebec, Canada

(Manuscript received 16 September 1992, in final form 12 December 1992)

ABSTRACT

An adjoint model is developed for variational data assimilation using the 2D semi-Lagrangian semi-implicit (SLSI) shallow-water equation global model of Bates et al. with special attention being paid to the linearization of the interpolation routines. It is demonstrated that with larger time steps the limit of the validity of the tangent linear model will be curtailed due to the interpolations, especially in regions where sharp gradients in the interpolated variables coupled with strong advective wind occur, a synoptic situation common in the high latitudes. This effect is particularly evident near the pole in the Northern Hemisphere during the winter season. Variational data assimilation experiments of "identical twin" type with observations available only at the end of the assimilation period perform well with this adjoint model. It is confirmed that the computational efficiency of the semi-Lagrangian scheme is preserved during the minimization process, related to the variational data assimilation procedure.

1. Introduction

The variational data assimilation method using the adjoint technique was implemented by Courtier (1985), Derber (1985), Lewis and Derber (1985), Le Dimet and Talagrand (1986), Thacker and Long (1988), and Courtier and Talagrand (1990) on simple models such as the shallow-water equation models. Recently this method has been extended and applied to various three-dimensional operational multilevel primitive equation models such as at National Meteorological Center (NMC), European Centre for Medium-Range Weather Forecasts (ECMWF), as well as to the National Aeronautics and Space Administration (NASA) Goddard Laboratory for Atmospheres (GLA) fourth-order A-grid finite-difference model by Navon et al. (1990, 1992), Thépaut and Courtier (1991), and Chao and Chang (1992), respectively.

Due to the iterative nature of the large-scale unconstrained minimization process required by this method, the computational efficiency of integrating the models turns out to be a crucial factor when operational application is concerned. One way to improve the computational efficiency is to seek a temporal discretization scheme that allows the use of large time steps without introducing computational instability. After the early work of Krishnamurti (1962), a vast amount of research efforts (Bates and McDonald 1982; McDonald and Bates 1987, 1989; Ritchie 1988; Côté and Staniforth 1988; Staniforth and Côté 1991; Bates et al. 1992) has been directed toward employing the Lagrangian approach for treating the advection terms while retaining a regular discretization in space, this approach being known as the semi-Lagrangian scheme. By doing so, one may obtain practically unconditional stability (Robert 1981, 1982) for the time integration. It is, therefore, natural to attempt to perform variational data assimilation using semi-Lagrangian models.

However, many problems need to be tackled in this line of research. As an initial effort in this direction, we have derived the adjoint model for the semi-La-

* Current affiliation: General Science Corporation, Seabrook, Maryland.

Corresponding author address: Dr. I. M. Navon, Department of Mathematics and Supercomputer Computations Research Institute, The Florida State University, Tallahassee, FL 32306-4052.

grangian semi-implicit (SLSI) two-time-level finite-difference shallow-water equation model of Bates et al. (1990) (referred to as the BSHB model hereafter) using a direct solver of Moorthi and Higgins (1993). In section 2, we discuss the model development and the formulation of the data assimilation problem after a brief description of the forward model. Section 3 is dedicated to the examination of the validity of the tangent linear model, including a theoretical analysis and some numerical results. We present results of variational assimilation experiments in section 4 and discuss various related issues. Finally in section 5 conclusions are drawn.

2. Model developments and assimilation formulation

The BSHB model consists of the shallow-water momentum and continuity equations, which are written as

$$\left(\frac{d\mathbf{V}}{dt}\right)_H = -\nabla(\Phi - cD) - f\mathbf{k} \times \mathbf{V} \qquad (2.1)$$

$$\frac{d\Phi}{dt} = -\bar{\Phi}D - \Phi' D, \qquad (2.2)$$

where $\mathbf{V}$ is the horizontal velocity vector ($=u\mathbf{i} + v\mathbf{j}$); $\mathbf{k}$ is the unit vector in the vertical direction; ∇ is the horizontal gradient operator; Φ is the geopotential; f is the Coriolis parameter; c is the coefficient of divergence damping; D is the divergence; $(d/dt)_H$ is the horizontal component of the Lagrangian derivative; $\bar{\Phi}$ is a constant mean geopotential; and Φ' is the perturbation of geopotential from the mean. The readers are referred to Bates et al. (1990) for details.

The variational data assimilation approach using the adjoint technique aims to minimize a cost function J consisting of a weighted lack of fit between model and observations, which can be written as

$$J[\mathbf{X}(t_0)] = \frac{1}{2}\sum_{r=0}^{R} [\mathbf{H}\mathbf{X}(t_r) - \mathbf{Z}(t_r)]^{\mathrm{T}}\mathbf{W}(t_r) \times [\mathbf{H}\mathbf{X}(t_r) - \mathbf{Z}(t_r)], \qquad (2.3)$$

where $\mathbf{X}(t_r)$ is a vector of dimension N containing all model variables, which consist of two horizontal velocity components and the geopotential, over all grid points at time t_r. Here R is the number of time levels for the analyzed fields in the assimilation window; $\mathbf{Z}(t_r)$ represents observational data used for the assimilation purposes; $\mathbf{H}$ is a transformation matrix that maps the model variables to the observations. In this study, the observational data are model outputs defined at model grid points, so $\mathbf{H}$ is an identity matrix. Here $\mathbf{W}(t_r)$ is an $N \times N$ diagonal matrix of weighting coefficients. The values for its elements are usually determined by a dimensional scaling of the various variables, relative importance, and quality of the dataset and other considerations. In its most general form, $\mathbf{W}(t_r)$ can be taken as the inverse covariance matrix of the observation errors. There are other choices for the inner product used in the cost function, such as an energy norm inner product used by Courtier and Talagrand (1990) and Thépaut and Courtier (1991).

The gradient of the cost function with respect to the control variables (in our case the initial condition) is obtained by integrating the adjoint model backward in time. The derivation of the adjoint model consists of first linearizing the model about some basic state and then finding the adjoint operators of the tangent linear model, as described in Navon et al. (1992) and Li et al. (1991). The correctness of both the tangent linear model and its adjoint needs to be checked. The tangent linear model check is performed by comparing the difference between two model integrations started from two slightly different initial states with the perturbations predicted by the tangent linear model. These should be very close to each other, with the differences decreasing with the size of the initial perturbations (Thépaut and Courtier 1991). An observed rule in the check of correctness of adjoint operators is that the acceptable inaccuracy be limited to machine round-off errors. Finally, the correctness of the gradient obtained with the adjoint model needs to be checked before it is used in the minimization procedure. Such a gradient check is formulated using a Taylor expansion of the cost function

$$\Phi(\alpha) = \frac{J(\mathbf{X} + \alpha\mathbf{h}) - J(\mathbf{X})}{\alpha\mathbf{h}^{\mathrm{T}}\nabla J(\mathbf{X})} = 1 + O(\alpha). \qquad (2.4)$$

Here α is a small scalar and $\mathbf{h}$ is a vector of unit length in an arbitrary direction. Usually $\mathbf{h}$ is taken to be ($\mathbf{h} = \nabla J \|\nabla J\|^{-1}$), a vector in the gradient direction, so that the variation of the variables yields a consistent scaling and the check is not shifted to certain model variables due to computer precision. For values of α that are small but not too close to the machine zero, one should obtain a value of $\Phi(\alpha)$, which linearly approaches 1 as α decreases (Navon et al. 1992). Due to the characteristics of the Taylor expansion, the residual of the above formula should linearly approach 0 (Thépaut and Courtier 1991).

We used a limited memory quasi-Newton algorithm for the unconstrained minimization described in Nocedal (1980) and Liu and Nocedal (1989), which was also used in Navon et al. (1992), hereafter called the L-BFGS method. It was concluded in Zou et al. (1993) that the L-BFGS algorithm is one of the best candidates for robust large-scale unconstrained minimization problems typically arising in meteorological models (see also Navon and Legler 1987). Due to the different physical dimensions, the values for the geopotential and velocity fields in the BSHB model span a range of six orders of magnitude. The components of the gra-

dient of the cost function with respect to the geopotential field could be three orders of magnitude smaller than those with respect to the horizontal velocity field. A rough scaling (Navon and de Villiers 1983; Courtier and Talagrand 1990) is applied to precondition the gradient and the control variables. The scaling factors in this study are 1.0 for the velocity field components and 3.0×10^2 for the geopotential field.

It is important to point out that interpolation routines are used for the estimation of model variables at departure points and midtrajectory points in the SLSI model. In order to determine which grid points should be used as the surrounding points for the interpolation, one needs to take the integers of the variables defining a nongridpoint location, and these integers then become the subscripts of the arrays storing the model variables at the surrounding grid points. The location of such a particular nongridpoint location is a function of the velocity at grid points. Such operations appear at first to be "discontinuous" and thus nondifferentiable. A more in-depth analysis actually shows that the dependency relation between the model variables in such operations is actually continuous despite the fact that discontinuous algebraic steps are introduced. In other words, the nongridpoint location is a continuous function of the velocity field and the interpolated model variables at the nongridpoint location continuously depend on its location as well as on the values for model variables at surrounding grid points. Such operations should be differentiable. Therefore, it is justifiable to perform the linearization and the adjoint derivation. This point will be presented in more detail in the next section when we discuss the validity of the tangent linear model. As a simple illustration, we can consider a linear interpolation problem in one dimension. The description of the problem, the interpolation program, the linearization program, and the corresponding adjoint code are presented in the Appendix.

3. Validity of the tangent linear model

For time intervals greater than the limit of validity of the tangent linear model, the validity of the use of the adjoint model technique in variational data assimilation can be questioned (Lacarra and Talagrand 1988; Rabier and Courtier 1992; Errico and Vukicevic 1992). Therefore, it is very important to examine the extent of the validity of the tangent linear model and also to demonstrate how interpolation routines employed by the semi-Lagrangian scheme affect the validity property.

a. Theoretical aspects

Consider the following 1D problem,

$$\frac{dU}{dt} = 0, \tag{3.1a}$$

$$\frac{dX}{dt} = U, \tag{3.1b}$$

which we wish to solve with the semi-Lagrangian method. If X_j stands for a grid point, (3.1a) and (3.1b) imply that

$$U(X_j, t + \Delta t) = U(X_j(t), t), \tag{3.2}$$

with $X_j(t) = X_j - \int_t^{t+\Delta t} U[X(t'), t]\,dt'$. Approximating the integral in a crude manner, one may write

$$X_j(t) = X_j - U(X_j, t)\Delta t. \tag{3.3}$$

This expression can be made more realistic in order to correspond to the discretized formulation actually employed in the semi-Lagrangian method. Using (3.3), (3.2) is seen to assume the form

$$U(X_j, t + \Delta t) = B_{ji}[U(X_j, t)]U(X_i, t). \tag{3.4}$$

The subscript j denotes an arrival point and i a departure point. If $X_{i_1} < X_j(t) < X_{i_1+1}$, the elements of the interpolation matrix $\mathbf{B}$, in a linear interpolation case, are such that

$$B_{ji} = \begin{cases} 0, & \text{if} \quad i \neq i_1 \quad \text{and} \quad i \neq i + 1 \\ B_{ji_1} = 1 - \alpha, \\ B_{ji_1+1} = \alpha, \end{cases}$$

with $\alpha = [X_j(t) - X_{i_1}]/(X_{i_1+1} - X_{i_1})$. This is similar to the example presented in the Appendix. Even though cubic interpolations are used in the model along with linear ones, the discussion here is based on this simple upstream linear interpolation formulation without loss of generality.

Equation (3.3) is nonlinear and the tangent linear model is obtained from it by linearization. Considering an infinitesimal perturbation δU of the control variable U at time t, (3.4) implies that, to first order,

$$\delta U(X_j, t + \Delta t) = B_{ji}[U_s(X_j, t)]\delta U(X_i, t) + B_{ji}[\delta U(X_j, t)]U_s(X_i, t), \tag{3.5}$$

where the subscript s refers to the trajectory obtained by integrating the direct model (3.4). The first term of (3.5) is the same interpolation matrix defined in the forward model integration; therefore, it can be considered known. In practice, however, one recovers the interval from the stored value of U_s.

The second term in (3.5) is associated with variations in the interpolated value of U_s due to an infinitesimal change in U. If $\xi_s = X_j - U_s(X_j, t)\Delta t$ is the location used in the direct integration, this variation is

$$\delta\xi_s = -\delta U(X_j, t)\Delta t. \tag{3.6}$$

This change being infinitesimal implies that $\xi_s + \delta\xi_s$ will remain in the same interval as ξ_s unless $\alpha = 0$, an event that can occur with zero probability. The second term in (3.5) is also associated with the U_s field, the advective velocity itself. Therefore, it is in regions with sharp gradients in the interpolated fields coupled with strong advective winds that the tangent linear model approximation is likely to break down due to the interpolation effect.

It follows that $B_{ji}[\delta U(X_j, t)]$ will be such that

$$B_{ji} = \begin{cases} 0, & \text{if } i \neq i_1 \text{ and } i \neq i_1 + 1 \\ -\delta\alpha, & \\ \delta\alpha, & \end{cases}$$

with

$$\delta\alpha = -\frac{\delta U(X_j, t)\Delta t}{X_{i_1+1} - X_{i_1}}.$$

The most significant point about (3.6) is that since the size of the $\delta\xi_s$ is related to the time step Δt, one can expect this limit to vary with the magnitude of the time step. Moreover, for a finite-difference model, the zonal grid size becomes smaller toward higher latitudes. It is then expected that it is more probable to have $\xi_s + \delta\xi_s$ lie outside of the interval where ξ_s is located, that is, when approaching the polar regions.

As is well known, higher-order interpolations involve a larger number of grid points as well as larger intervals, making it more likely for $\xi_s + \delta\xi_s$ to remain in the same interval as ξ_s, compared with the linear interpolations. We expect that use of higher-order interpolations may extend the limit of the validity of the tangent linear model.

b. *Numerical results*

Following Rabier and Courtier (1992), let us write the atmospheric state vector at the initial time t_0 as

$$x(t_0) = x_0(t_0) + \delta x(t_0),$$

where $x_0(t_0)$ stands for the basic state and $\delta x(t_0)$ for the perturbation. The initial basic state, of which the geopotential field is shown in Fig. 6a, is chosen as a model state predicted from an ECMWF analysis. More details concerning how this initial field is constructed can be found in section 4a. The perturbations are taken to be random fields over all the model's grid points, with the magnitude of the two velocity components ranging from -3.0 to 3.0 m s^{-1}, and the geopotential from -300.0 to 300.0 m^2 s^{-2}. The magnitude of such initial perturbations is far from being negligible. Another property of these initial perturbations is that they focus on the smallest resolvable scale by the model grid resolution. Since the zonal grid size becomes smaller toward the polar regions, we have sharper gradients or larger variations of the interpolated variables in those regions.

The evolution of x is given by the integration of the model M between times t_0 and t_n as

$$x(t_n) = M(t_n, t_0)[x(t_0)] = M(t_n, t_0)[x_0(t_0) + \delta x(t_0)],$$

whereas the first-order evolution of the perturbation $\delta x(t_0)$ is the result of the integration of the tangent linear model R obtained by linearization of the nonlinear model M in the vicinity of the trajectory whose initial condition is the basic state $x_0(t_0)$:

$$\delta x(t_n) = R(t_n, t_0)\delta x(t_0).$$

In order to check the validity of the tangent linear model, we compare the total perturbation

$$N[\delta x(t_0)] = M(t_n, t_0)[x_0(t_0) + \delta x(t_0)] - M(t_n, t_0)[x_0(t_0)]$$

to its linear component

$$L[\delta x(t_0)] = R(t_n, t_0)\delta x(t_0).$$

The difference between the two is denoted as

$$D[\delta x(t_0)] = N[\delta x(t_0)] - L[\delta x(t_0)].$$

In order to quantify this comparison, we defined the square of the norm to be

$$\|\mathbf{x}\|^2 = \mathbf{x}^{\mathrm{T}} W \mathbf{x}$$

in accordance with what is used in the cost function for the data assimilation problem. The relative error of the tangent linear model is then defined as the ratio $\|D\|/\|L\|$.

We consider here the sensitivity of this relative error to the size of the time step used for the integration. Figure 1 presents the evolution of the relative errors for different time steps as the model is integrated. The tangent linear model remains quite accurate with time steps equal to 10 and 30 min; the value of the error reaches a level of only about 10% after 96 h. These results confirm the correctness of the coding of the tangent linear model. It is interesting to note that after 48 h, the error with a time step of 30 min becomes slightly smaller than that with a time step of 10 min. Such a marginal difference can be attributed to round-off errors that accumulate faster due to the higher number of operations involved in the integration with a smaller time step. As the time step increases to 1, 1.5, and

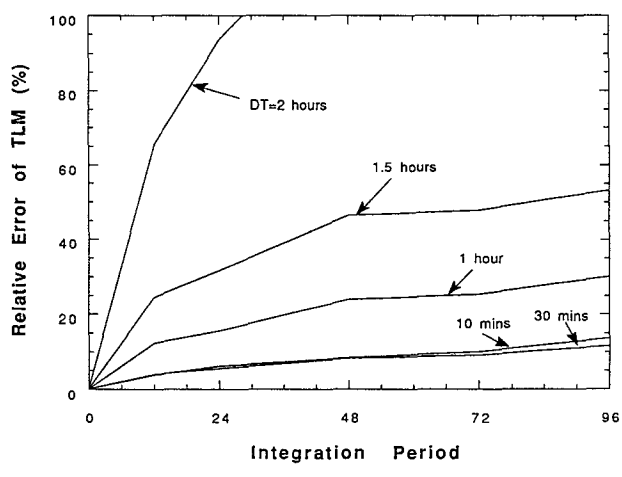

FIG. 1. Evolution of relative error of the tangent linear model for model integrations with various time-step sites.

2 h, the relative error increases monotonically and somewhat dramatically. With a time step of 2 h, the relative error becomes as large as 100% after 24 h of integration. The relative error, with a time step of 1.5 h, attains more than 40% after 48 h. The theoretical expectation as discussed previously is thus confirmed; that is, when the time step for integration is too large, the limit of validity of the tangent linear model is shortened.

As mentioned before, the random initial perturbations we used created sharp variations near the polar regions. We would expect that the error of the tangent linear model compared with the full model arises mainly due to differences occurring at high latitudes. Figure 2 presents the geopotential part of the difference field D after 48 h for the case with a time step equal to 1.5 h. It can be clearly seen that the errors are located mainly near the poles. However, the error distribution displays an asymmetric pattern, with the errors located near the North Pole being much larger than those near the South Pole. Since the initial perturbations are randomly and uniformly generated and the model grids are symmetric about the equator, one should expect other reasons for this asymmetric behavior. Figure 3 shows the zonal mean of the zonal wind of the initial basic state. The magnitude of the mean zonal wind near the southern pole is much smaller than that near the northern pole, which is not surprising since this dataset was collected during the Northern Hemisphere winter. As illustrated before, the errors are amplified by the advective velocity. This provides an explanation for the presence of larger errors near the northern pole.

So far, we have demonstrated the variability of the validity of the tangent linear model with various sizes of the time step used in the SLSI model. This suggests that when performing variational data assimilation with semi-Lagrangian models, the time step used for the model integration should also be constrained by the limit of validity of the tangent linear model due to interpolations as well as due to the particular grid geometry employed in the model. Even though it is not possible at this point to provide a universally and quantitatively optimal time step for variational data assimilation with the SLSI model, the results presented here do shed some light on the way the choice of time step should be carried out. This result is also likely to be valid in a baroclinic model, if not more important, since baroclinic instability induces sharp gradients in frontal structures.

4. Data assimilation experiments

a. General description

Numerical experiments of variational data assimilation with the SLSI shallow-water equation model on the sphere were carried out using a resolution of $(\Delta\theta, \Delta\lambda) = (7.5°, 7.5°)$, respectively. The model was initialized with the 500-mb geopotential and horizontal velocity fields of the ECMWF analysis fields of 0000 UTC 15 January 1979. A 12-h forecast was made and the model output at the end of this forecast period was used as the initial condition for a model integration used to generate "observations," of which the geopotential field is shown in Fig. 6a. The end of this 12-h period constitutes the initial time for the data assimilation experiments, and the reason for having such a 12-h preparatory integration is to ensure a more consistent initial condition to avoid the spinup effect. A 3-h integration is made from this initial condition, and the predicted fields are used as the first guess of the initial condition for all the following assimilation experiments. Such a first guess bears a typical error on the order of magnitude of difference between two consecutive atmospheric states separated by 3 h. The error field for the geopotential, for illustration purposes, is presented in Fig. 6b. For all numerical experiments, unless indicated otherwise, the divergence damping coefficient c is chosen to be 5.0×10^6 m^2 s^{-1}.

The cost function has the form described by (2.3). The weights are selected to be 10^{-3} for the geopotential

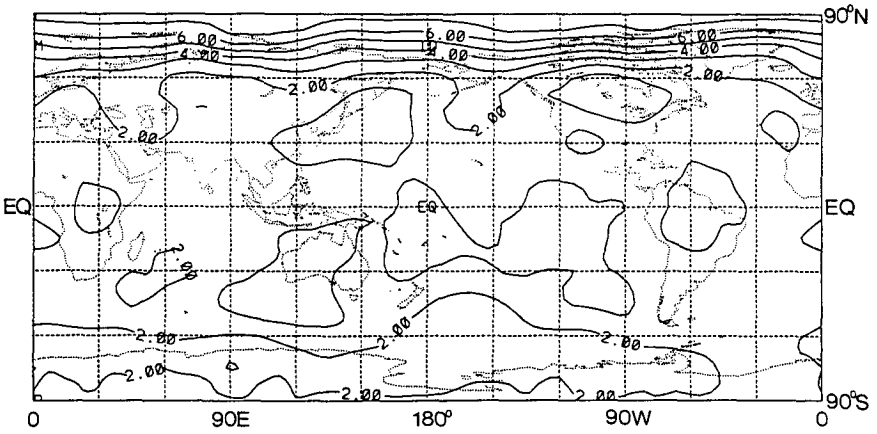

FIG. 2. The geopotential of the difference field between the perturbations calculated from the full model and the perturbations calculated from the tangent linear model (contour interval is 2.0 m^2 s^{-2}).

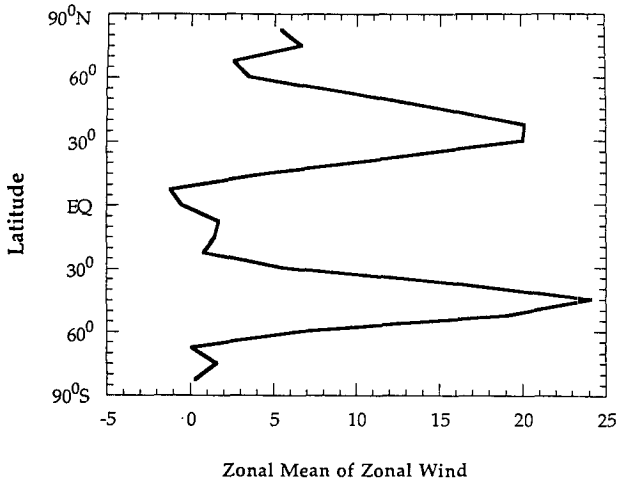

FIG. 3. Zonal average of the zonal velocity (m s^{-1}).

and 1.0 for the two horizontal velocity components, respectively. All the experiments performed in this study are of "identical twin" type; that is, the "observations" are generated by the same model.

b. *Test of the model*

A gradient check as shown in Fig. 4a indicates good behavior. We have also checked the residual of $\Phi(\alpha)$ in (2.4), and the criterion that the residual approaches zero linearly is satisfied, as shown in Fig. 4b. However, in case of localized errors, most likely to occur at special points such as the boundary points and the poles in our model, the gradient checks may not be able to detect them. Such errors could negatively impact the behavior of the minimization and the retrieval of the initial fields. Hence, additional examinations must be carried out.

A crucial test of the overall performance of the variational data assimilation consists of examining the behavior of the minimization and the retrieval of the initial state by inserting data only at the end of the assimilation period. An experiment with an assimilation period of 3 h is carried out. Figure 5 displays the evolution of the cost function and the gradient norm normalized by their respective initial values versus the number of minimization iterations. The values of the cost function and its gradient norm are reduced by four orders and three orders of magnitude after 40 minimization iterations, respectively, and results after more iterations show further reduction. The reduction of the initial distance function is smaller (about 2.5 orders of magnitude), as indicated in Fig. 7. In other words, the rate of convergence of the initial distance function lagged behind that of the cost function itself. As explained by Thépaut and Courtier (1991), the dynamics of the model is the reason for this loss of conditioning. However, the reduction of distance function after 40 iterations is significant and the retrieved initial field is acceptable; the difference field for the retrieved geopotential as compared to the reference field is shown in Fig. 6c.

c. *Length of assimilation interval*

The loss of the conditioning of the initial distance function, caused by the dynamics of the model as discussed earlier, is expected to be enhanced by the length of an assimilation interval. Producing experiments with observations available only at the end of the assimilation interval of 3, 6, and 9 h, respectively, should enable us to demonstrate the presence of the conditioning problem. Figure 7 displays the initial distance function $\langle x(t_0) - x_{\text{ref}}(t_0), x(t_0) - x_{\text{ref}}(t_0) \rangle$ versus the number of iterations as the minimization proceeds. The loss of conditioning, as the length of the assimilation interval increases, becomes very evident. For the experiments with a 6-h interval, the retrieved initial field is acceptable after 40 minimization iterations, while a

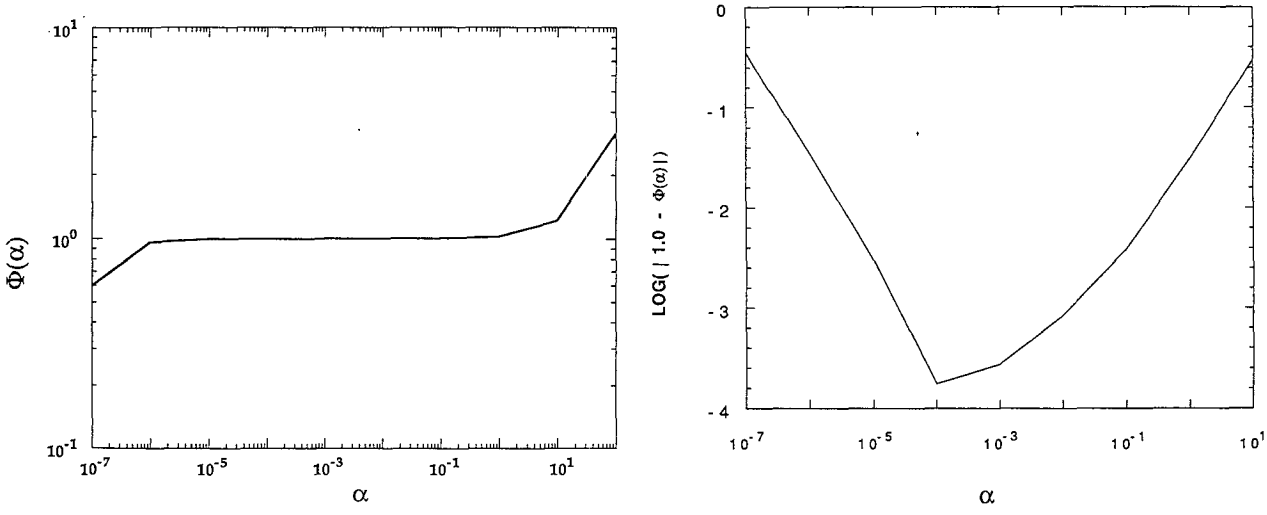

FIG. 4. Gradient check of the adjoint model: (a) $\Phi(\alpha)$, (b) $\log[|1.0 - \Phi(\alpha)|]$.

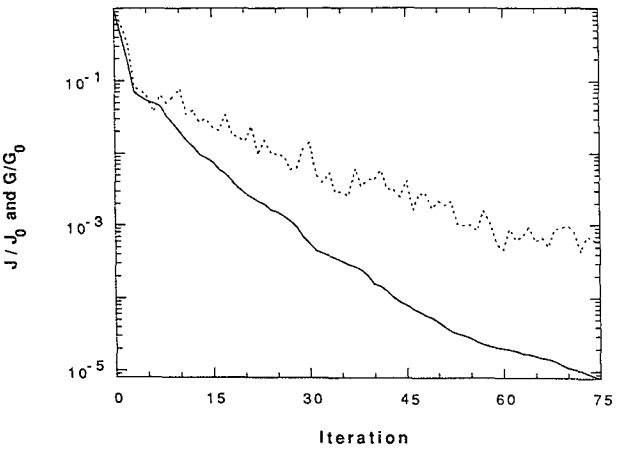

FIG. 5. Cost function and gradient norm normalized by their respective initial values for the experiment with observations available only at the end of the 3-h assimilation interval.

significant degradation in the quality of the retrieval is detected for the experiment with a 9-h interval. A much higher number of iterations in the minimization is required for the 9-h interval experiment to attain the same rms level as that obtained by the experiment with a 3-h interval after 40 iterations. These results are consistent with Thépaut and Courtier (1991) and Zou et al. (1992).

An examination of the consecutive retrieved initial fields (not shown here) indicates that the major differences related to the retrieved initial fields with different assimilation intervals are mainly evident on the small scales of motion, which makes sense dynamically. The time scale is smaller for the small-scale features, which, being governed by the dynamics, then experience a larger change over a given period of time and are therefore subject to more loss of conditioning.

d. Impact of the divergence damping coefficient

It was found in Bates et al. (1990) that in the absence of an initialized start it was necessary to use an appreciable value of the divergence damping coefficient c in order to suppress short-wave gravity waves. Since the initial state used by us for the variational data assimilation experiments is well balanced, the damping coefficient could be smaller than the one used in Bates et al. (1990), but a nontrivial value of c is, however, necessary. It is interesting to examine how the magnitude of the divergence coefficient will impact upon the retrieval of the initial state.

With observations inserted only at the end of a 3-h assimilation interval but with different values of the divergence damping coefficient c ranging from 1.0×10^6 to 5.0×10^7 m^2 s^{-1}, which correspond to an e-folding time ranging from 5 to 0.1 h for the shortest wave resolved for divergence, the corresponding initial distance functions for these experiments are displayed in Fig. 8. It is clearly seen that the convergence rate becomes slower as the damping coefficient increases. This can be easily explained by noting that the divergence damping acts as a dissipation mechanism that remains dissipative in the backward adjoint integration. Information contained in the initial fields, especially in the small-scale features, will be damped so that the retrieval of this part of the initial state from the information at some later time becomes more difficult. This result confirms the results of Thépaut and Courtier (1991) where they examined the impact of horizontal diffusion. A thorough analysis on the impact of diffusion on the adjoint variational data assimilation method can be found in Li and Droegemeier (1993).

e. Impact of the time-step size

The major advantage of semi-Lagrangian models over Eulerian ones is their computational efficiency due to their not being subject to the CFL computational stability criterion. This is also the justification of our motivation in using semi-Lagrangian models in order to ease the computational cost involved in adjoint data assimilation. However, the question of whether the computational efficiency can be preserved during the iterative minimization is worth investigating, though the answer is obviously expected to be positive.

Two experiments are carried out, both being of identical-twin type. The assimilation window is 12 h and data are inserted every hour. The first guesses between the initial state are also identical. The only difference between the two experiments is that a time step of 1 h is used for one and a time step of 10 min is used for the other; hence, data are inserted every time step during the adjoint integration in the first experiment and every six time steps in the second experiment. Figure 9 shows the respective normalized cost functions for these two experiments versus the number of iterations. The two cost functions evolve similarly but not in an identical way. A positive feature is that the experiment with the 1-h time step converges slightly faster than the corresponding experiment with the 10-min time step. This could be explained as follows. First, the information of the forcing term where data were inserted every time step is passed to the gradient more directly. Second, a smaller number of computational operations is required for the integration with the larger time step, which results in less round-off error accumulation. Both factors contribute to a gradient of higher quality for the experiment with larger time step, a fact that slightly improves the minimization efficiency in terms of the number of iterations.

However, since the experiments are of identical-twin type and highly idealized, this result should not be generalized beyond computational aspect. It is only intended here to confirm the speculation that the computational efficiency in the forward integration by using semi-Lagrangian schemes is well preserved during the iterative minimization process related to variational data assimilation.

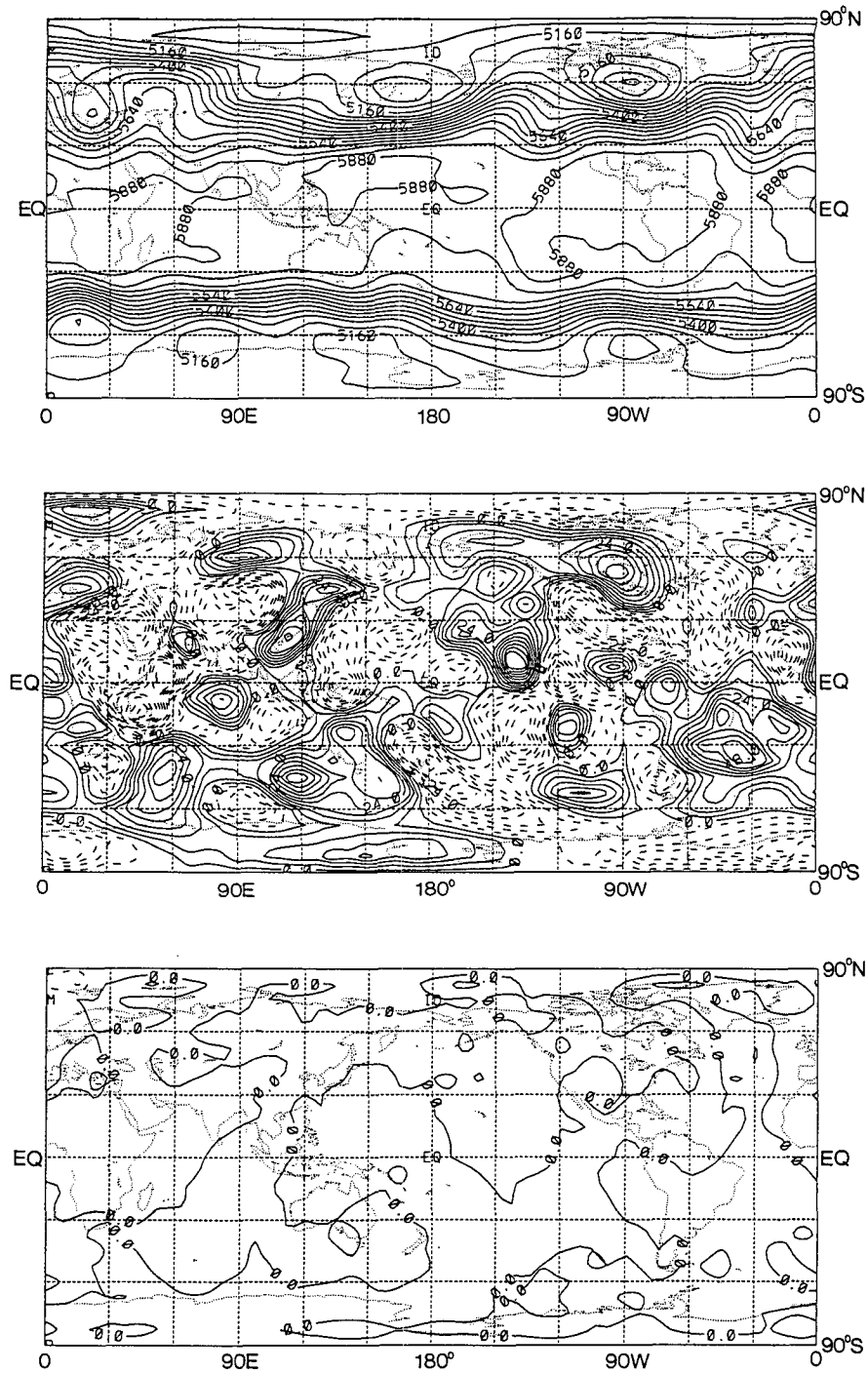

FIG. 6. (a) Initial condition of the reference geopotential. (b) Difference between the initial geopotential field and the reference before minimization (contour interval is 6.0 dam). (c) Difference between the initial geopotential field and the reference after 40 iterations of minimization (contour interval is 6.0 dam). Results are from the experiment where observations are available only at the end of the 3-h assimilation interval.

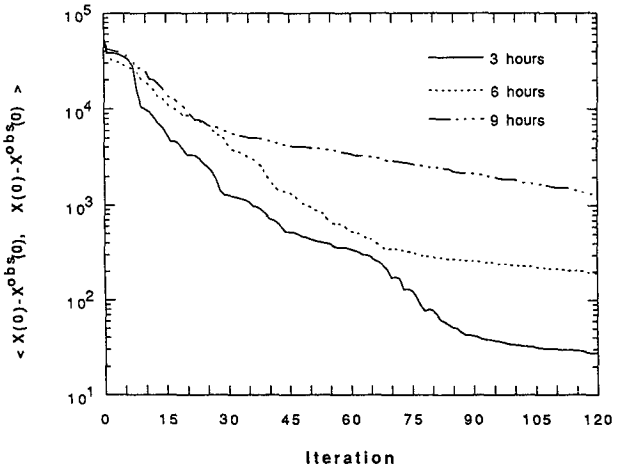

FIG. 7. Initial distance function with assimilation intervals equal to 3, 6, and 9 h.

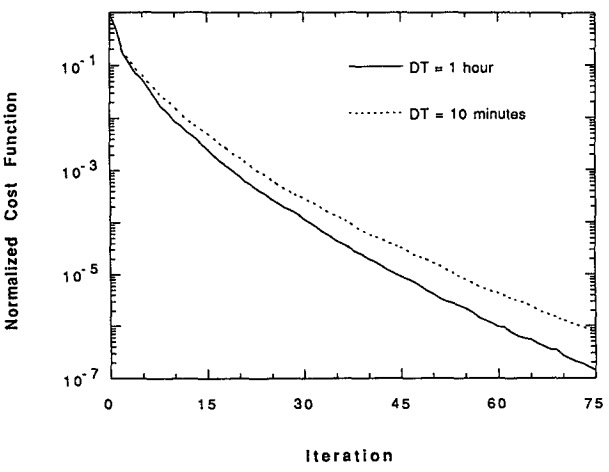

FIG. 9. Cost function normalized by its initial value with two different time-step sizes.

5. Concluding remarks

Adjoint model development and its application to variational data assimilation with a semi-Lagrangian semi-implicit model of the shallow-water equation on the sphere are found to be feasible. The seemingly "discontinuous" operations in the interpolation routine required in order to estimate the model variables at midtrajectory and departure points did not affect the adjoint model derivation. A special treatment was, however, required during the linearization process.

Due to the interpolations related to the semi-Lagrangian approach, the tangent linear model is more likely to break down in regions where sharp gradients prevail in the interpolated fields coupled with strong advective wind; such a situation is likely to occur in high latitudes near the pole, especially in the winter hemisphere. These effects increase, in a linear sense, as the temporal step size utilized in the model integration becomes larger. Since the validity of the adjoint approach is in question when the length of the assimilation interval exceeds the validity limit of the tangent linear model, these results impose a constraint on the choice of the temporal step size to be used in the adjoint variational data assimilation with semi-Lagrangian models.

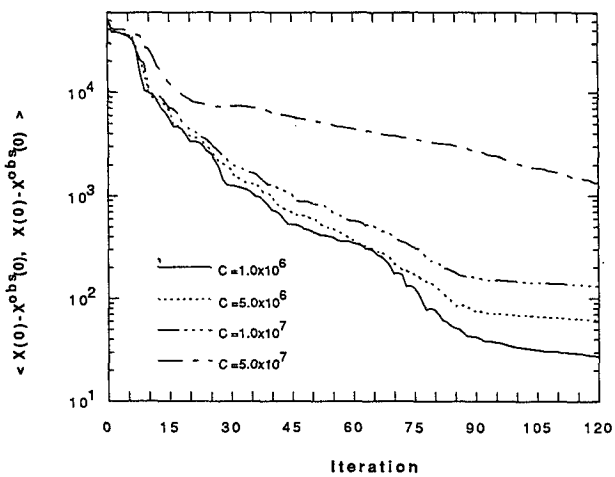

FIG. 8. Initial distance function with different values of the divergence damping coefficient c ($m^2 s^{-2}$). The assimilation interval is 3 h.

Data assimilation experiments with observations inserted only at the end of the assimilation period whose length is 3, 6, and 9 h, with the time step for integration being 1 h, all perform well in the minimization of the cost function and the retrieval of the initial state. Loss of conditioning in the initial distance function, as the assimilation interval increases, is identified. Numerical experiments have shown the impact of the magnitude of the divergence damping coefficient on the retrieval of the initial state, especially what concerns small-scale features. Two experiments were carried out for comparison purposes, both being of identical-twin type, with observations being inserted every hour, one with a time step of 1 h and the other with a time step of 10 min. It was found that the convergence rate of the minimization for the experiment with 1-h time step was slightly faster. This confirms the obvious expectation that the computational efficiency of semi-Lagrangian models is preserved in the large-scale unconstrained minimization process.

The results in this study are encouraging, illustrating the potential of semi-Lagrangian models for performing variational data assimilation. However, the impact of the interpolation on the limit of the validity of the tangent linear model, in the case of a baroclinic model with real data assimilation, needs to be further examined in order to ascertain whether large time steps allowed by semi-Lagrangian schemes have to be constrained by limits on the validity of the tangent linear

model to the extent that computational efficiency of semi-Lagrangian schemes may be lost. This first adjoint model development for variational data assimilation with a semi-Lagrangian semi-implicit model is the precursor of the development of a variational data assimilation system for a 3D semi-Lagrangian semi-implicit numerical weather prediction model (Bates et al. 1992), which is now in its final stages of implementation. These results will be reported in a forthcoming paper.

Acknowledgments. The first author wishes to thank Dr. X. Zou for some discussions concerning the strategy of adjoint coding. We are grateful to Dr. J. R. Bates who provided the forward 2D semi-Lagrangian semi-implicit model, and Dr. S. Moorthi and Dr. R. W. Higgins who provided the direct solver for the elliptic equation. Our thanks extend to Drs. J. Derber and M. Zupanski for some valuable remarks. The research effort of the first two authors was supported by NASA Grant NAG-5-1660, managed by Dr. K. Bergman at NASA Headquarters. Additional support was provided by the Supercomputer Computations Research Institute at Florida State University, which is partially funded by the Department of Energy through Contract No. DE-FC0583ER250000.

APPENDIX

Adjoint Derivation of Interpolation Routines

Let us conceptually construct the following problem. For some model variable Y, we have gridpoint values $Y(I)$ at gridpoint locations $X(I)$, where $[X(I) = I$, for $I = 1, 2, 3, \cdots, N]$, that are uniformly distributed with spacings of unit grid interval. Consider a location at X_P that is also a model product, that is, a function of the control variables. Point X_P is located somewhere between two grid points $X(J)$ and $X(J+1)$. Now, we wish to linearly interpolate $Y(J)$ and $Y(J+1)$, which are defined at the gridpoint locations to X_P, a nongridpoint location. The interpolated value Y_P will then become a model quantity.

The program for this operation is as follows:

$$
\begin{aligned}
X_P &= C\,(\text{some value}) \\
J &= \text{integer}\,(X_P) \\
\text{JP1} &= J + 1 \\
\alpha &= X_P - J \\
\beta &= \text{JP1} - X_P \\
Y_P &= \alpha Y(\text{JP1}) + \beta Y(J).
\end{aligned}
$$

The purpose of taking the integer of a model variable and using this integer as the index of some array whose elements are also model variables is to keep track of those model variables that are used for the interpolation. As long as the one-to-one variable correspondence order is preserved, the coding should remain correct. Therefore, the important issue here is to save the model variables and the integers and to use these quantities in the adjoint coding. This idea can be implemented in the tangent linear model.

If we denote the basic-state variable with overbars, the tangent linear model operation becomes

$$
\begin{aligned}
\bar{X}_P &= C \\
J &= \text{integer}\,(\bar{X}_P) \\
\text{JP1} &= J + 1 \\
\bar{\alpha} &= \bar{X}_P - J \\
\bar{\beta} &= \text{JP1} - \bar{X}_P \\
\alpha' &= X'_P \\
\beta' &= -X'_P \\
Y'_P &= \bar{\alpha} Y'(\text{JP1}) + \alpha' \bar{Y}(\text{JP1}) + \bar{\beta} Y'(J) + \beta' \bar{Y}(J).
\end{aligned}
$$

The corresponding adjoint program then becomes

$$
\begin{aligned}
\bar{X}_P &= C \\
J &= \text{integer}\,(\bar{X}_P) \\
\text{JP1} &= J + 1 \\
\bar{\alpha} &= \bar{X}_P - J \\
\bar{\beta} &= \text{JP1} - \bar{X}_P \\
\beta' &= Y'_P \bar{Y}(J) \\
Y'(J) &= \bar{\beta} Y'_P + Y'(J) \\
\alpha' &= Y'_P \bar{Y}(\text{JP1}) \\
Y'(\text{JP1}) &= \bar{\alpha} Y'_P + Y'(\text{JP1}) \\
X'_P &= -\beta' + X'_P \\
X'_P &= \alpha' + X'_P.
\end{aligned}
$$

REFERENCES

Bates, J. R., and A. McDonald, 1982: Multiply-upstream, semi-Lagrangian advective schemes: Analysis and application to a multilevel primitive equation model. *Mon. Wea. Rev.*, **110,** 1831–1842.

——, S. Moorthi, and R. W. Higgins, 1993: A global multilevel atmospheric model using a vector semi-Lagrangian finite-difference scheme. Part I: Adiabatic formulation. *Mon. Wea. Rev.*, **121,** 244–263.

——, F. H. M. Semazzi, R. W. Higgins, and R. W. Barros, 1990: Integration of the shallow water equations on the sphere using a vector semi-Lagrangian scheme with a multigrid solver. *Mon. Wea. Rev.*, **118,** 1615–1627.

Chao, W. C., and L. P. Chang, 1992: Development of a four-dimensional variational analysis system using the adjoint method at GLA: Part 1: Dynamics. *Mon. Wea. Rev.*, **120,** 1661–1673.

Côté, J., and A. Staniforth, 1988: A two-time-level semi-Lagrangian semi-implicit scheme for spectral models. *Mon. Wea. Rev.*, **116,** 2003–2012.

Courtier, P., 1985: Experiments in data assimilation using the adjoint

model technique. *Workshop on High-Resolution Analysis, ECMWF* 20 pp.

——, and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus,* **42A,** 531–549.

Derber, J. C., 1985: The variational 4-D assimilation of analyses using filtered models as constraints. Ph.D. thesis, University of Wisconsin—Madison, 142 pp.

Errico, R. M., and T. Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU–NCAR mesoscale model. *Mon. Wea. Rev.,* **120,** 1644–1660.

Krishnamurti, T. N., 1962: Numerical integration of primitive equations by quasi-Lagrangian advective scheme. *J. Appl. Meteor.,* **1,** 508–521.

Lacarra, J.-F., and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. *Tellus,* **40A,** 81–95.

Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus,* **38A,** 97–110.

Lewis, J. M., and J. C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constrains. *Tellus* **37A,** 309–322.

Li, Y., and K. K. Droegemeier, 1993: The influence of diffusion associated errors on the adjoint data assimilation technique. *Tellus,* in press.

——, I. M. Navon, S. Moorthi, and R. W. Higgins, 1991: The 2-D semi-implicit semi-Lagrangian global model: Direct solver, vectorization and adjoint model development. *Tech. Rep. FSU-SCRI-91-158,*

Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Mathematical Programming,* **45,** 503–528.

McDonald, A., and J. R. Bates, 1987: Improving the estimate of the departure point position in a two-time level semi-Lagrangian and semi-implicit scheme. *Mon. Wea. Rev.,* **115,** 737–739.

——, and ——, 1989: Semi-Lagrangian integration of a gridpoint shallow water model on the sphere. *Mon. Wea. Rev.,* **117,** 130–137.

Moorthi, S., and R. W. Higgins, 1993: Application of fast Fourier transforms to the direct solution of a class of two-dimensional elliptic equations on the sphere. *Mon. Wea. Rev.,* **121,** 290–296.

Navon, I. M., and R. de Villiers, 1983: Combined penalty multiplier optimization methods to enforce integral invariant conservation. *Mon. Wea. Rev.,* **111,** 1228–1243.

——, and D. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.,* **115,** 1479–1502.

——, X. Zou, K. Johnson, J. Derber, and J. Sela, 1990: Variational real-data assimilation with the NMC spectral model. Part I: Adiabatic model test. *Int. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Clermont-Ferrand, France, 341–348.

——, ——, J. Derber, and J. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.,* **120,** 1433–1446.

Nocedal, J., 1980: Updating quasi-Newton matrices with limited storage. *Math. Comput.,* **35,** 773–782.

Rabier, F., and P. Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteor. Soc.,* **118,** 649–672.

Ritchie, H., 1988: Application of the semi-Lagrangian method to a spectral model of the shallow water equations. *Mon. Wea. Rev.,* **116,** 1587–1598.

Robert, A., 1981: A stable numerical integration scheme for the primitive meteorological equations. *Atmos. Ocean,* **19,** 35–46.

——, 1982: A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive meteorological equations. *J. Meteor. Soc. Japan,* **60,** 319–324.

Staniforth, A., and J. Côté, 1991: Semi-Lagrangian integration schemes for atmospheric models—A review. *Mon. Wea. Rev.,* **119,** 2206–2223.

Thacker, W. C., and R. B. Long, 1988: Fitting dynamics to data. *J. Geophys. Res.,* **93,** 1227–1240.

Thépaut, J.-N., and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive equation model. *Quart. J. Roy. Meteor. Soc.,* **117,** 1225–1254.

Zou, X., I. M. Navon, and F. X. Le Dimet, 1992: Incomplete observations and control of gravity waves in variational data assimilation. *Tellus,* **44A,** 273–296.

——, ——, M. Berger, K. H. Phua, T. Schlick, and F. X. Le Dimet, 1993: Numerical experience with limited-memory quasi-Newton methods for large-scale unconstrained minimization. *SIAM J. Optimization,* in press.