Four-Dimensional Variational Data Assimilation Experiments with a Multilevel Semi-Lagrangian Semi-Implicit General Circulation Model

YONG LI*

NASA/Goddard Laboratory for Atmospheres, Greenbelt, Maryland

I. MICHAEL NAVON

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

WEIYU YANG AND XIAOLEI ZOU**

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

J. R. BATES, S. MOORTHI, AND R. W. HIGGINS

NASA/Goddard Laboratory for Atmospheres, Greenbelt, Maryland

(Manuscript received 6 July 1993, in final form 18 October 1993)

ABSTRACT

Four-dimensional variational data assimilation (VDA) experiments have been carried out using the adiabatic version of the NASA/Goddard Laboratory for Atmospheres semi-Lagrangian semi-implicit (SLSI) multilevel general circulation model. The limited-memory quasi-Newton minimization technique was used to find the minimum of the cost function. With model-generated observations, different first-guess initial conditions were used to carry out the experiments. The experiments included randomly perturbed initial conditions, as well as different weight matrices in the cost function.

The results show that 4D VDA works well with various initial conditions as control variables. Scaling the gradient of the cost function proves to be an effective method of improving the convergence rate of the VDA minimization process.

The impacts of the length of the assimilation interval and the time density of the observations on the convergence rate of the minimization have also been investigated. An improved assimilation was obtained when observations were available in selected segments of the assimilation window. Moreover, our 4D VDA experiments with the SLSI model confirm the results obtained by Navon et al. and Li et al. concerning the impact of the length of the assimilation window. The choice of an adequate time distribution of observations along with an appropriate length of assimilation interval is an important issue that will be further investigated.

1. Introduction

The adjoint method has become widely used in variational data assimilation (VDA) experiments and is soon to be implemented operationally. The method has been applied by Courtier (1985), Derber (1985), Lewis and Derber (1985), Le Dimet and Talagrand (1986), Thacker and Long (1988), and Courtier and Talagrand (1990) to barotropic models. Recently it has been extended and applied to various three-dimensional primitive equation models such as the National Meteorological Center (NMC) spectral model, the European Centre for Medium-Range Weather Forecasts (ECMWF) spectral model, as well as to the National Aeronautics and Space Administration/Goddard Laboratory for Atmospheres (NASA/GLA) fourth-order A-grid finite-difference model, by Navon et al. (1990, 1992), Thépaut and Courtier (1991), and Chao and Chang (1992), respectively. Extensions to operational models, including segments of the physical packages, have recently been implemented successfully by Zou et al. (1993). Thépaut et al. (1993) have conducted VDA experiments with conventional observations and have carried out a comparison with optimal interpolation (OI), showing the advantage of 4D VDA. Ehrendorfer (1992) has compared VDA and Kalman filtering for idealized cases and has shown VDA to be advan-

* Also affiliated with General Sciences Corporation, a subsidiary of Science Applications International Corporation.

** Current affiliation: NCAR/MMM/MPS, Boulder, Colorado.

Corresponding author address: Dr. I. Michael Navon, SCRI-4052, 415 Science Library, The Florida State University, Tallahassee, FL 32306-4052.

© 1994 American Meteorological Society

tageous compared with Kalman filtering. The conclusions from these studies are very encouraging and constitute a major step toward operational application of the adjoint method using real data.

Due to the iterative nature of the large-scale unconstrained minimization process required by the adjoint method, which necessitates repeated integrations of the model and its adjoint over the span of the assimilation window, computational efficiency in integrating the model is one of the crucial factors determining operational application. One approach to this problem is to seek a temporal discretization scheme allowing the use of large time steps without introducing computational instability. Following the early work of Wiin-Nielsen (1959) and Krishnamurti (1962), a vast amount of research (Robert 1981, 1982; Bates and McDonald 1982; McDonald and Bates 1987, 1989; Ritchie 1988; Côté and Staniforth 1988; Staniforth and Côté 1991; Bates et al. 1990; Bates et al. 1993) has been dedicated to the semi-Lagrangian approach. Using semi-Lagrangian semi-implicit schemes, one may obtain practically unconditional computational stability (Robert 1981, 1982) for the time integration. A semi-Lagrangian scheme has been adopted for the moisture equation in the new National Center for Atmospheric Research Community Climate Model (CCM2), following the work of Williamson and Rasch (1989) and Rasch and Williamson (1990), and has been shown to contribute to an improved climate model (Rasch and Williamson 1991). A semi-Lagrangian spectral general circulation model (GCM) (Ritchie 1991) has been introduced for operational use in Canada, and a modified version of it has been adopted for operational use at ECMWF (Hortal and Simmons 1991; Simmons 1991; Temperton 1991). It is therefore natural to attempt to take advantage of the semi-Lagrangian semi-implicit scheme when performing 4D variational data assimilation.

However, a number of problems need to be solved in this area of research. An initial effort in this direction was conducted by Li et al. (1991, 1993), who derived the adjoint model for the semi-Lagrangian semi-implicit (SLSI) two-time-level finite-difference shallow-water equations model of Bates et al. (1990) (to be referred to as the BSHB model hereafter). This model used the direct solver of Moorthi and Higgins (1993). It was concluded that the adjoint derivation can proceed, despite the fact that the interpolations seem to involve algebraically discontinuous operations, as long as special treatment is applied during the stage of deriving the tangent linear model. It was also shown by conducting experiments of the "identical twin" type that the computational efficiency of the semi-Lagrangian semi-implicit scheme can be preserved during the process of minimization. Though the theoretical and numerical experiments of Li et al. (1993) have shown that the interpolation routines may limit the validity of the 2D semi-Lagrangian tangent linear model for large time steps, thus implying a constraint on either the maximum allowable time steps for model integration or on the length of the assimilation window, the study has demonstrated a potential advantage of semi-Lagrangian schemes for VDA using the adjoint technique.

Encouraged by the results obtained with the BSHB global shallow-water model and its adjoint, we extended our efforts to VDA experiments with the 3D adiabatic version of the NASA/GLA SLSI GCM developed by Bates et al. (1993), referred to as the BMH model hereafter. This paper presents the development of the 3D adjoint model and its application to a series of 4D VDA experiments.

A gradient check was conducted to ensure that the model is correct and can be used to perform the experiments. Impacts of various initial conditions and weighting matrices in the cost function were examined, as well as the effects of scaling the gradient. The 4D VDA experiments performed well with random perturbations. The results of the experiments also suggest that, for some assimilation problems, performing the minimization process with appropriate scaling of the gradient of the cost function allows much improved convergence rates of the minimization and a better quality of the assimilation. Additionally, we investigate in this paper the effects of different lengths of the assimilation window and the time density of the observations on the convergence rate of the cost function.

This paper presents the first adjoint model developed so far for a 3D semi-Lagrangian GCM, which constitutes a test-bed for VDA with such models. The plan of the paper is as follows. In section 2 we provide a brief description of the BMH model. The tangent linear model and its adjoint, as well as a verification check of their correctness, are provided in sections 3 and 4, respectively. In section 5, we present a series of 4D variational data assimilation experiments and discuss various issues related to the length of the assimilation interval and the time density of the observations. A summary and conclusions are presented in section 6.

2. The BMH model

The BMH model is a two-time-level SLSI finite-difference GCM that uses a σ coordinate in the vertical and regular latitude–longitude coordinates in the horizontal. The vertical discretization is based on a Lorenz grid (Lorenz 1960), while the horizontal discretization is based on a fully staggered C grid (Arakawa and Lamb 1977). In this paper, the horizontal resolution for the model is $(\Delta\theta, \Delta\lambda) = (4°, 5°)$ and there are eight vertical layers. The time step for the integration is chosen to be 30 min. For further details concerning the discretized model, we refer to Bates et al. (1993).

3. The tangent linear model

To derive the adjoint model, we need first to linearize the forward model in the vicinity of a basic state, which

is on a model trajectory. This leads to the tangent linear model. The notation used here follows Rabier and Courtier (1992).

To verify the correctness of the tangent linear model, we compared the output of each subroutine of the tangent linear model with its counterpart in the original forward model. To verify the full tangent linear model, we employed a more quantitative method, described herein.

The evolution of $\mathbf{x}$ is given by the integration of the model $\mathbf{M}$ between times t_0 and t_n as

$$\mathbf{x}(t_n) = \mathbf{M}(t_n, t_0)[\mathbf{x}(t_0)] = \mathbf{M}(t_n, t_0)[\mathbf{x}_0(t_0) + \delta\mathbf{x}(t_0)], \quad (3.1)$$

whereas the first-order evolution of the perturbation $\delta\mathbf{x}(t_n)$ is the result of the integration of the tangent linear model $\mathbf{R}$:

$$\delta\mathbf{x}(t_n) = \mathbf{R}(t_n, t_0)\delta\mathbf{x}(t_0). \quad (3.2)$$

We then compare the total perturbation

$$\mathbf{N}[\delta\mathbf{x}(t_0)] = \mathbf{M}(t_n, t_0)[\mathbf{x}_0(t_0) + \delta\mathbf{x}(t_0)] - \mathbf{M}(t_n, t_0)[\mathbf{x}_0(t_0)] \quad (3.3)$$

with its linear component

$$\mathbf{L}[\delta\mathbf{x}(t_0)] = \mathbf{R}(t_n, t_0)\delta\mathbf{x}(t_0). \quad (3.4)$$

The difference between the two is denoted as

$$\mathbf{D}[\delta\mathbf{x}(t_0)] = \mathbf{N}[\delta\mathbf{x}(t_0)] - \mathbf{L}[\delta\mathbf{x}(t_0)]. \quad (3.5)$$

To quantify this comparison, we choose a norm whose square is defined by

$$\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{W}\mathbf{x} \quad (3.6)$$

in accordance with the norm used in the inner product of the cost function for the data assimilation problem. The relative difference between the tangent linear model and the nonlinear forward model is then defined as the ratio $\|\mathbf{D}\|/\|\mathbf{L}\|$. We first examine different components of $\|\mathbf{D}\|/\|\mathbf{L}\|$ according to individual model variables contributions $[u, v, T'/\bar{T}, (\ln p_s)']$.

A 48-h model integration was carried out starting from an ECMWF analysis. The output of this 48-h preparatory integration was then used as the data source to verify the tangent linear model. As in Rabier and Courtier (1992), we chose the zonal average fields of the 48-h preparatory integration as the basic-state initial condition $[\mathbf{x}_0(t_0)]$, while the departure of the zonal average fields multiplied by a variable parameter α serves as the perturbation of the initial condition $\delta\mathbf{x}(t_0)$. The amplitudes of the different components of this perturbation are very large when $\alpha = 1$: at the $\sigma = 0.26$ level, the zonal wind perturbation reaches up to 38 m s^{-1}; at the $\sigma = 0.38$ level, the meridional wind perturbation attains 51 m s^{-1}; that of the temperature at the lowest σ level is about 36 K; and the maximum perturbation of surface pressure is -429 hPa, due to the orography of the Tibetan Plateau. The tangent linear model check with these strongly perturbed initial conditions can shed light on possible errors. The period of integration t_n is taken to be 12 h.

TABLE 1a. Correlation coefficients between **N** field and **L** field.

α	u	v	$\frac{T'}{T}$	$(\ln p_s)'$
1.0	0.9389941	0.9088381	0.8985242	0.9556994
10^{-1}	0.9986423	0.9943771	0.9970486	0.9993822
10^{-2}	0.9981816	0.9988431	0.9659738	0.9993876
10^{-3}	0.9999996	0.9999990	0.9999994	0.9999999
10^{-4}	1.0000000	0.9999999	1.0000000	1.0000000
10^{-5}	1.0000000	1.0000000	1.0000000	1.0000000

Table 1a presents the correlation between the **N** field and the **L** field for various values of the parameter α, while Table 1b displays the relative error $\|\mathbf{D}\|/\|\mathbf{L}\|$. From these tables we see that all correlation coefficients between the nonlinear output fields **N** and the linear output fields **L** for each of the variable fields exceed 90% and reach values near unity when α is less than or equal to 0.1. As α decreases, the relative errors decrease to very small values. The correlation coefficients reach to seven- to eight-digit accuracy close to unity, and the relative error values are at 10^{-4} scale when α is equal to 10^{-4} or 10^{-5}. These results provide a reliable indication about the correctness of the tangent linear model code.

To assess the impact of the length of the integration period on the validity of the tangent linear model, we carried out a check for different lengths, up to 120 h. For convenience, we averaged the correlation coefficients of the four model variables and used the norm defined in (3.6) for $\|\mathbf{D}\|$ and $\|\mathbf{L}\|$ to calculate the relative error. The diagonal component values of the weighting matrix used are $\mathbf{W}_u = 10^{-2}\mathbf{I}$ (s^2 m^{-2}), $\mathbf{W}_v = 10^{-2}\mathbf{I}$ (s^2 m^{-2}), $\mathbf{W}_{T'/\bar{T}} = 2 \times 10^3\mathbf{I}$, and $\mathbf{W}_{(\ln p_s)'} = 5 \times 10^3\mathbf{I}$ (ln^{-2} hPa), respectively. We chose the values of the parameter α as 1.0, representing a strong perturbation, and as 0.1, representing the normal perturbation.

Figure 1a presents the correlation coefficients between the **N** fields and the **L** fields for various values of the parameter α with respect to different lengths of

TABLE 1b. Relative error $\|\mathbf{D}\|/\|\mathbf{L}\|$ (%).

α	u	v	$\frac{T'}{T}$	$(\ln p_s)'$
1.0	34.85	42.09	44.12	28.77
10^{-1}	5.22	10.63	7.63	3.44
10^{-2}	6.05	4.81	26.40	3.44
10^{-3}	8.85×10^{-2}	0.14	0.11	4.41×10^{-2}
10^{-4}	2.77×10^{-2}	4.14×10^{-2}	2.67×10^{-2}	1.49×10^{-2}
10^{-5}	1.71×10^{-2}	2.16×10^{-2}	1.38×10^{-2}	1.34×10^{-2}

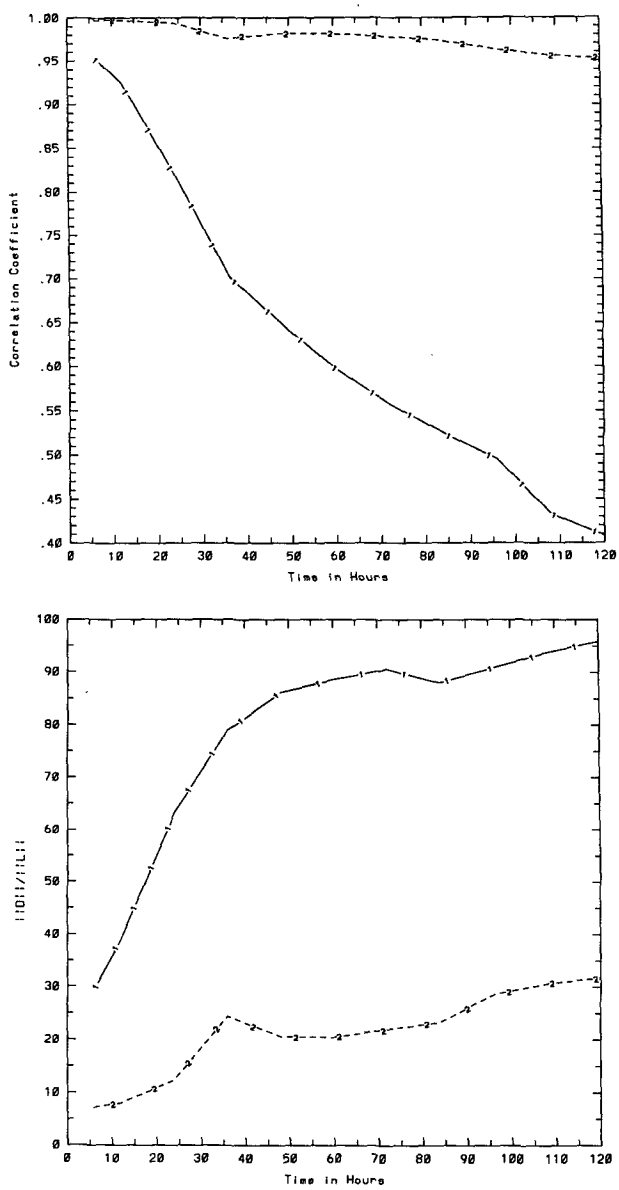

FIG. 1. Correlation coefficient between (a) the **N** fields and the **L** fields, and the relative error (b) $\|\mathbf{D}\|/\|\mathbf{L}\|$ with respect to integration periods. Curve 1 corresponds to $\alpha = 1$, and curve 2 corresponds to $\alpha = 0.1$.

the integration period. Figure 1b displays the relative error $\|\mathbf{D}\|/\|\mathbf{L}\|$ curves. Considering the $\alpha = 0.1$ curve we deduce that the error of the tangent linear model output is small when normal perturbations are used; that is, the scales of the perturbations of wind, temperature, and surface pressure are at 10^0 m s^{-1}, 10^0 K, and 10^1 hPa, respectively. Even for integration periods of up to 120 h, we find that the correlation coefficient exceeds 95% and the relative error is 31.76%. These numerical results show that the tangent linear model well approximates the nonlinear forward model for up to 5 days when the initial perturbations are not too strong. For strongly perturbed initial conditions ($\alpha = 1$), the validity of the tangent linear model decreases as the length of the integration period is increased. With these strong perturbations, the validity limit of the tangent linear model is less than 1 day. In general, such strong perturbations are not encountered in real operational problems, so we conclude that the tangent linear model is a good approximation for the nonlinear forward modeling process.

4. Adjoint model development

We chose the cost function J as follows:

$$J[\mathbf{X}(t_0)] = \frac{1}{2}\sum_{r=0}^{R}[\mathbf{X}(t_r) - \mathbf{X}^{\text{obs}}(t_r)]^{\mathrm{T}} \times \mathbf{W}(t_r)[\mathbf{X}(t_r) - \mathbf{X}^{\text{obs}}(t_r)], \quad (4.1)$$

where $\mathbf{X}(t_r)$ is the $N[=M(3K+1)]$ component vector containing values of $[u, v, T'/\bar{T}, (\ln p_s)']$, with which the BMH model is initialized, over all grid points and at all vertical levels at time t_r; M is the number of grid points at each vertical level; K is the number of vertical levels; R is the number of time levels for the analyzed fields in the assimilation window; t_r is a certain observation time in the assimilation window; $\mathbf{X}^{\text{obs}}(t_r)$ is the N-component vector of analyzed values of $\mathbf{X}$ over all grid points on all levels at time t_r; and $\mathbf{W}(t_r)$ is an $N \times N$ diagonal weighting matrix, where $\mathbf{W}_u$, $\mathbf{W}_v$, $\mathbf{W}_{T'/\bar{T}}$, and $\mathbf{W}_{(\ln p_s)'}$ are diagonal submatrices consisting of weighting factors for each variable, respectively. Their respective values (as used in gradient check calculation) are $\mathbf{W}_u = 10^{-4}\mathbf{I}$ (s^2 m^{-2}), $\mathbf{W}_v = 10^{-3}\mathbf{I}$ (s^2 m^{-2}), and $\mathbf{W}_{T'/\bar{T}} = 10^{1}\mathbf{I}$, $\mathbf{W}_{(\ln p_s)'} = 10^{1}\mathbf{I}$ (ln^{-2} hPa).

Developing an adjoint model consists of linearizing the forward model and constructing the adjoint operators from the tangent linear model. With the adjoint model, one obtains the value of the gradient of the cost function with respect to the initial conditions by integrating the nonlinear model forward in time and its adjoint backward in time. For a detailed derivation of the adjoint model and the verification of its correctness, see Navon et al. (1992). A complete discussion related to the special problems of managing adjoint calculations for the grid interpolations in the semi-Lagrangian model is presented in the paper of Li et al. (1993). Since the BMH model consists of thousands of lines of code, any minor coding error may cause the final gradient of the cost function with respect to the control variables to be erroneous. Therefore, it is advantageous to verify the correctness of the linearization and adjoint coding segment by segment. Each segment may consist of either a subroutine or of several DO loops.

The correctness of the adjoint of each operator can be checked by applying the following identity (Navon et al. 1992):

$$(\mathbf{AQ})^{*\mathrm{T}}(\mathbf{AQ}) = \mathbf{Q}^{*\mathrm{T}}[\mathbf{A}^{*\mathrm{T}}(\mathbf{AQ})], \quad (4.2)$$

where $\mathbf{Q}$ represents the input of the original code and A represents either a single DO loop or a subroutine (see Navon et al. 1992). The left-hand side (lhs) involves only the tangent linear code, while the right-hand side (rhs) involves also the adjoint code ($\mathbf{A}^{*\mathrm{T}}$). If (4.2) holds, the adjoint code is correct when compared with the tangent linear model. In practice the identity (4.2) holds up only to machine accuracy. For a Cray computer that has intrinsic double precision, the lhs and the rhs of Eq. (4.2) should be equal up to the 11th significant digit.

Even after having successfully checked the adjoint of each operator matrix and finally the whole model according to the above identity, the adjoint model may still be subject to fatal coding errors. These errors are usually caused by linearization errors, which could not be detected in the process verifying the linearization code, partly due to the fact that the check itself is of an approximate nature and partly due to the choice of the basic state and other model parameters. The verification of the final gradient obtained through the integration of the adjoint model backward in time serves as an overall verification of the correctness of the adjoint model code and proves to be an essential step, in our experience. For a detailed derivation of the gradient check formula, see Navon et al. (1992).

The gradient check is shown in Fig. 2a. The value of the function $\phi(\alpha)$ equals unity to a high degree of accuracy when the parameter α varied from 10^{-2} to 10^{-11}, and obeys the monotonically decreasing rule when α decreases over 14 orders of magnitude.

The residual of $\phi(\alpha)$ [i.e., $|\phi(\alpha) - 1|$] was also checked and is displayed in Fig. 2b. We found that the residual approaches zero linearly. The check of the gradient verifies that the adjoint model is correct and can safely be used to perform 4D VDA experiments.

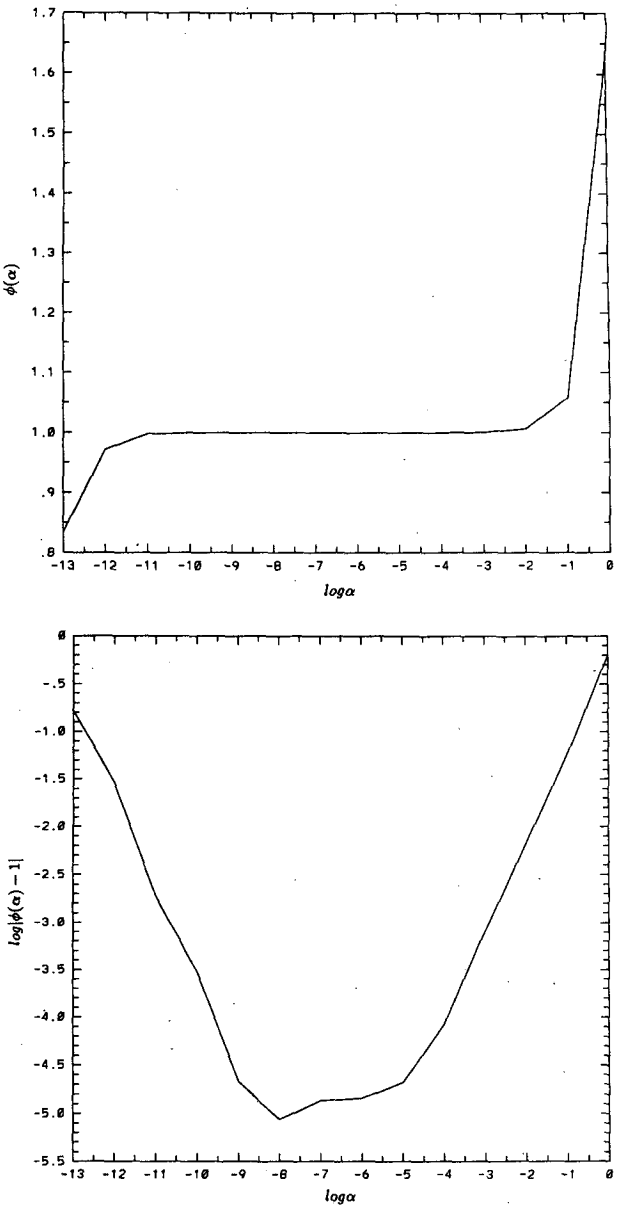

FIG. 2. Variation of (a) the function $\phi(\alpha)$ and (b) variation of the $\log|\phi(\alpha) - 1|$ with respect to $\log\alpha$.

5. Four-dimensional variational data assimilation experiments

a. *Observations, initial conditions, cost function, and scaling*

We carried out a series of 4D VDA experiments to verify the properties of the BMH model and its adjoint. A 4D VDA experiment consists of integrating the nonlinear forward model starting from a first-guess initial condition and then integrating the adjoint model backward in time. A forcing term is added whenever an observation is encountered, to obtain the values of the cost function and its gradient with respect to the control variable vector, which in this case is the initial condition vector. Then through an iterative large-scale unconstrained minimization process, the retrieved fields of the initial condition that minimizes the cost function (measuring the weighted lack of fit between model solution and observations) are obtained. The large-scale unconstrained minimization algorithm used throughout this paper is the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm of Liu and Nocedal (1989).

1) OBSERVATIONS

To ensure a consistent initial condition, we integrated the model for 48 h starting from the ECMWF analysis of 0000 UTC 15 January 1979. The 48-h preparatory integration was necessary to obtain a "realistic" initial condition devoid of large gravity wave

oscillations. The model output at the end of this integration period was used as the $t = 0$ "observations" and the model output from another 3-h continuous integration was used to generate $t = 3$ h observations. The observations consisted of the complete fields of the velocity components, temperature, and surface pressure. The $t = 0$ observations constitute the initial conditions for the 4D VDA experiments. If one reduces the preparatory integration time to 12 h, the output fields will display locally sharp gradients, due to the lack of initialization. For instance, the 500-hPa temperature field displays a region of very strong gradients located between south Asia and the Tibetan Plateau region.

2) FIRST-GUESS INITIAL CONDITIONS

We used three types of first-guess initial conditions to carry out our 4D VDA experiments. The first was obtained by multiplying the $t = 0$ observation fields by a constant coefficient A; for instance, $A = 1.1$. This initial condition is called IC_A.

The second was obtained by multiplying the $t = 0$ observation fields by a randomly perturbed coefficient vector $\mathbf{B}$; that is,

$$\mathbf{X}^{\text{initial}} = \mathbf{X}^{\text{obs}}(t_0) \times \mathbf{B} \tag{5.1a}$$

$$B_i = C_1 + C_2 D_i, \quad i = 1, 2, \cdots, N, \tag{5.1b}$$

where $\mathbf{X}^{\text{initial}}$ is the $N\,[=M(3K + 1)]$ component vector, containing values of $[u, v, T'/\bar{T}, (\ln p_s)']$, with which the BMH model is initialized, over all grid points and at all vertical levels at time t_0; M is the number of grid points at each vertical level; K is the number of vertical levels; D_i is a random coefficient whose values range from 0 to 1 with uniform distribution; C_1 and C_2 are constant coefficients—for instance, $C_1 = 0.75$ and $C_2 = 0.5$; $\mathbf{X}^{\text{obs}}(t_0)$ is the N-component vector of analyzed values of $\mathbf{X}$ over all grid points on all levels at time t_0. This initial condition is called IC_B.

The third initial condition consists of the output fields resulting from a 6-h integration carried out from the $t = 0$ observation fields. This initial condition is called IC_C.

3) WEIGHTING AND SCALING THE COST FUNCTION

The cost function J consists of a weighted least-squares difference between the model forecast and the observations, and is given by (4.1). The t_r consists of a 3-h interval. The observation fields are available at the beginning and the end of the assimilation interval.

The weights in the cost function serve two purposes: they reflect confidence in the quality of the observed data and scale J to be a nondimensional quantity (see Navon et al. 1992). In the 4D VDA experiments, two types of weighting matrices are used. The first type of weighting matrix is invariant with respect to the vertical levels. It is calculated by the following formula:

$$\mathbf{W}_{f,a} = \frac{1}{\max_{i,j,k} |f^{\text{obs}}_{i,j,k}(t_0) - f^{\text{obs}}_{i,j,k}(t_R)|^2}, \tag{5.2}$$

where i, j represent the horizontal grid points and k denotes the vertical levels; f represents the fields of u, v, $T'/\bar{T}$, or $(\ln p_s)'$, respectively. For the model-generated observation fields, we choose $t_0 = 0$ h and $t_R = 3$ h. The second type of weighting matrix uses weights that vary with vertical level. It is given by the following formula:

$$\mathbf{W}_{f,b}(\sigma) = \frac{1}{\max_{i,j} |f^{\text{obs}}_{i,j}(t_0) - f^{\text{obs}}_{i,j}(t_R)|^2}. \tag{5.3}$$

Let us discuss the difference between $\mathbf{W}_{f,a}$ and $\mathbf{W}_{f,b}(\sigma)$. With the weighting matrix $\mathbf{W}_{f,a}$, each variable uses the same constant weight $\mathbf{W}_{f,a}$. There are regions of large departure (i.e., difference between the initial conditions and the observations) at some vertical levels, and the cost function structure will be impacted mainly by these regions. It is reasonable to assume that the large-scale unconstrained minimization process will be weighted in favor of the larger departure values; that is, the minimization process will result in large adjustments occurring mainly at the levels where large departures occur. The other levels will presumably be less impacted during the large-scale minimization process. For the $\mathbf{W}_{f,b}(\sigma)$ case, the $\mathbf{W}_{f,b}(\sigma)$ weights will cause the departures at each vertical level to exhibit nearly the same order of importance in the cost function. It is thus expected that during the VDA minimization process all vertical levels may experience almost equal adjustments irrespective of the size of the departure values. However, it is clear that larger departure values will experience smaller adjustments than those obtained by using the $\mathbf{W}_{f,a}$ weighting matrix.

Scaling is a crucial issue in the success of nonlinear unconstrained optimization, and there is a rich literature pertaining to nonlinear programming problems (see Navon and de Villiers 1983; Navon and Legler 1987). A recent contribution to this area is a preconditioning algorithm for large-scale minimization tested with a certain degree of success by Zupanski (1993). Adequate scaling factors are a simple but crude way to do preconditioning, since ultimately we need to know the Hessian matrix of the cost function to efficiently deal with small-scale features. Knowing the spectrum of the Hessian of the cost function allows us to use sophisticated preconditioning aimed at reducing the condition number of the Hessian and more importantly—by means of preconditioning techniques such as incomplete factorization and polynomial preconditioning—to modify the spectrum of eigenvalues of the Hessian in such a way as to obtain several clusters of the eigenvalues. Iterative methods of optimization such as conjugate-gradient-like methods perceive each cluster as a single eigenvalue, and therefore their rate of convergence is greatly improved

by preconditioning. Also, any reduction in the condition number of the Hessian improves the rate of convergence that is asymptotically proportional to the condition number (see Axelsson and Barker 1984). We carried out 4D VDA experiments with and without scaling of the cost function and its gradient. The numerical results reveal that a proper scaling can markedly improve the convergence rate of the VDA minimization process.

b. *Four-dimensional VDA experiments without scaling the gradient of the cost function*

First, we carried out 4D VDA experiments with the three aforementioned initial conditions. No scaling of the cost function and its gradient was implemented in these experiments.

For the IC_A initial condition case, the coefficient used was $A = 1.1$. For the IC_B randomly perturbed initial condition case, the coefficients in (5.1) are $C_1 = 0.9$ and $C_2 = 0.2$, respectively.

For the third initial condition experiment IC_C, the effects of using both weighting matrices $\mathbf{W}_{f,a}$ and $\mathbf{W}_{f,b}$ in the cost function were tested separately. Since the results for both cases were found to be very similar, we restrict ourselves to presenting results only for the case with the $\mathbf{W}_{f,b}$ weighting matrix. All the other 4D VDA experiments presented in this paper use the $\mathbf{W}_{f,b}$ weighting matrix.

The numerical results obtained with the 4D VDA show that the cost function decreases steeply during the first few iterations of the unconstrained minimization process and finally converges to a value about 20%–30% (for the IC_A and IC_B cases) or 70% (for the IC_C case) of its initial value (Fig. 3). The logarithm of the gradient of the cost function decreases gradually by about two orders of magnitude (Fig. 4). The cost function could not be further reduced, even in the case where a very small departure from the initial condition ($A = 1.01$, IC_A case) is used as observations. The reason the cost function cannot be further reduced is that the minimization process in those cases appears to be effective in minimizing the lack of fit of the temperature and the surface pressure fields but less so in minimizing the lack of fit of the wind fields (Fig. 3); that is, the wind field components in the cost function were reduced much less than either the temperature or the surface pressure field components in all 4D VDA experiments that did not include scaling of the cost function and its gradient.

By analyzing the gradient of the cost function (Fig. 5), we see clearly that the lengths of the u and v components of the gradient are much shorter than the other two field components—that is, the temperature and surface pressure fields. We thus conclude that an adequate scaling of the gradient is necessary for obtaining a good retrieval.

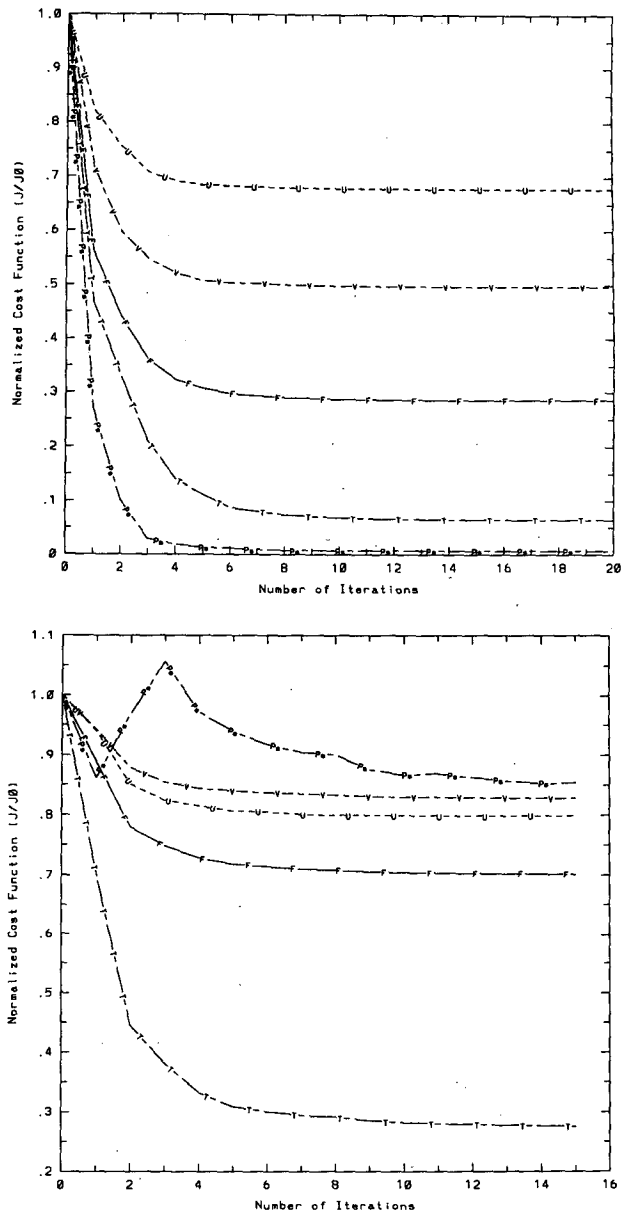

FIG. 3. Variation of each normalized component of the cost function as well as the total cost function versus the number of minimization iterations: (a) IC_B initial condition—$C_1 = 0.9$, $C_2 = 0.2$; (b) IC_C initial condition.

From the difference fields between the retrieved initial fields at the completion of the minimization process and the observations (Fig. 6), it is evident that the wind fields undergo only a small adjustment while the temperature and the surface pressure fields experience appreciable adjustments.

Comparing the results of the IC_A and the IC_B cases, we found that they are rather similar. Four-dimensional VDA worked well with randomly perturbed initial conditions (Fig. 6), while the rate of convergence of the minimization was not seriously affected.

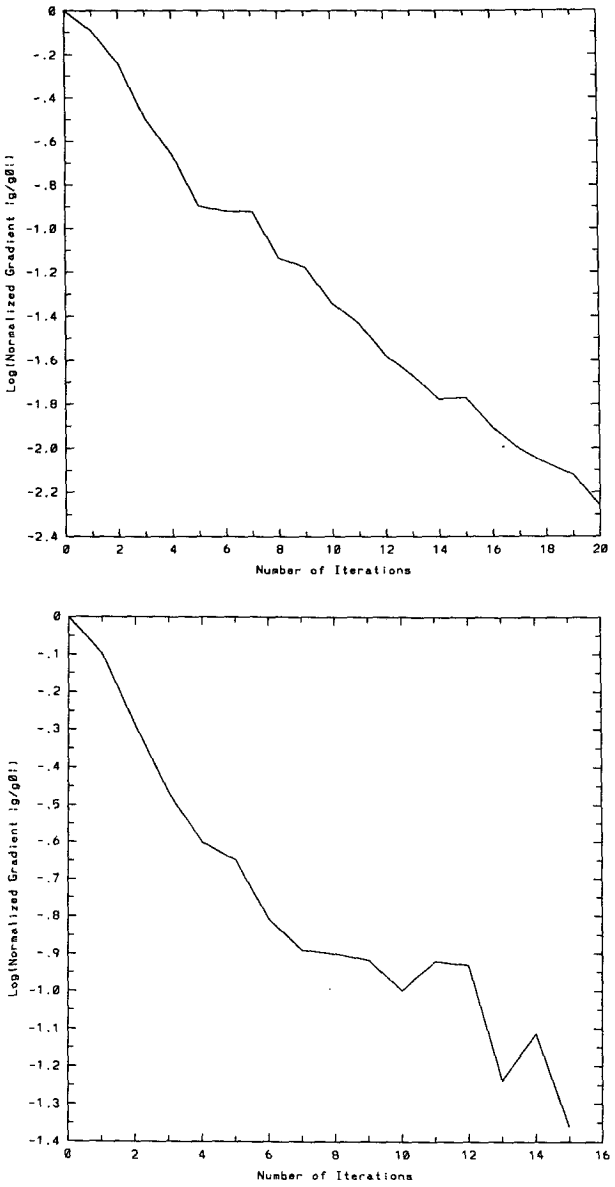

FIG. 4. Variation of the log of the normalized gradient of the cost function with the number of iterations: (a) IC_B initial condition—$C_1 = 0.9$, $C_2 = 0.2$; (b) IC_C initial condition.

Since the values of different control variables in the model have different scales, a proper weighting matrix in the cost function is necessary to balance the small-scale variables. Without it, the VDA minimization results with respect to small-scale variables will yield worse results. We carried out a 4D VDA experiment using the IC_B initial condition without any weight in the cost function (i.e., we set $W = I$) and found (Fig. 7) that the small-scale components, $T'/\bar{T}$ and $(\ln p_s)'$, of the cost function cannot be reduced to a satisfactory level; therefore, the total cost function cannot be reduced to a satisfactory value.

c. *Four-dimensional VDA experiments with scaling of the cost function and its gradient*

According to Gill et al. (1981), one distinguishes between four methods of scaling: by variable (which converts the variables from units that reflect the physical feature of the problem to units that display desirable properties for the minimization problem), constraint, gradient, and Hessian. Normally, scaling by variable is used because it is simple and saves computation time. For the problem at hand, however, the

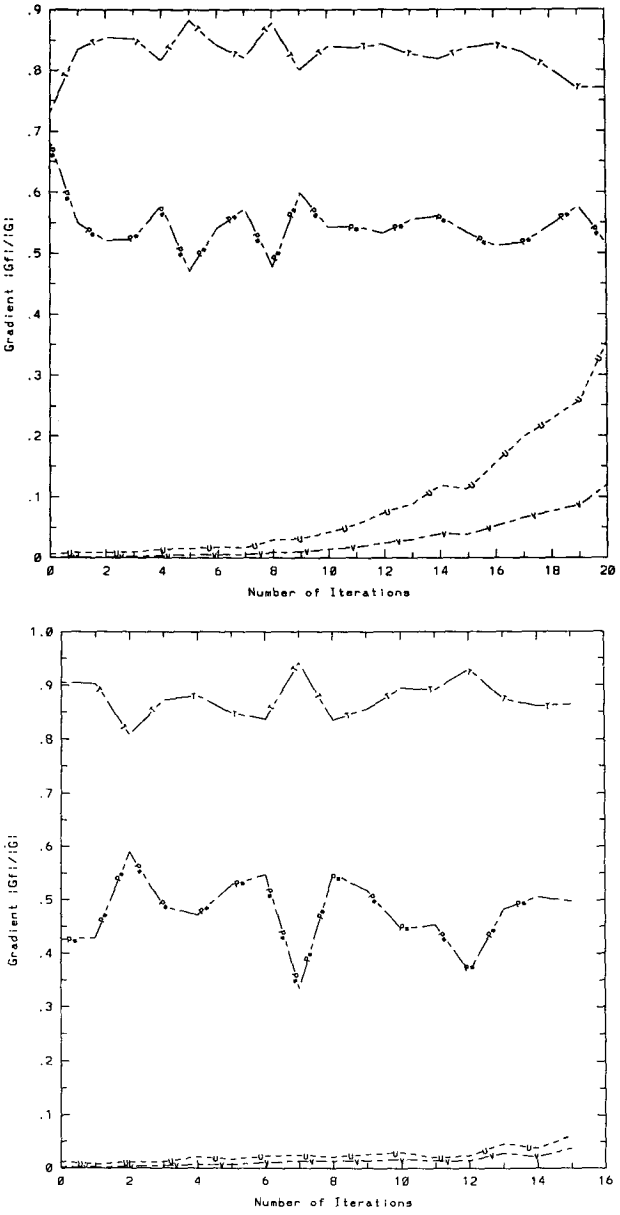

FIG. 5. Variation of each component of the gradient of the cost function ($|\nabla J_f|/|\nabla J|$) with the number of iterations: (a) IC_B initial condition—$C_1 = 0.9$, $C_2 = 0.2$; (b) IC_C initial condition.

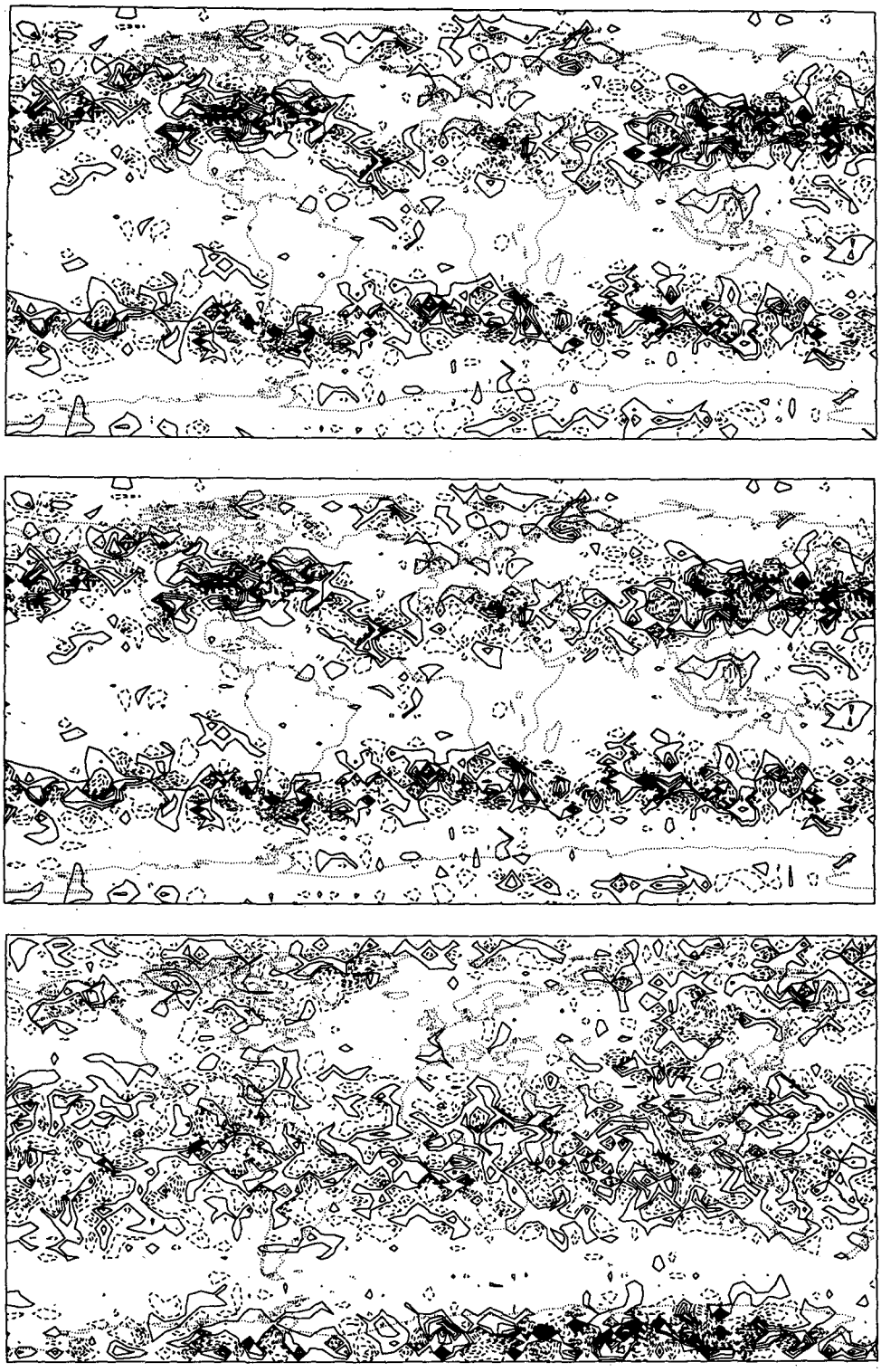

FIG. 6. Difference between the initial condition field [(a), (c), and (e)], the retrieval field after 20 minimization iterations [(b), (d), and (f)], and the corresponding "observation" field at 500 hPa. Using IC_B initial condition, $C_1 = 0.9$, $C_2 = 0.2$. (a) and (b) The zonal wind contour interval is 0.5 m s^{-1}. (c) and (d) The temperature contour interval is 0.5 K. (e) and (f) The surface pressure contour interval is 1 hPa.

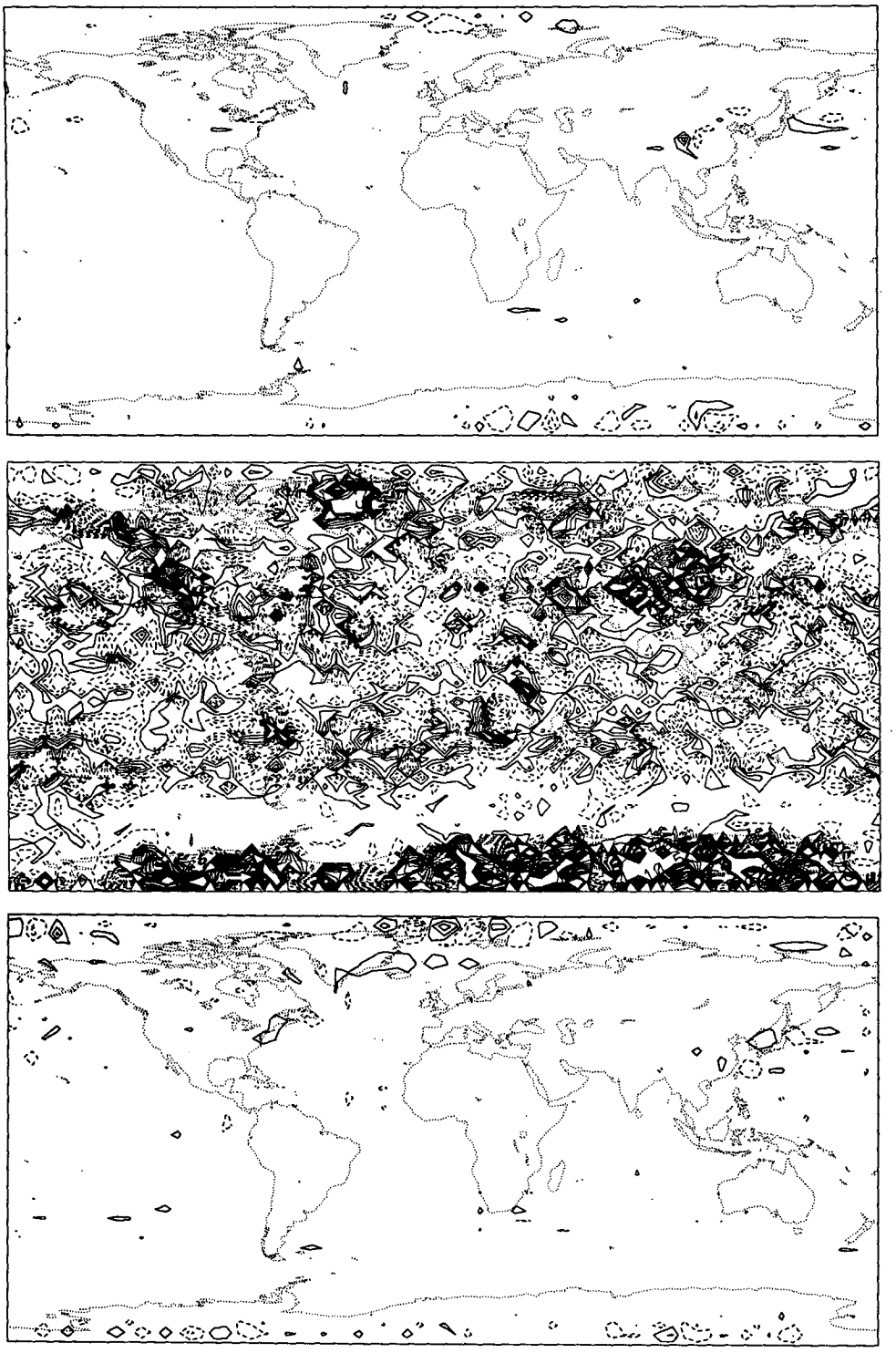

FIG. 6. (*Continued*)

relative magnitudes of the gradient field components of the cost function differ from each other by at least two orders of magnitude. This problem manifests itself in the fact that some components of the cost function cannot be reduced to a satisfactory value during the retrieval process. We therefore carried out some 4D VDA experiments using a gradient scaling and found that this gave much better results.

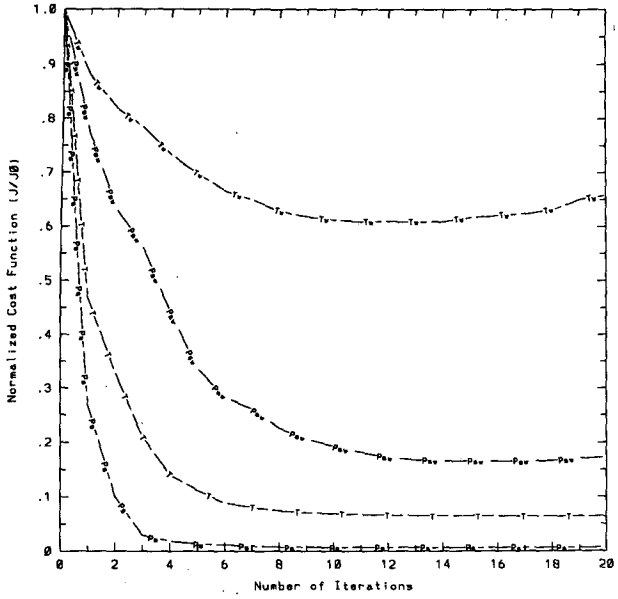

FIG. 7. Variation of the T'/T and $(\ln p_s)'$ normalized components of the cost function versus the number of minimization iterations. The subscript w on the curves denotes the case without the weight matrix ($\mathbf{W} = \mathbf{I}$) in the cost function.

Scaling the gradient adjusts the components to be about the same order of magnitude. This requires a change in the scaling coefficients based on the value of each variable component of the cost function gradient at each step of the minimization.

In our gradient scaling experiments, we used a rough method. Before carrying out each step of the minimization calculation, we implemented on each variable the following linear transformation:

$$\mathbf{X} = \mathbf{L}\mathbf{Y}, \tag{5.4}$$

where the diagonal matrix $\mathbf{L}$ consists of diagonal submatrices $[\mathbf{L}_u, \mathbf{L}_v, \mathbf{L}_{T'/\bar{T}}, \mathbf{L}_{(\ln p_s)'}]$ and the gradient of the cost function with respect to the new variables $\mathbf{Y}$ is

$$\nabla J_Y = \mathbf{L}\nabla J_X, \tag{5.5}$$

and we have

$$\mathbf{Y} = \mathbf{L}^{-1}\mathbf{X}. \tag{5.6}$$

The matrix $\mathbf{L}$ should be chosen so as to reduce the difference of each variable component vector length in ∇J_Y. We chose a gradient scaling coefficient that is $\mathbf{L}_u = 30\mathbf{I}$, $\mathbf{L}_v = 10\mathbf{I}$ (IC_A case) or $30\mathbf{I}$ (IC_B and IC_C cases), $\mathbf{L}_{T'/\bar{T}} = \mathbf{I}$, and $\mathbf{L}_{(\ln p_s)'} = \mathbf{I}$. Using this method, we repeated the aforementioned 4D VDA experiments with different initial conditions and found that the results exhibited a remarkable improvement as far as the rate of convergence of the minimization is concerned.

The results of the 4D VDA experiment with gradient scaling turned out to be markedly better than those obtained without gradient scaling in terms of convergence rate; that is, the cost function and its components were reduced to much lower values, especially the wind velocity field component (see Fig. 8, compared to Fig. 3). For the IC_C case, whereas the cost function could be reduced only to 70.1% of its initial value without gradient scaling, it was now reduced to 29.6% of its initial value. This result reveals that the 4D VDA model may deal with different complex problems as long as an appropriate scaling procedure of the cost function and its gradient is included. The adjusted wind com-

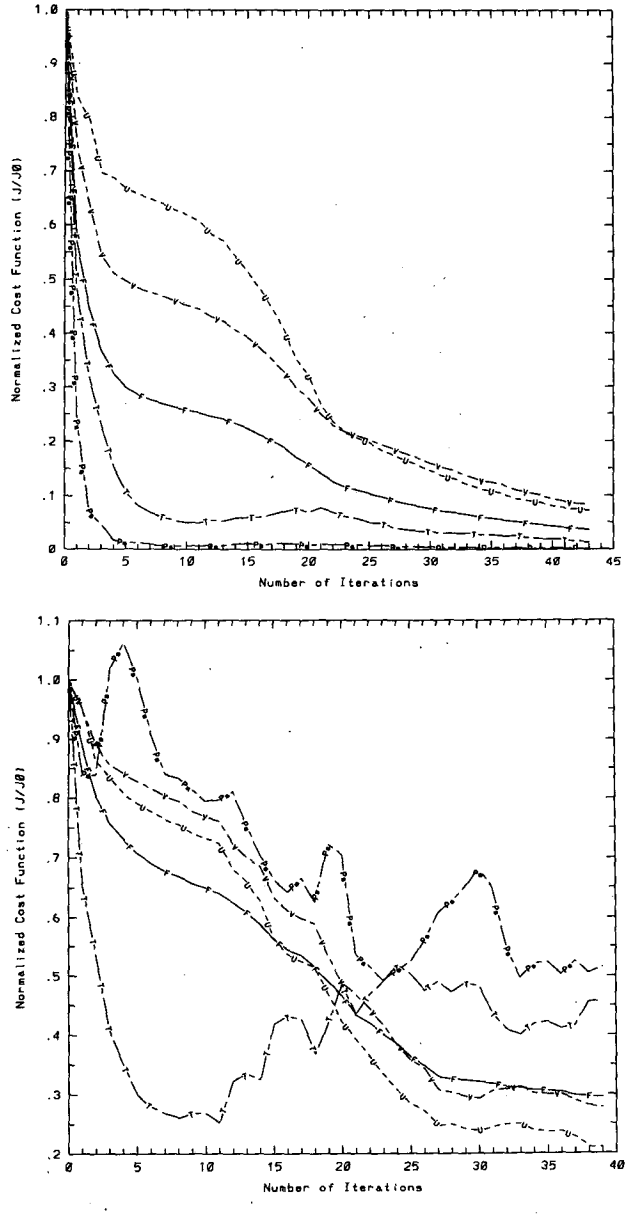

FIG. 8. Variation of each normalized component of the cost function as well as the total cost function with the number of iterations. Scaling of the gradient of the cost function is performed: (a) IC_B initial condition—$C_1 = 0.9$, $C_2 = 0.2$; (b) IC_C initial condition.

ponents of the gradient of cost function are also strengthened (Fig. 9, compared to Fig. 5).

Figure 10 presents the differences between the retrieval fields (at the final iteration step of the minimization process) and the observation fields of the zonal wind at 500 hPa. Comparing each retrieval field (figures omitted) with the case where no gradient scaling was applied (Fig. 6), we find that the gradient scaling allows us to obtain improved retrieval fields.

From the above results, we conclude that the minimization process performs much better for all the variable fields when scaling of the gradient of the cost function is used. The gradient scaling experiment resulted in a successful implementation of the 4D VDA for model-generated data.

d. Applicability of VDA to different initial conditions

To check the applicability of this 4D VDA scaling procedure to different types of initial conditions, we designed an experiment in which a more intensely randomized perturbation of the initial conditions was used. We chose $C_1 = 0.75$, $C_2 = 0.5$ in (5.1), and $\mathbf{L}_v = 30\mathbf{I}$.

From the differences between the initial conditions and the observations (Fig. 11), we conclude that these perturbation fields are stronger and more complex. Through the minimization process with gradient scaling, these intensely random perturbations were reduced to a much lower level and a successful retrieval was obtained (Fig. 12), with a remarkable reduction in all components of the cost function (Fig. 13). We also carried out a 4D VDA experiment using this initial condition without gradient scaling and found the results to be much worse than those with gradient scaling.

To illustrate that the 4D VDA minimization can be stopped before the cost function achieves its asymptotic rate of decrease, we integrated the nonlinear forward model from the initial observations, the first-guess initial conditions and the retrieved initial conditions, stopping after 10, 20, 30, and 40 minimization iterations. Figure 14 presents the time integrations of the surface pressure at a given grid point. We note that even after 10 iterations, when the cost function has decreased to only about 26% of its initial value, most of the information has been retrieved. After 20 iterations, the difference between the observations and retrievals becomes very small. Only small differences in the amplitude can be discerned, with almost no phase discrepancy. After 40 iterations, the time variation is similar to the curve resulting from integration from the observations. Thus, the 4D VDA minimizes most large-amplitude perturbations in the first 10–20 iterations, while in subsequent iterations only small-scale small-amplitude features are being assimilated.

To ensure that variational data assimilation with the present model can be carried out for a large period of synoptic significance (e.g., 24 h), a 4D VDA experiment with a 24-h assimilation window was performed.

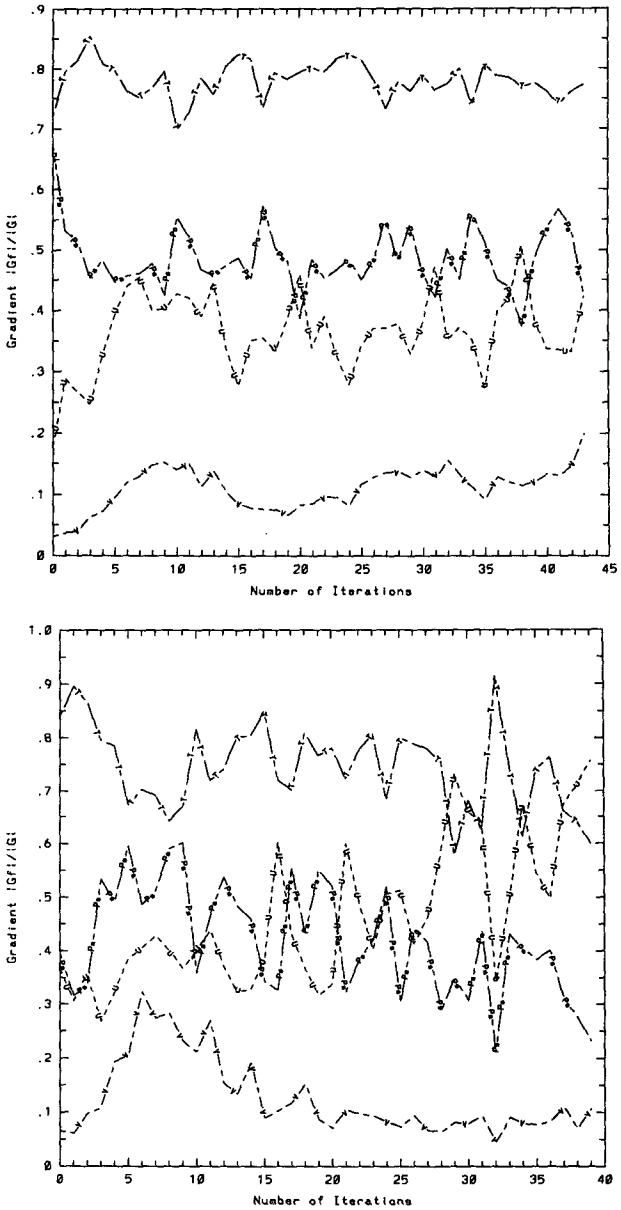

FIG. 9. Variation of each component of the gradient of the cost function ($|\nabla J_f|/|\nabla J|$) with the number of iterations, including scaling the gradient of the cost function: (a) IC_B initial condition—$C_1 = 0.9$, $C_2 = 0.2$; (b) IC_C initial condition.

Apart from the length of the assimilation window, all the parameters are identical to the VDA experiment performed with the 3-h window of assimilation.

Figure 15 displays the variation of each normalized component of the cost function as well as the total normalized cost function for the 24-h period of assimilation with the number of iterations. Compared with Fig. 13 (which corresponds to a 3-h window of assimilation), we note only a worsening in the convergence rate of the minimization for the horizontal wind com-

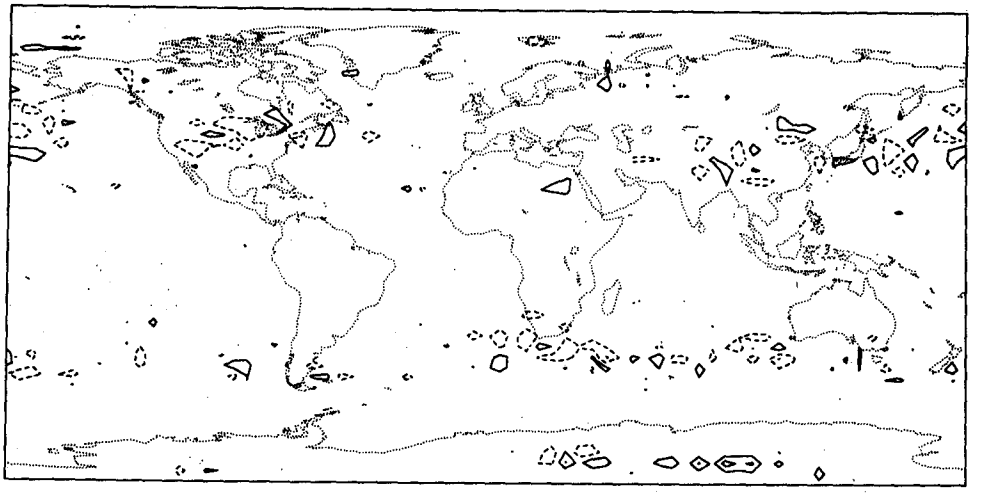

FIG. 10. Difference between the retrieval zonal wind field after 42 minimization iterations and the corresponding observation field at 500 hPa. Using IC_B initial condition, $C_1 = 0.9$, $C_2 = 0.2$. Scaling of the gradient of the cost function is performed. The contour interval is 0.5 m s^{-1}.

ponents, while the temperature and surface pressure components of the cost function display a better convergence rate. The total cost function remains almost unchanged. The total cost function as well as its components display a steep rate of decrease during the first minimization iterations for the 24-h assimilation window. A similar feature is observed for the retrieval fields for both the 24- and 3-h cases.

e. *Impact of the length of assimilation interval*

Li et al. (1993) carried out a set of VDA experiments with the BSHB shallow-water model to investigate the impact of varying the length of the assimilation window. Their results suggest that as the length of the window increases, the number of minimization iterations will increase for the same prescribed convergence criterion. They concluded that the effort expended in minimizing the cost function over the window of assimilation goes mainly to balancing small-scale motions. These results are consistent with those obtained by Thépaut and Courtier (1991) and Navon et al. (1992).

To see whether their conclusion is also valid for the present 3D SLSI model, we carried out a number of similar 4D VDA experiments to check the impact of the length of assimilation interval on the rate of convergence of the cost function minimization.

We used the IC_C initial condition and assimilation intervals of lengths 1, 3, and 6 h. The model-generated observations were available at every time step in the whole assimilation window. A scaling of the cost function and its gradient was included in these experiments with $\mathbf{L}_0 = 30\mathbf{I}$.

Figure 16 displays the evolution of the normalized cost function (J/J_0) with respect to the number of iterations. We found our results to be consistent with those of Li et al. (1993) and Navon et al. (1992): as the length of the assimilation interval window increases, the convergence rate of the minimization decreases. The minimization for the 1-h window resulted in much lower values than those obtained with the 3- and 6-h windows, for the same number of iterations. Normalized cost functions were employed throughout.

f. *Impact of the time distribution of the observations on the convergence rate*

In the above experiments, the observations were available at every time step in the window of assimilation. One may assume that if the time density of observations increases, the convergence rate will also increase. To check this idea, we carried out a number of 4D VDA experiments in which observations were available either in selected segments or in the whole of the assimilation window.

The length of the window was chosen to be 6 h and we included the gradient scaling. Apart from the beginning ($t = t_0$) and the end of the assimilation interval, observations were available at every time step in the segments between 0 and 0 h (i.e., no additional observations), 0 and 1 h, 0 and 3 h, and, finally, 0 and 6 h (which included the whole window of assimilation). The initial condition was taken to be the IC_C condition.

Figure 17 displays the evolution of the normalized cost function (J/J_0), with the number of iterations for the experiments, where observations were available in different selected segments of the window of assimilation, namely, between 0 and 0, 0 and 1, 0 and 3, and 0 and 6 h.

We see that, compared to the case where no additional observations were included (the 0–0-h case), the convergence rate clearly does increase in the case

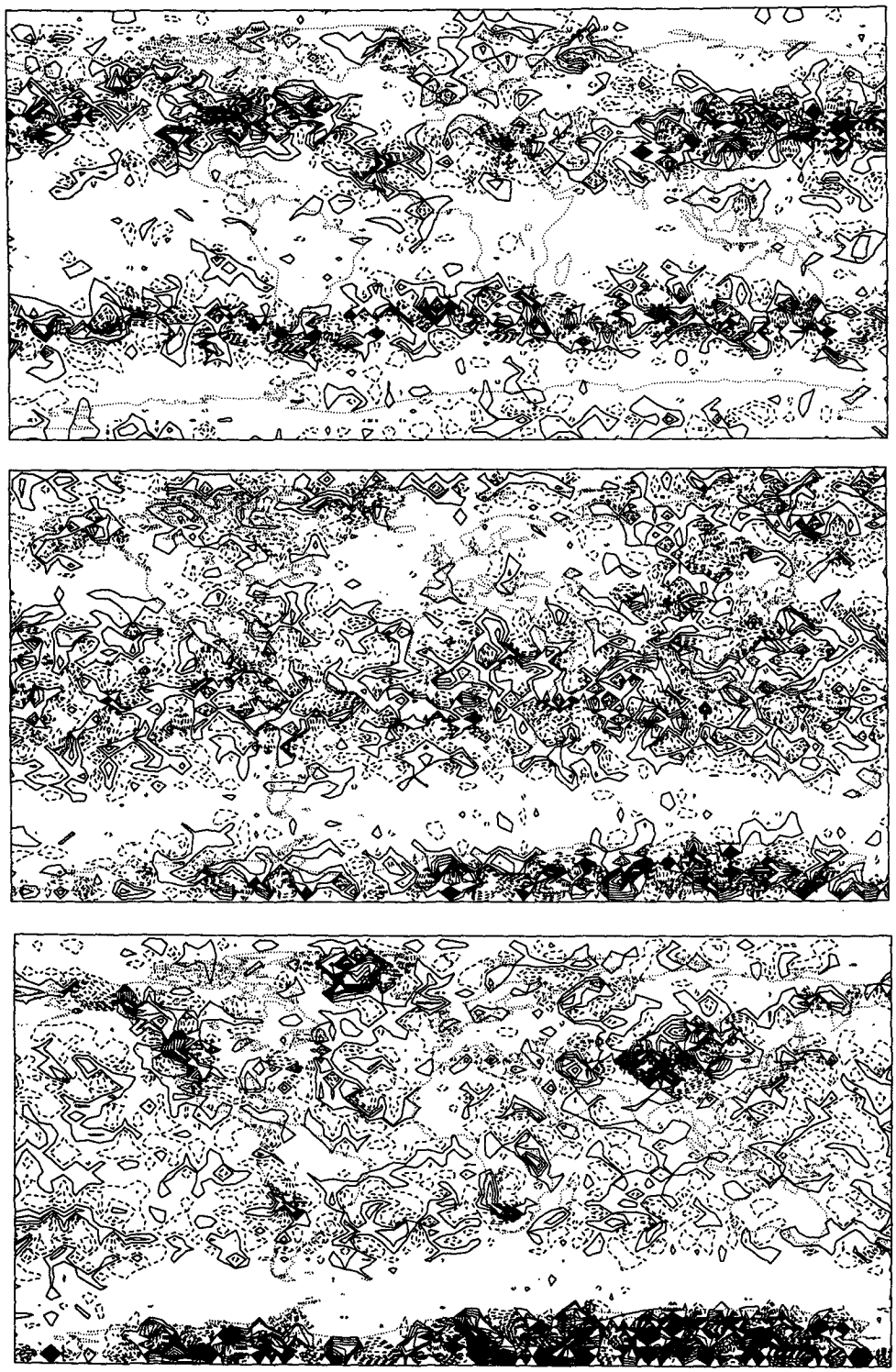

FIG. 11. Difference between the initial condition field and the corresponding observation field at 500 hPa, using the IC_B initial condition: $C_1 = 0.75$, $C_2 = 0.5$. (a) Zonal wind, contour interval is 1 m s^{-1}. (b) Temperature, contour interval is 1 K. (c) Surface pressure, contour interval is 4 hPa.

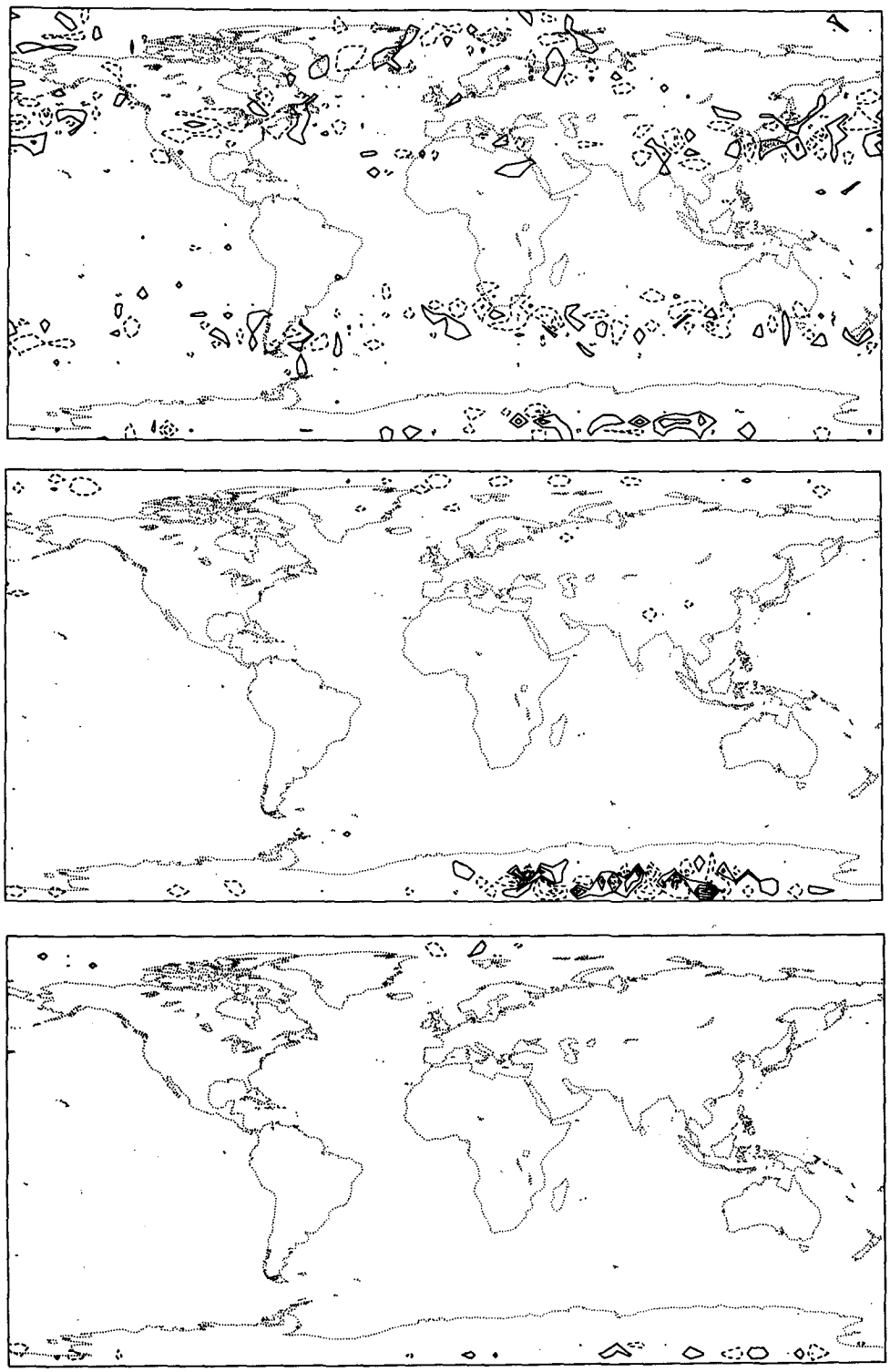

FIG. 12. Difference between the retrieval field after 42 minimization iterations and the corresponding observation field at 500 hPa, using the IC_B initial condition: $C_1 = 0.75$, $C_2 = 0.5$. Including scaling of the gradient of the cost function. (a) Zonal wind, contour interval is 1 m s^{-1}. (b) Temperature, contour interval is 1 K. (c) Surface pressure, contour interval is 4 hPa.

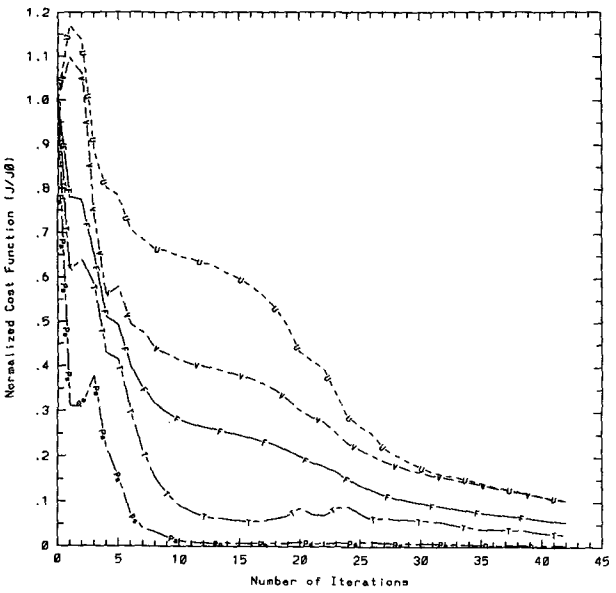

FIG. 13. Variation of each normalized component of the cost function as well as the total cost function versus the number of iterations, using IC_B initial condition: $C_1 = 0.75$, $C_2 = 0.5$, including scaling of the gradient of the cost function.

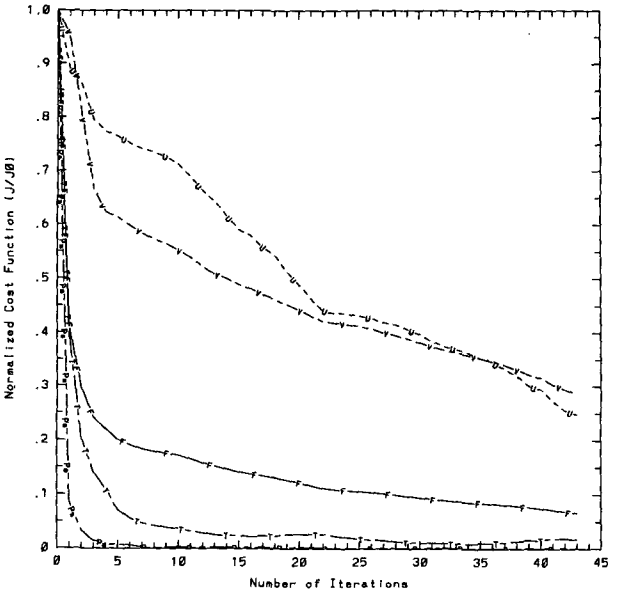

FIG. 15. Variation of each normalized component of the cost function as well as the total cost function versus the number of iterations, using IC_B initial condition. $C_1 = 0.75$, $C_2 = 0.5$. Including scaling of the gradient of the cost function. The assimilation window is 24 h.

where additional observations were added for the segment between 0 and 1 h. However, it decreases in the experiments where additional observations were available between 0–3 and 0–6 h (the whole window of assimilation). The convergence rate in the experiment where observations are available for the entire window of assimilation (0–6 h) is larger than that for the corresponding 0–3-h case. The retrieval initial condition fields also show the same features (figures omitted).

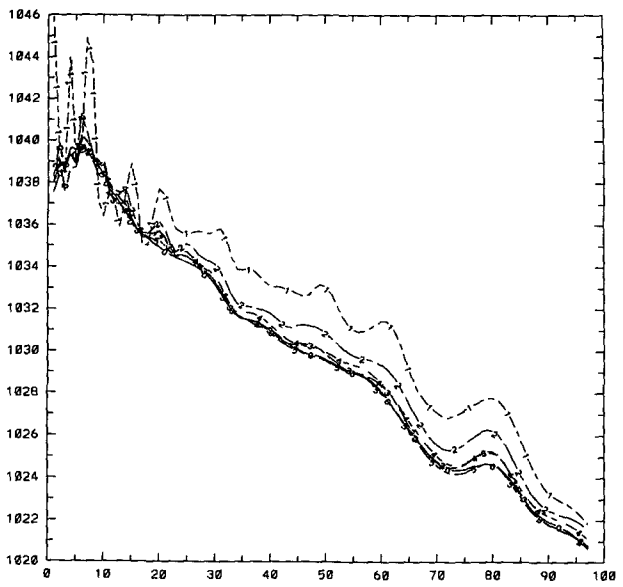

FIG. 14. Time variations of the surface pressure at a chosen grid point from initial observations (marked by 0), the initial guess (marked by 1), and the retrieval after 10 minimization iterations (marked by 2), 20 minimization iterations (marked by 3), 30 minimization iterations (marked by 4), and 40 minimization iterations (marked by 5), respectively.

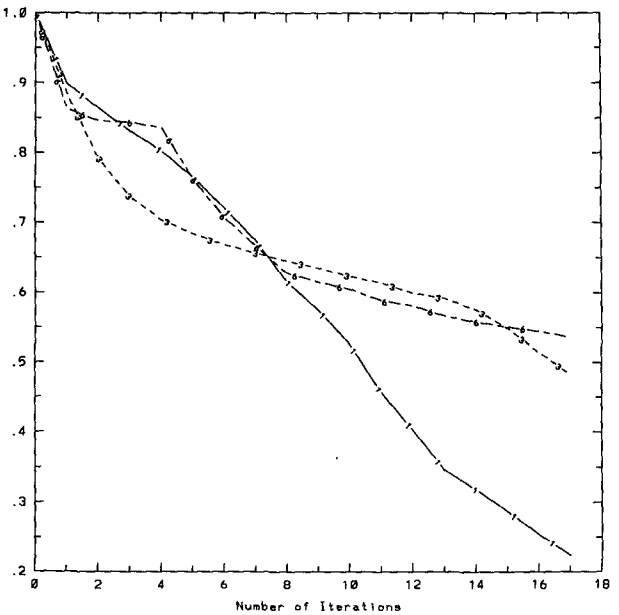

FIG. 16. Variation of the normalized cost function (J/J_0) versus the number of iterations. The lengths of the assimilation intervals are 1, 3, and 6 h, respectively, using IC_C initial condition.

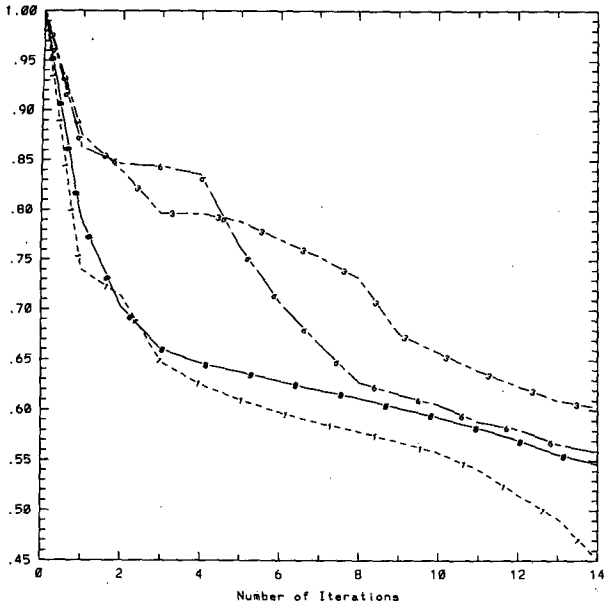

FIG. 17. Variation of the normalized cost function (J/J_0) versus the number of iterations for a window of assimilation whose length is 6 h. Observations are added at each time step in the assimilation window for the segments between 0 and 0 h, 0 and 1 h, 0 and 3 h, and 0 and 6 h, respectively, using IC_C initial condition.

These results suggest that the convergence rate of VDA minimization cannot increase simply as the time density of observations increases. There appears to exist a preferential way of adding observations into the assimilation window that can lead to a faster convergence rate.

6. Summary and conclusions

This paper presents the development of the adjoint of the adiabatic version of the 3D NASA/GLA semi-Lagrangian semi-implicit GCM developed by Bates et al. (1993), together with the result of several 4D VDA experiments using the adjoint. These experiments involved various types of first-guess initial conditions and weight matrices, as well as tests with or without scaling of the cost function and its gradient. This is to the best of our knowledge the first adjoint model developed for a 3D semi-Lagrangian GCM.

The tangent linear model and its adjoint for the BMH model were derived, and their correctness was verified. The 4D VDA experiments with model-generated observations performed well, as judged by the satisfactory quality of the retrievals obtained even with intensely randomized perturbations. Two different types of weight matrices were tested in the cost function, one invariant with respect to vertical level, the other using weights varying with level. We found the impact of changing the weight matrix in the cost function to be rather insignificant.

Our experiments also reveal that scaling of the cost function and its gradient is very effective in improving the convergence rate of the minimization process.

An initial condition with randomized perturbations was used to test the performance of the 4D VDA. We found that a satisfactory retrieval was obtained with a fast convergence rate of the minimization process.

Comparing the results obtained by stopping the iteration procedure at various stages of the minimization, we see that the variational data assimilation adjusts most of the large amplitude perturbations in the first 10–20 iterations. During the subsequent iterations, only small-scale small-amplitude features are being assimilated. This yields information as to when the minimization procedure could be stopped economically without affecting the overall quality of the retrieval. This can have a crucial impact on computational efficiency and hence on operational implementation.

Another set of experiments was carried out to investigate the effects of varying the length of the assimilation window. Our results confirm those of Navon et al. (1992) and Li et al. (1993).

Finally, a set of experiments was carried out to check the effects of the time density of the observations on the convergence rate of the minimization process. The experiments were designed with additional observations available in either a segment or the whole of the assimilation window. The results suggest that the convergence rate of VDA minimization does not simply increase as the number of observations increases in a segment of the window. Rather, we conclude that by adding observations only into preferential segments of the window one can obtain a faster convergence rate. Choosing an appropriate time distribution of observations in the VDA process turns out to be an important issue in 4D VDA applications. This deserves further research, as does the issue of choosing proper time weighting of the observations in the assimilation window.

The availability of the adjoint model of the SLSI GCM provides an important tool for further research in variational data assimilation with real observations. The preliminary numerical results reported here demonstrate the feasibility of using a semi-Lagrangian GCM with large time steps for efficiently carrying out 4D VDA.

Acknowledgments. The research of the first three 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 The Florida State University, which is partially funded by the Department of Energy through Contract DE-FC0583ER250000.

The authors would like to acknowledge the beneficial comprehensive comments of Prof. Olivier Talagrand from the Laboratoire de Meteorologie Dynamique du C.N.R.S., Ecole Normale Superieure, Paris, France, on a first draft of this paper.

REFERENCES

Arakawa, A., and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods in Computational Physics,* Vol. 17, *General Circulation Models of the Atmosphere,* J. Chang, Ed., Academic Press, 173–265.

Axelsson, O., and V. A. Barker, 1984: *Finite element solution of boundary value problems. Theory and computation.* Academic Press, 432 pp.

Bates, J. R., and A. McDonald, 1982: Multiply-upstream, semi-Lagrangian advective schemes: Analysis and application to a multi-level 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,* European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, U. K., 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.

Ehrendorfer, M., 1992: Four-dimensional data assimilation: Comparison of variational and sequential algorithms. *Quart. J. Roy. Meteor. Soc.,* **118,** 673–713.

Gill, P. E., W. Murray, and M. H. Wright, 1981: *Practical Optimization.* Academic Press, 401 pp.

Hortal, M., and A. J. Simmons, 1991: Use of reduced Gaussian grids in spectral models. *Mon. Wea. Rev.,* **119,** 1057–1074.

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

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., 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, 56 pp.

——, ——, P. Courtier, and P. Gauthier, 1993: Variational data assimilation with a semi-Lagrangian semi-implicit global shallow water equations model and its adjoint. *Mon. Wea. Rev.,* **121,** 1759–1769.

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

Lorenz, E. N., 1960: Energy and numerical weather prediction. *Tellus,* **12,** 364–373.

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 separable 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 invariants 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, WMO, 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.

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

Rasch, P. J., and D. L. Williamson, 1990: On shape-preserving interpolation and semi-Lagrangian transport. *SIAM J. Sci. Stat. Comput.,* **11,** 656–687.

——, and ——, 1991: The sensitivity of a general circulation model climate to the moisture transport formulation. *J. Geophys. Res.,* **96,** 13 123–13 137.

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

——, 1991: Application of the semi-Lagrangian method to a multilevel spectral primitive-equations model. *Quart. J. Roy. Meteor. Soc.,* **117,** 91–106.

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.

Simmons, A. J., 1991: Development of a high-resolution semi-Lagrangian version of the ECMWF forecast model. Numerical methods in atmospheric models, Vol. II, 281–324. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]

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

Temperton, C., 1991: A new high-resolution global spectral model for medium-range numerical weather prediction. *Proc. IMACS '91 13th World Congress on Computation and Applied Mathematics,* Dublin, 572 pp.

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.

——, D. Vasiljevic, and P. Courtier, 1993: Variational assimilation of conventional meteorological observations with a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.,* **119,** 153–186.

Wiin-Nielsen, A., 1959: On the application of trajectory methods in numerical forecasting. *Tellus,* **11,** 180–196.

Williamson, D. L., and P. J. Rasch, 1989: Two-dimensional semi-Lagrangian transport with shape preserving interpolation. *Mon. Wea. Rev.,* **117,** 102–129.

Zou, X., I. M. Navon, and J. Sela, 1993: Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus,* **45A,** 370–387.

Zupanski, M., 1993: A preconditioning algorithm for large scale minimization problems. *Tellus,* **45A,** 478–492.