Control of Gravitational Oscillations in Variational Data Assimilation

X. ZOU

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

I. M. NAVON

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

J. SELA

NOAA, NMC, World Weather Building, Washington, D.C.

(Manuscript received 2 December 1991, in final form 29 April 1992)

ABSTRACT

Variational four-dimensional data assimilation, combined with a penalty method constraining time derivatives of the surface pressure, the divergence, and the gravity-wave components is implemented on an adiabatic version of the National Meteorological Center's 18-level primitive equation spectral model with surface drag and horizontal diffusion. Experiments combining the Machenhauer nonlinear normal-mode initialization procedure and its adjoint with the variational data assimilation are also presented. The modified variational data-assimilation schemes are tested to assess how well they control gravity-wave oscillations.

The gradient of a penalized cost function can be obtained by a single integration of the adjoint model. A detailed derivation of the gradient calculation of different penalized cost functions is presented, which is not restricted to a specific model.

Numerical results indicate that the inclusion of penalty terms into the cost function will change the model solution as desired. The advantages of the use of simple penalty terms over penalty terms including the model normal modes results in a simplification of the procedure, allowing a more direct control over the model variables and the possibility of using weak constraints to eliminate the high-frequency gravity-wave oscillations. This approach does not require direct information about the model normal modes. One of the encouraging results obtained is that the introduction of the penalty terms does not slow the convergence rate of the minimization process.

1. Introduction

There is currently an increasing interest in various aspects, both applied and theoretical, of variational data assimilation. Encouraging success in the application of variational data assimilation involving simple 2D models are due, for instance, to Courtier (1984), Lewis and Derber (1985), LeDimet and Talagrand (1986), and Courtier and Talagrand (1990). More realistic experiments with 3D models are due to Thépaut and Courtier (1991), Navon et al. (1992), and Chao and Chang (1992). These research efforts employ optimal control methods, using initial conditions as control variables, and attempt to obtain optimal initial conditions by reducing a quadratic measure of lack of fit between a model's forecast and distributed data in some time window. The aforementioned results indicate that adjoint variational data assimilation is one of the most versatile tools for assimilating data distributed in space and time.

Variational data assimilation imposes no limitation on the characteristics of the data to be assimilated, and allows, in principle, the incorporation of all kind of data: humidity, wind strength and direction, vorticity, temperature measured by meteorological stations, satellite-measured radiance, etc. It is well known that in a forecast with the primitive equations, high-frequency gravity-wave "noise" will result if the initial fields are not suitably adjusted. Due to the high frequency and small amplitude of some gravity waves, they are difficult to predict and may distort the prediction of the more important meteorological features through nonlinear interactions. Therefore, a natural problem arising in the advance of variational data assimilation is how we will be able to filter the gravitational oscillations contained in the data being assimilated in the framework of variational data assimilation.

The issue of controlling gravity oscillations has a long history of importance in initialization. The need for initialization was first observed in Richardson's fa-

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

© 1993 American Meteorological Society

mous experiment where he attempted to use the primitive equations for producing a numerical forecast of weather in Europe. He lacked sufficient data for his model initial conditions. The errors in the initial conditions resulted in huge dynamic tendencies. Essentially, a portion of the errors in the initial conditions are interpreted by the model as due to the presence of unrealistic inertia–gravitational waves that subsequently propagate throughout the forecast and appear as gravitational "noise." Some appropriate initialization schemes for using the primitive equations without unrealistic noise have become desirable since that time.

Hinkelmann (1951) was the first to have shown that the amplitude of these unwanted high-frequency waves can be reduced by simply using geostrophic values for the initial wind field. Later, Charney (1955) showed that better results can be obtained when the initial wind and geopotential fields are related by the balance equation. Phillips (1960) showed that gravitational noise could be further reduced within forecasts that used the primitive equations by constraining the initial condition to satisfy a balance quasigeostrophic omega equation. It was found, however, that quasigeostrophic approximations also altered the prediction of meteorologically significant features. Subsequently, more effective procedures were designed to limit gravity-wave activity, and these procedures were based on nonlinear normal-mode initialization (NNMI). Machenhauer (1977) and Baer (1977) were the first to describe NNMI. Machenhauer considered the prognostic equations for amplitudes of gravitational modes. He showed that the adiabatic nonlinear forcing term has a strong, slowly varying component. This yields a correspondingly slow response, which approximately satisfies a nonlinear balance equation. He showed also how a solution to this new balance condition could be determined and applied to the initialization problem. Machenhauer's work was then extended by Baer and Tribbia (1977). Machenhauer's scheme was used predominately, while the scheme of Baer and Tribbia provided a more suitable theoretical framework for many problems.

Many other initialization methods have been developed since the advent of NNMI. The use of NNMI, however, provides a benchmark with which to compare results and a theoretical framework with which to explain methodologies and reasons for success.

Our main concern in this paper is not to use initialization per se, but to find a simple and efficient way to control residual high-frequency gravitational noise in the solution of variational data assimilation, while assimilating data that might be contaminated by some observational errors. Some of the ideas, however, of initialization schemes can be combined with the variational data assimilation. In some sense, our approach is philosophically related to the one presented by Miyakoda et al. (1978) that drew on the work of dynamical initialization of Miyakoda and Moyer (1968). The NNMI technique can also be used as a benchmark here for comparing the efficiency of filtering gravitational oscillations by the present new method developed with variational data assimilation.

One of the first attempts to damp the amplitude of high-frequency gravity oscillations in a variational data-assimilation process was made by Courtier and Talagrand (1990) using a shallow-water model. They combined a penalty term and an NNMI procedure with a variational data-assimilation scheme. The penalty term required the vanishing of the time tendency of the gravity-mode component of the initial data. Courtier and Talagrand concluded that gravity-wave noise can be efficiently eliminated by adding such a penalty term to the cost function, and by introducing into the variational process a nonlinear normal-mode initialization algorithm and its adjoint.

In this paper, we present a variational data-assimilation scheme, using the National Meteorological Center's (NMC's) spectral multilevel primitive equation model with horizontal diffusion and surface drag (Sela 1980). Penalty terms, which do or do not depend on the model's normal modes, are added to the cost function. In such a variational data-assimilation scheme, formulated with one or several penalty terms, the retrieved initial state will (a) fit the data as closely as possible, and (b) satisfy approximately some imposed dynamical constraints that will control the high-frequency gravity-wave oscillations. In section 2, three types of penalty terms are described. The first two penalty terms represent constraints on the first time derivatives of the surface pressure and divergence. The third penalty term constrains the second time derivative of the surface pressure. The NNMI scheme and its adjoint operation applied to the NMC model are presented in section 3. A penalty term constraining the time tendency of gravity-wave components is also presented in section 3. A detailed derivation of the calculation of the gradient of different penalized cost functions using the adjoint model is presented in appendixes A, B, and C. The derivation does not depend on any specific numerical model. Discussion of the results of numerical experiments with each penalty term or in various combinations is presented in section 4. These experiments are designed to study the amount of residual gravity-wave oscillations contained in the retrieved meteorological fields, as well as the computational cost and convergence rate of the proposed procedure. Summary and conclusions are presented in section 5.

2. Application of penalty-function method

The ideas of gravity-wave control used in classical variational initialization (Sasaki 1970) can be incorporated into variational data assimilation by the addition of a quadratic penalty term to the cost function. For instance, in the framework of classical variational initialization, one way of suppressing the external

gravity waves is to adjust the wind field with the strong dynamical constraint requiring the integrated mass divergence $\nabla \cdot \ln p_s \mathbf{v}$ to vanish, where p_s is the surface pressure and $\mathbf{v}$ the horizontal wind (see Barker et al. 1977; Sasaki et al. 1979). The classical variational method thus attempts to minimize a Lagrangian function

$$L = J + \lambda \nabla \cdot (\ln p_s \mathbf{v}), \tag{2.1}$$

where J is the cost function measuring the lack of fit between the model and the analyzed data and λ is a vector of Lagrange multipliers. Setting the coefficients of the first variation of all the independent variables and the Lagrange multiplier to zero leads to a set of coupled Euler–Lagrange equations, usually of elliptic type, which can be solved iteratively (see Thacker and Long 1988).

Instead of solving the problem of the augmented Lagrangian function L defined in (2.1), we add a quadratic penalty term to the basic cost function J. The basic cost function is the weighted sum of squares of the differences between the model's solution and analyzed data in the assimilation window. The penalized cost function is defined as

$$J_f = J + r\left(\frac{\partial \ln \mathbf{p}_s}{\partial t}\right)^{\mathrm{T}}\left(\frac{\partial \ln \mathbf{p}_s}{\partial t}\right), \tag{2.2}$$

where $(\ )^{\mathrm{T}}$ represents the transpose and $\mathbf{r}$ the penalty parameter.

The penalty-function methods transform the basic optimization problem into an alternative formulation such that numerical solutions are sought by solving a sequence of unconstrained minimization problems.

There are several reasons for the appeal of the penalty-function formulation. First is that the sequential nature of the method allows a gradual approach to criticality of the constraints. This means that if evaluation of the cost function and the gradient of the constrained is computationally difficult, we can use coarse approximations during the early stages of optimization and when the unconstrained minima of the penalized function are far away from the optimum. Finer and more detailed analysis approximations will be used during the final stages of the approximation. Second is that the algorithms for unconstrained minimization of rather arbitrary functions are well studied and generally quite reliable.

It is shown in optimization theory (Gill et al. 1981) that the problem of solving a constrained minimization problem via the quadratic penalty-function approach is equivalent, in the limit of large penalty parameters, to the Lagrangian problem if one chooses the multiplier

$$\lambda(r) = 2r\frac{\partial \ln \mathbf{p}_s}{\partial t}[x^*(r)], \tag{2.3}$$

where $x^*(r)$ is the unconstrained minimum of (2.2). This means that associated with every penalty parameter r there is a Lagrange multiplier vector that is determined after the unconstrained minimization is performed.

a. First-order time-tendency constraints

High-frequency fluctuations in the surface pressure tendency reflect the presence of external gravity waves. A way to suppress these external gravity waves is to impose as a constraint the vanishing of $\partial \ln p_s/\partial t$. A quadratic penalty term of the form

$$J_{p1} = r_1 \sum_{i=0}^{R-1}\left[\frac{\partial \ln \mathbf{p}_s(t_i)}{\partial t}\right]^{\mathrm{T}}\left[\frac{\partial \ln \mathbf{p}_s(t_i)}{\partial t}\right] \tag{2.4}$$

thus augmented the cost function J, where $\mathbf{p}_s$ is an M-dimensional vector, M is the total number of Gaussian grid points used in the definition of the cost function, and R is the number of time steps in the time interval spanning the window of assimilation. Variational data assimilation with a penalty term J_{p1} included in the cost function means that while minimizing the distance between the model solution and data, the high-frequency oscillations presented in the surface pressure field are also controlled.

We may also impose a divergence constraint to damp high-frequency oscillations of internal gravity waves in a manner similar to (2.4). We define a penalty term

$$J_{p2} = r_2 \sum_{i=0}^{R-1}\left[\frac{\partial \mathbf{D}(t_i)}{\partial t}\right]^{\mathrm{T}}\left[\frac{\partial \mathbf{D}(t_i)}{\partial t}\right], \tag{2.5}$$

where $\mathbf{D}$ is the $M \times K$ dimensional vector of divergence and K is the total number of vertical levels.

The penalized cost function can assume any of the following form

$$J + J_{p1}, \tag{2.6a}$$

$$J + J_{p2}, \tag{2.6b}$$

$$J + J_{p1} + J_{p2}. \tag{2.6c}$$

The penalty term J_{p1} (J_{p2}) can also be viewed as an example of the explicit inclusion of dynamical residuals $\partial \ln \mathbf{p}_s/\partial t(\partial D/\partial t)$ in variational assimilation (Bennett and Miller 1991). It is known from their theoretical conclusions that the model residual vanishes at time t_0.

Before assessing the performance of the quadratic penalty method in damping the high-frequency gravity waves present in the initial fields, we require the calculation of the gradient of the penalized cost functions. From (2.6), we have

$$\nabla(J + J_{p1}) = \nabla J + \nabla J_{p1}, \tag{2.7a}$$

$$\nabla(J + J_{p2}) = \nabla J + \nabla J_{p2}, \tag{2.7b}$$

$$\nabla(J + J_{p1} + J_{p2}) = \nabla J + \nabla J_{p1} + \nabla J_{p2}. \tag{2.7c}$$

Therefore, the gradient of a penalized cost function is equal to the gradient of the nonpenalized cost function

plus the gradient of the penalty terms. The calculation of ∇J by integrating the adjoint model backward in time is described in detail by Navon et al. (1992). We need only to derive the evaluation of ∇J_{p1} and ∇J_{p2}. For a detailed derivation, see appendix A.

From the detailed derivation in appendix A, we found out that the gradient of the penalized cost function can be obtained by integrating the same adjoint model as the one used in obtaining the gradient of the nonpenalized cost function. The only difference is that more forcing terms are added to the right-hand side of the adjoint-equations model besides the weighted differences between the model solution and data (see Fig. 1). These forcing terms consist of the time differences of the penalized variables [see Eqs. (A.7)–(A.10)].

b. Second-order time-tendency constraint

In order to further damp out the high-frequency gravity-wave oscillations, second-time-derivative constraints may be imposed. Interest in higher-order balance schemes was first discussed by Hinkelmann (1951). Barker et al. (1977) proposed using higher-order time-tendency derivatives to achieve a better elimination of Lamb waves.

In the context of NNMI, it was proposed that instead of considering $dG/dt = 0$, we use the condition

$$\frac{d^nG}{dt^n} = 0$$

for some $n > 1$ or some estimate of d^nG/dt^n $(n > 1)$

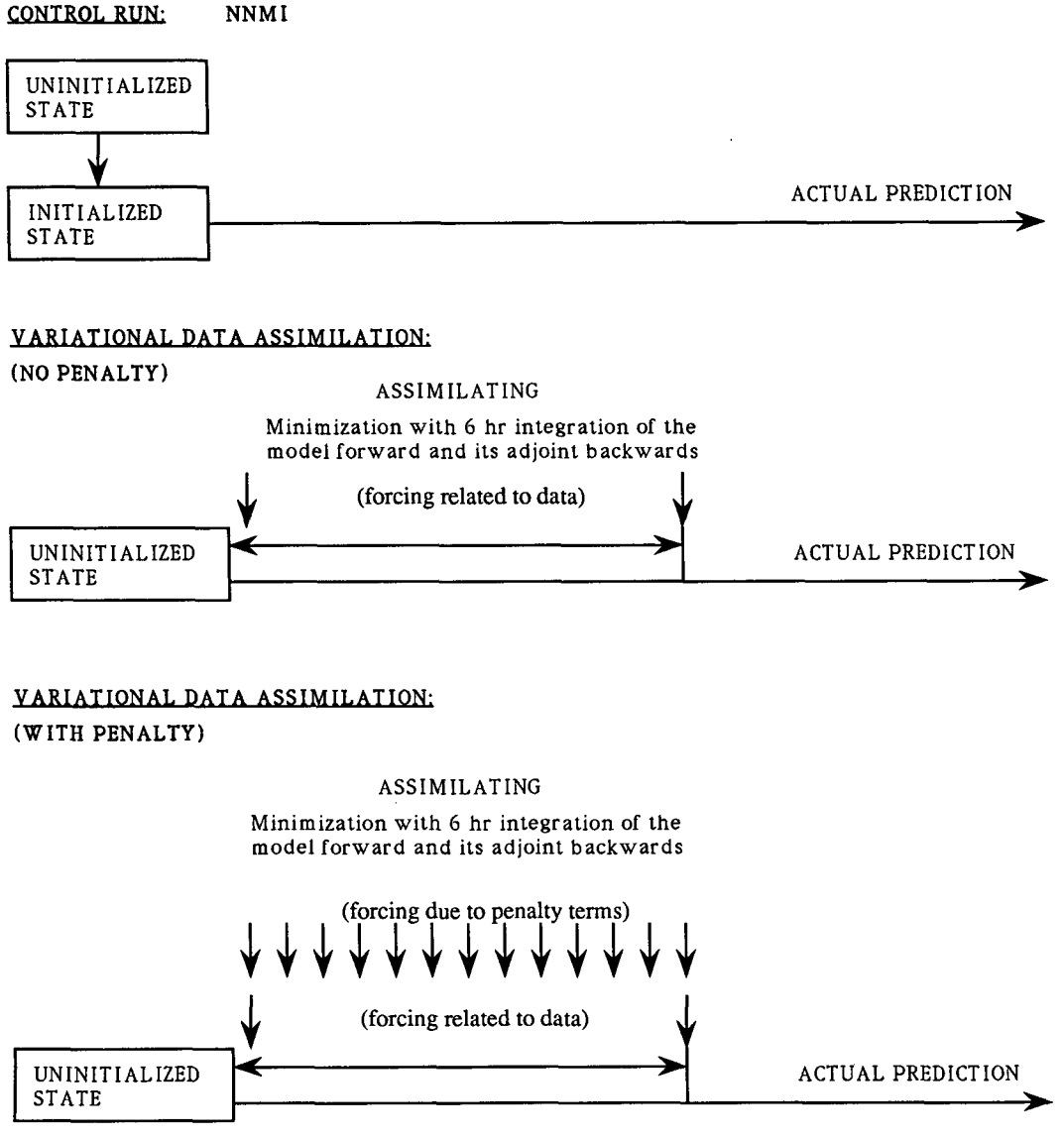

FIG. 1. A schematic diagram of the assimilation–forecast cycle.

based on slowly propagating components only, where G is the gravity-wave component of the flow. Such schemes were discussed by Baer and Tribbia (1977), Machenhauer (1982), Tribbia (1984), and Errico (1989b). See also Errico's survey of NNMI schemes (1989a).

In practical numerical weather prediction, however, in most cases higher-order balance schemes change subsequent forecasts very little with respect to other analysis or model errors because order ϵ^n adjustments are small for small ϵ (Rossby number) and large n (see Errico 1989a).

It is not difficult to find a situation when the integrated mass divergence is zero, and yet external gravity waves may still form. This situation is most evident when the initial data for the model are assimilated with a steep terrain and a static atmosphere (Barker et al. 1977). To delay and reduce the formation of external gravity waves, the initial fields may be chosen so that the tendency of the integrated mass divergence is also zero. In order to achieve this condition, we may add an additional quadratic penalty term of the form

$$J_{p3} = r_3\left(\frac{\partial^2 \ln p_s}{\partial t^2}\right)^{\mathrm{T}} \frac{\partial^2 \ln p_s}{\partial t^2} \tag{2.8}$$

to the cost function in the framework of variational data assimilation. Using a second-order finite-difference approximation for the second time derivative, J_{p3} assumes the form

$$J_{p3} = \frac{r_3}{(\Delta t)^4} \sum_{i=1}^{R-1} [\mathbf{P}(t_i)]^{\mathrm{T}} \mathbf{P}(t_i), \tag{2.9}$$

where

$$\mathbf{P}(t_i) = \ln p_s(t_i + \Delta t) + \ln p_s(t_i + \Delta t) - 2 \ln p_s(t_i). \tag{2.10}$$

The derivation of the evaluation of the gradient of the penalty term J_{p3}, which was presented in appendix B, is similar to the calculation of ∇J_{p1}. Only the forcing terms added to the right-hand side of adjoint-equations model are different.

3. Inclusion of the NNMI and the penalty term $r_4\|dz/dt\|^2$

a. NNMI and its adjoint

The explicit Machenhauer nonlinear NNMI (Sela 1982) is introduced at each step of the minimization before the forward integration of the model, and the fields before the initialization are taken to be the control variables. At each iteration of the minimization procedure, we first apply the NNMI procedure to the initial fields and integrate the direct model to obtain the value of the cost function and the value of the forcing terms. We then integrate the adjoint model backward in time, and finally carry out the adjoint operation of the explicit NNMI procedure to obtain the gradient of the cost function.

The Machenhauer NNMI is an iterative procedure that was implemented in the form of

$$\mathbf{x}_k(t_0) = \mathbf{x}_{k-1}(t_0) - \Delta\mathbf{x}_{k-1}(t_0)$$
$$\Delta\mathbf{x}_{k-1}(t_0) = \mathbf{N}\mathbf{x}_{k-1}(t_0), \quad k = 1, 2, \cdots, \tag{3.1}$$

where $\mathbf{N}$ represents the result of applying all the operator matrices in the NNMI code to obtain the correction term $\Delta\mathbf{x}_{k-1}(t_0)$ from $\mathbf{x}_{k-1}(t_0)$. Here, $\mathbf{x}$ is the vector of spectral coefficients of the state variable $\mathbf{X}$. This NNMI iterative algorithm is stopped after a finite number of iterations. The condition that the initial time tendency of the gravity component be zero cannot be exactly enforced in practice. It was experimentally determined at NMC that two Machenhauer iterations using the first four vertical modes produce acceptable changes in the initial conditions and result in very smooth surface pressure integrations.

The adjoint of the initialization scheme can thus be written as

$$\hat{\mathbf{x}}_{k-1}(t_0) = (\mathbf{I} - \mathbf{N}^{\mathrm{T}})\hat{\mathbf{x}}_k(t_0), \tag{3.2}$$

where $\hat{\mathbf{x}}$ represents the corresponding adjoint variable of $\mathbf{x}$. The code for the adjoint explicit NNMI was written by following the logic used in building the adjoint-model equations (Navon et al. 1992).

The evaluation of the gradient of the cost function when the NNMI procedure is included in the variational data assimilation is done in two steps. Step one the ordinary backward time integration of the adjoint model. Step 2 applies the adjoint of the NNMI operator to the output of the adjoint-model integration.

The NNMI operator has the property of a projection operator onto the slow manifold along a direction parallel to the gravity modes' subspace. This operator is partially invertible due to its being iterative (Courtier and Talagrand 1990; Thépaut and Courtier 1991).

The adjoint of the NNMI projects the gradient of the cost J on the Rossby modes along a direction parallel to the orthogonal of a tangent to the slow manifold. This should cause (in the absence of invertibility) the gradient of the cost function with respect to the gravity components to vanish (Thépaut and Courtier 1991). For a few iterations, the NNMI operator remains invertible under the adjoint operation, and increasing only slightly the number of Machenhauer iterations has the effect of delaying the stage at which the variational minimization process starts reconstructing the gravity-wave part of the state. One of the purposes of the present research is to evaluate this invertibility using the adiabatic version of the NMC operational model.

b. The penalty term $r_4\|dz/dt\|^2$ and its gradient

A state vector $\mathbf{x}$ is here represented by its gravity part $\mathbf{z}$ and its Rossby part $\mathbf{y}$. A penalty term of the form

$$J_{p4} = r_4 \left\| \frac{d\mathbf{z}(t_0)}{dt} \right\|^2, \tag{3.3}$$

similar to that implemented by Courtier and Talagrand (1990), has been added to the cost function J for comparison with these authors' results obtained using other penalty terms.

The norm of $d\mathbf{z}/dt$ in (3.3) is defined by an inner product $\langle \cdot, \cdot \rangle$, which can either be

$$\|\mathbf{x}\|_2^2 \equiv \langle \mathbf{x}, \mathbf{x} \rangle = \frac{1}{2} \sum [\chi^2 + \psi^2 + (\ln p_s)^2 + T^2], \tag{3.4}$$

or

$$\|\mathbf{x}\|_E^2 \equiv \langle \mathbf{x}, \mathbf{x} \rangle = \frac{1}{2} \sum \left[u^2 + v^2 + RT_r(\ln p_s)^2 + \frac{C_p}{T_r} T^2 \right] \frac{\partial p}{\partial \sigma} = \frac{1}{2} (\mathbf{WSHx})^{\mathrm{T}} \mathbf{WSHx}, \tag{3.5}$$

where

$$\mathbf{H} = \begin{pmatrix} \mathbf{I} & & \\ & \mathbf{I} & \\ & & \mathbf{F} \end{pmatrix}, \quad \mathbf{W} = \frac{\partial p}{\partial \sigma} \begin{pmatrix} RT_r\mathbf{I} & & \\ & C_p/T_r\mathbf{I} & \\ & & \mathbf{I} \end{pmatrix}.$$

Here, $\mathbf{F}$ is an operator that computes the spectral coefficients of the wind-field components (u,v) from the divergence and vorticity fields and $\mathbf{S}$ is a transform operator from spectral to grid space. The first inner product is an $\mathcal{L}_2$ norm, and the latter is an energy norm that is a quadratic invariant of the linearized primitive equations in the vicinity of a state of rest defined by a uniform surface pressure p_0 and a constant temperature T_r. As discussed by Thépaut and Courtier (1991), the choice of an energy form defines a physical measure of atmospheric fields as well as a natural weighting for the different variables of the gravity components.

This penalty process by (3.3) does not ensure the vanishing of the gravity modes' time tendencies in the subsequent tendency computation due to the nonlinear nature of the problem. The constraint on the time tendency of the gravity component can also be applied to all the time steps in the window of assimilation instead of being applied at only the initial time. Instead of (3.3), the following term can be added to the cost function J:

$$J_{p4m} = \frac{1}{2} r_4 \sum_{i=0}^{R} \left\| \frac{d\mathbf{z}(t_i)}{dt} \right\|^2, \tag{3.6}$$

where $t_i = t_0 + i\Delta t$, $\Delta t = 1800$ s.

Implementation of the minimization of $J + J_{p4}$ or $J + J_{p4}$ requires the explicit computation of the gradient of J_{p4} or J_{p4m} with respect to the initial model fields. Thus, we presented in appendix C the derivation of ∇J_{p4} and ∇J_{p4m}. It was concluded that the process of calculating the gradient of the penalty term J_{p4} consists mainly of a projection of the time tendency of the initial state on the gravity manifold and the adjoint operation of this projection [see Eq. (C.10)]. The calculation of ∇J_{p4m} can be obtained by a single integration of the adjoint model, where forcing terms (C.17) are added to the right-hand side of the adjoint-equations model at each time step.

4. Results

Three previously mentioned types of penalty terms have been used, first independently and then in combination, in order to reduce the amount of gravity-wave activity present in the assimilated fields.

Initial conditions for the model and analyzed data were obtained from the global data-assimilation scheme at NMC. The variational data assimilation was carried out adiabatically, as well as with a surface drag and horizontal diffusion terms included in the model. Two uninitialized state vectors, separated by a 6-h time interval, were taken as analyzed data to be assimilated. The cost function was defined as

$$J[\mathbf{X}(t_0)] = \frac{1}{2} [\mathbf{X}(t_0) - \mathbf{X}^{\mathrm{obs}}(t_0)]^{\mathrm{T}} \mathbf{W} \times [\mathbf{X}(t_0) - \mathbf{X}^{\mathrm{obs}}(t_0)] + \frac{1}{2} [\mathbf{X}(t_1) - \mathbf{X}^{\mathrm{obs}}(t_1)]^{\mathrm{T}} \times \mathbf{W}[\mathbf{X}(t_1) - \mathbf{X}^{\mathrm{obs}}(t_1)], \tag{4.1}$$

where $\mathbf{X}$ is the state vector, and $\mathbf{W}$ is a diagonal weighting matrix whose values are calculated as the inverse of the maximum squared difference between the two analyzed data fields at times t_0 and t_1. Such a weighting renders the different terms in the cost function to be of the same order of magnitude, which will result in a better convergence rate of the minimization process. The time difference between the two analyzed times is 6 h. A limited-memory quasi-Newton method (Navon and Legler 1987) due to Liu and Nocedal (1989) was employed throughout the minimization experiments. This method has proved to be very efficient and robust for large-scale unconstrained minimization.

The testing of the accuracy of the adjoint model was carried out using the following identity check that applies at the level of each subroutine:

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

where $\mathbf{A}$ represents the tangent linear model, $\mathbf{A}^{\mathrm{T}}$ the corresponding adjoint model, and $\mathbf{Q}$ and $\mathbf{AQ}$ represent the input and output of the linear mode, respectively. Since the penalization of the cost function does not alter the adjoint model, we did not include a further discussion on the accuracy of adjoint model. Due to the presence of the additional forcing terms resulting from the penalty terms, only the calculation of the gradient of the cost function needs to be rechecked.

The following expression is derived from a Taylor expansion of the form:

$$J_f[\mathbf{X}(t_0) + \alpha\nabla J_f] = J_f[\mathbf{X}(t_0)] + \alpha(\nabla J_f)^{\mathrm{T}}\nabla J_f + O(\alpha^2), \quad (4.2)$$

where J_f represents any of the penalized cost function. From (4.2) we obtain

$$\Phi(\alpha) \equiv \frac{J_f[\mathbf{X}(t_0) + \alpha\nabla J_f] - J_f[\mathbf{X}(t_0)]}{\alpha(\nabla J_f)^{\mathrm{T}}\nabla J_f} = 1 + O(\alpha). \quad (4.3)$$

Therefore, the Φ is defined in terms of α, the cost function and the gradient of the cost function.

If the values of the cost function and the gradient are correctly calculated, the value of $\Phi(\alpha)$ will linearly approach 1 with α decreasing in a certain range value of α covering several orders of magnitude. In this check, if we find that the value of Φ is linearly approaching a constant $C \neq 1$, it means that the gradient calculation is erroneous. The error occurs mainly in the data management in the direct model (write out) and the adjoint model (read in), which results in the added forcing being wrong. If there is no linearity in the variation of the value of Φ with decreasing values of α, then the errors are liable to have occurred in the calculation of the penalty terms.

Figure 2 shows the variation of the value of the function $\Phi(\alpha)$ with decreasing values of α. We note that for values of α between 10^{-7} and 10^{-15}, which are not too close to the machine zero, the relation

$$\Phi(\alpha) \approx 1$$

was verified for all the three cases. The correctness of the gradient calculation was therefore verified.

a. Penalty method with first-order constraints

The first-time-derivative constraints of (2.4) and (2.5) were tested, first independently, then jointly, and the rates of convergence of the penalization schemes were examined. Since the penalization scheme is trying to reduce the value of the sum of the squares of the time tendencies of the surface pressure (J_{p1}/r_1) or divergence (J_{p2}/r_2), the behavior of these variables as a function of the iteration number constitutes a useful measure of the convergence rate. Figure 3 shows the behavior of J_{p1}/r_1 for different values of the penalty parameter r_1 in the experiment minimizing $J + J_{p1}$. We note that when smaller values of the penalty parameter were used, a smaller decrease in J_{p1}/r_1 was observed. When the penalty parameter $r_1 \geq 10^{13}$, all the plot lines for J_{p1}/r_1 as a function of the iteration number coincide and a decrease of almost two orders of magnitude in its value was attained. Figure 4 shows the variation of J_{p2}/r_2, the constraint on the time tendency of divergence, with the number of iterations in the experiment minimizing the penalized cost $J + J_{p2}$ with $r_2 = 10^{19}$. A decrease of one and one-half orders

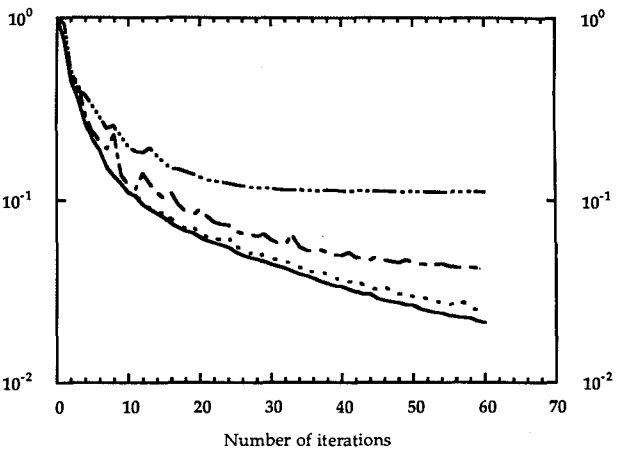

FIG. 3. The normalized value of J_{p1}/r_1 with the number of iterations. Dash-dot line: $r_1 = 10^{10}$. Dash-dot line: $r_1 = 10^{11}$. Dotted line: $r_1 = 10^{12}$. Solid line: $r_1 = 10^{13}$.

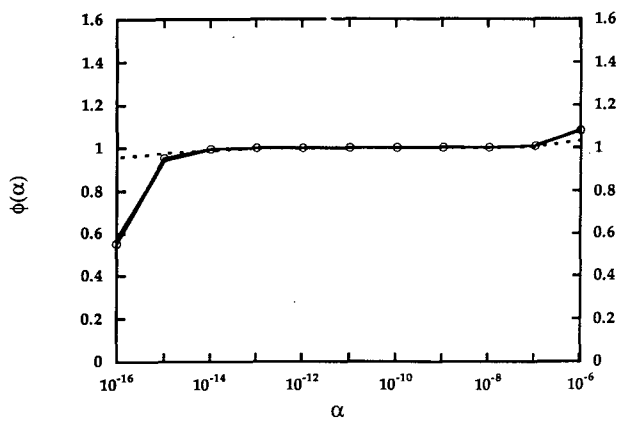

FIG. 2. Verification of the gradient of the cost function penalized by J_{p1} (solid line), J_{p2} (dotted line), and J_{p3} (circled line), respectively.

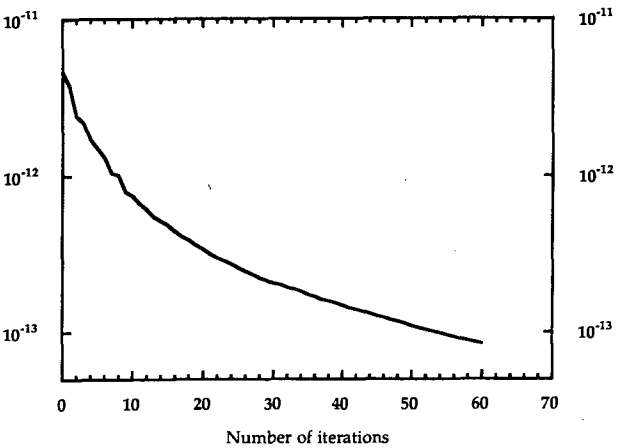

FIG. 4. The value of J_{p2}/r_2 as a function of the number of iterations.

of magnitude was obtained during the variational data-assimilation process.

In order to illustrate the performance of the penalty method, we present in Table 1 the amount of residual gravity-wave component in the retrieved initial state. The results obtained after application of NNMI provide a physical reference as to how much residual gravity-wave component should be present in the initial state. We see that the amount of residual gravity-wave component in the retrieved initial state after the minimization without inclusion of a penalty term is larger than that in the analyzed data. The solution of the minimization with both penalty terms J_{p1} and J_{p2} added to the cost function contains a lesser amount of residual gravity component than the one contained in the NNMI initialized data.

Figure 5 shows the time variation of the surface pressure at a point in the Indian Ocean (point 1) during 24-h integrations using the adiabatic version of the NMC model. The oscillating curve shows the result of an integration with the retrieved initial conditions without penalty term added to the cost function. The smooth curves show the results with penalized initial data. The nonpenalized forecast displays the presence of considerable amounts of high-frequency oscillations, with a peak-to-trough amplitude of as much as 7 hPa. When J is penalized by J_{p1}, the oscillations are reduced, and as the value of the penalty parameter increases, the amplitudes of the high-frequency oscillations decrease. The impact of increasing the penalty parameter r_1 beyond $r_1 = 10^{15}$ is almost imperceptible. Figure 6 shows the evolution of the surface pressure from retrieved initial conditions with J_{p1}, or J_{p2}, or both, added to the cost function J at point 1, and point 2, near the Rocky Mountains. The solid line in Fig. 6 represents the forecast without penalty terms, the dashed line with the first penalty J_{p1}, the dash-dot line with the second penalty term J_{p2}, and the dotted line with both the first and second penalty terms included in the cost function. At both points, the high-frequency oscillations contained in the analyzed data are mostly damped out after the penalized variational data assimilations are carried out. The minimization of the cost function $J + J_{p1} + J_{p2}$ with $r_1 = 10^{15}$ and $r_2 = 10^{19}$ yielded the strongest damping of the gravity-wave os-

TABLE 1. Energy norm of $dz(t_0)/dt$ contained in different retrieved initial states after the minimization.

Initial state	Penalty term	Iterations/ CPU (min)	Energy norm of $dz(t_0)/dt$
RETNOP		60/48	3.1541
	(3.3)	60/55	5.3506
	(3.3)	60/55	0.2156
	(energy norm)	60/109	0.0956
RETDGDT		20/51	0.2751
	(3.14)	30/76	0.1928
		60/146	0.0636
	(3.14) (energy norm)	60/141	0.4637
RETDIQ	(2.4) + (2.6)	60/61	0.0763
data1	observation at t_0		2.5618
data1.i	NNMI data 1		0.1533

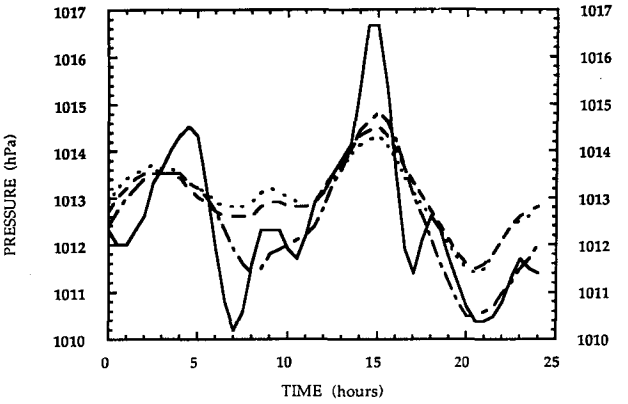

FIG. 5. Time traces of surface pressure at grid point 1. Solid line: no penalty. Dash-dot line: $r_1 = 10^{11}$. Dashed line: $r_1 = 10^{13}$. Dotted line: $r_1 = 10^{15}$.

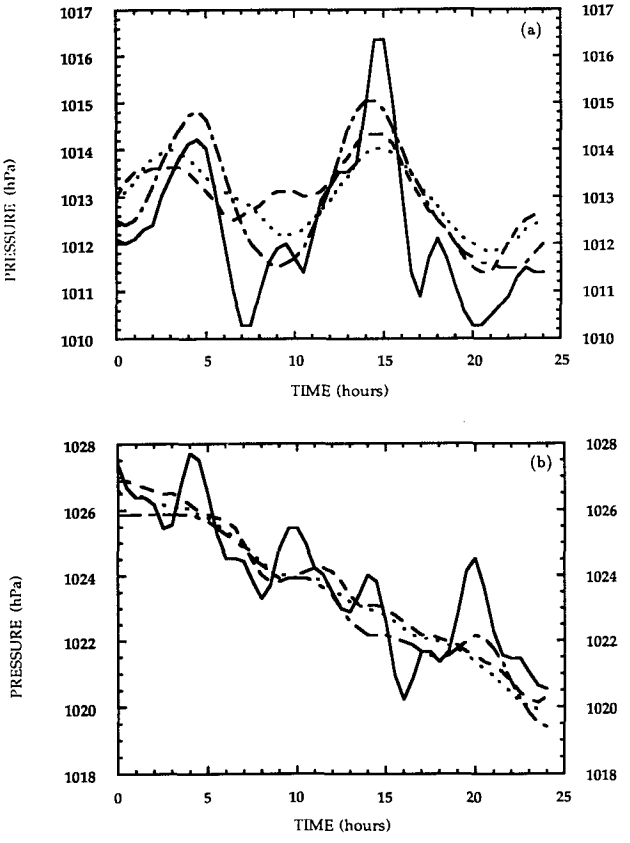

FIG. 6. Time traces of surface pressure at (a) point 1, and (b) point 2. Solid line: no penalty term. Dashed line: with the penalty term J_{p1}. Dash-dot line: with the penalty term J_{p2}. Dotted line: with both the penalty terms $J_{p1} + J_{p2}$.

cillations. Thus, the combination of both the constraints on surface pressure and divergence fields gives the best result for suppressing high-frequency gravity-wave oscillations.

Figure 7 is similar to Fig. 6 and displays the behavior of the divergence field. We note that large oscillations appeared in the integrations without penalty terms or with only the first penalty term included in the cost function. The forecast, with the second penalty term included in the cost function, yielded improved damping of the short-period oscillations. When the second penalty was combined with the first, high-frequency gravity-wave activity was satisfactorily controlled.

The rms of the divergence of the retrieved initial conditions with and without the inclusion of the penalty term is displayed in Fig. 8. The rms of the retrieved initial divergence field without any penalty term (solid line) is significantly reduced when penalty terms are included in the cost function. Again, the minimization with the penalty term J_{p2} for the divergence field included in the cost function, seems to yield the best result. The minimization with penalty on the surface pressure only, results in an improvement in the rms of the divergence only at the lower levels.

We examined the effect of the penalty method by studying maps of the divergence field without any penalty term, with only the first penalty term, with only the second penalty term, and with both the first and

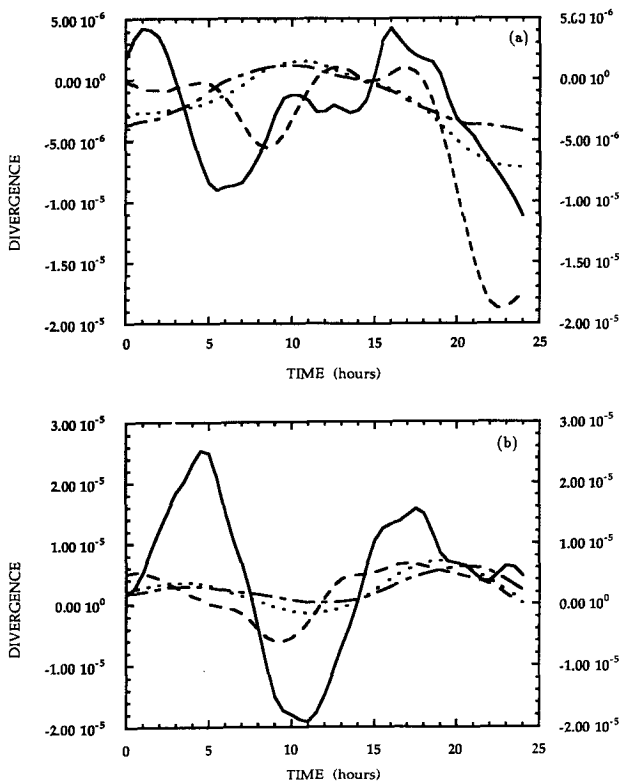

FIG. 7. As Fig. 5 for the divergence at the fifth model level.

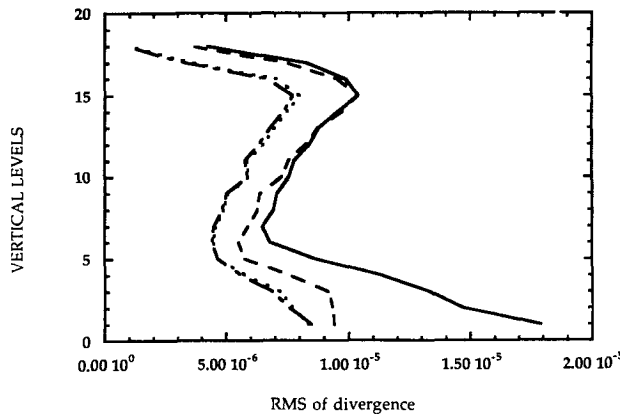

FIG. 8. Root-mean-square of the divergence at all vertical levels. Solid line: no penalty term. Dashed line: with the penalty term J_{p1}. Dash-dot line: with the penalty term J_{p2}. Dotted line: with both the penalty terms $J_{p1} + J_{p2}$.

second penalty terms included. The local changes of surface pressure are generally less than 3 hPa, an acceptable degree of adjustment. The divergence field obtained without a penalty term appears to be very disorganized and is generally too intense. The application of the penalty method results in a divergence field that has much more coherence and is synoptically reasonable, despite the neglect of diabatic effects.

To confirm that the Rossby modes are not degraded after the high-frequency gravity wave oscillations have been completely eliminated by the penalty procedure, and furthermore, that once these oscillations have been eliminated, they do not seem to develop, a closer inspection of the integration results was carried out. A 3-day forecast was made from the NNMI data and the retrieved initial states without a penalty term and with the two penalty terms (2.4) and (2.6), using the adiabatic version of the NMC spectral model with surface drag and horizontal diffusion. Figure 9 shows the time variation of the surface pressure at point 1 and point 2 during this period. It is seen that the value from the retrieved initial state without a penalty term oscillates around a slowly varying value from the retrieved initial state with penalty terms at both points (solid and dotted lines in Fig. 9). This fact indicates that the gravity oscillations present in the integration with the nonpenalized retrieval neither grow nor influence significantly the slowly varying part of the fields. This fact is confirmed by a comparison between the penalization result and NNMI result (dashed line in Fig. 9). The time variation of the surface pressure is nearly the same at point 2 for the results obtained from the retrieved initial state with penalty terms and from that obtained from the nonlinear normal-mode initialized state, with the former being even smoother than the latter. At point 1, the time variation of the penalty result is smooth and exhibits a 12-h oscillation. After NNMI, however, both the large-amplitude 12-h oscillations and the low-amplitude high-frequency oscillations are reduced. Af-

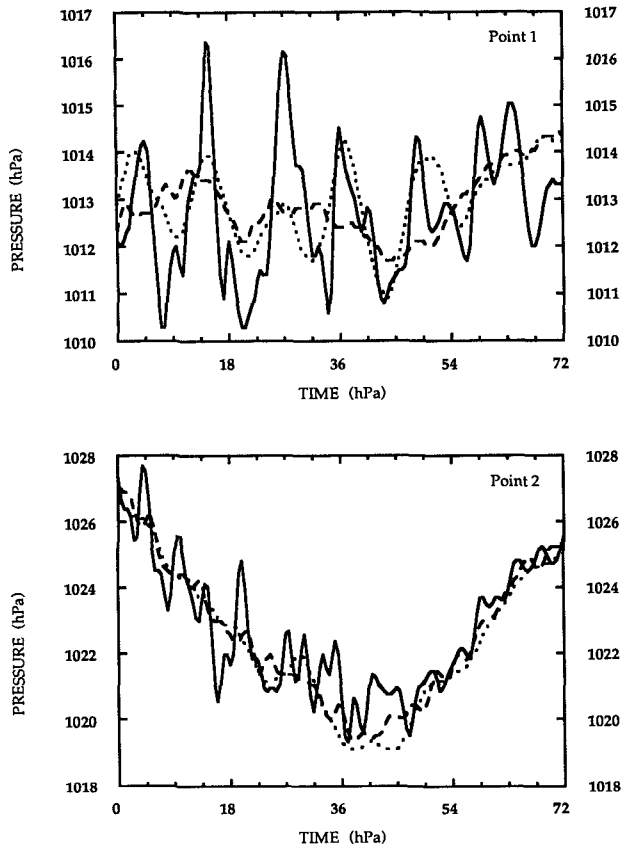

FIG. 9. Time variation of 3-day forecasts at point 1 and point 2. Solid line: no penalty term. Dotted line: with penalty terms $J_{p1} + J_{p2}$. Dashed line: NNMI.

ter two and one-half days, the time variation is nearly the same.

The dotted line in Fig. 10 shows the rms for the divergence field. It is seen that the value approaches a constant mean value that is smaller than the value of rms without penalty terms. Contours in Fig. 11 show the streamfunction at model level 9, which is near 500 hPa. We found that large-scale features remain the same with or without the inclusion of the penalty terms. This turns out to be also the case for the distributions of surface pressure, velocity potential, temperature, and divergence fields.

In view of the experimental results, we are led to the conclusion that the adequate penalization of a cost function is able to suppress undesirable high-frequency gravity-wave oscillations. The constraints of requiring the vanishing of the surface pressure and divergence at each time step were implemented weakly by the penalty method. The Rossby waves are not affected by the penalty procedure. It is also worth noting that no additional computational cost is required to calculate the values of the penalized cost function and its gradient, and the number of iterations and function evaluations is about the same as that required by the minimization without a penalty term.

It should be emphasized that in all the above experiments, the penalty terms were added to the cost function at each time step in the 6-h time assimilation window. To estimate the difference between this constraint and a constraint applied at the initial time only, we conducted an experiment minimizing the cost function

$$J + r'_1\left[\frac{\partial \ln p_s(t_0)}{\partial t}\right]^2.$$

The minimization was performed successfully, and a decrease in the penalty term of better than two orders of magnitude was achieved (after same number of iterations). A subsequent 24-h forecast from the retrieved initial state, however, reveals that gravity oscillations still persist. Apparently, to obtain a satisfactory damping of the high-frequency gravity-wave oscillations by applying the constraints only at the initial time would require more iterations. It also appears that the implementation of a penalty term requiring the vanishing of the time tendency of a field (either $\ln p_s$ or D) at every time step in the assimilation window allows us to overcome the slow convergence to a smooth solution of the penalty term requiring the vanishing of the time tendency at the initial time only. Moreover, no significant computational effort is required for the former case since the adjoint model is the same. The additional computational cost results from additional forcing terms during the adjoint integration.

b. Results using NNMI and the penalty $r_4\|dz/dt\|^2$

We now discuss the results obtained using NNMI and its adjoint in order to control gravity-wave oscillations in the forecast. In the experiments using this method, no penalty term was added to the cost function. An NNMI procedure was introduced at each iteration of minimization before the integration of the model. Using this method, the initial value of the time

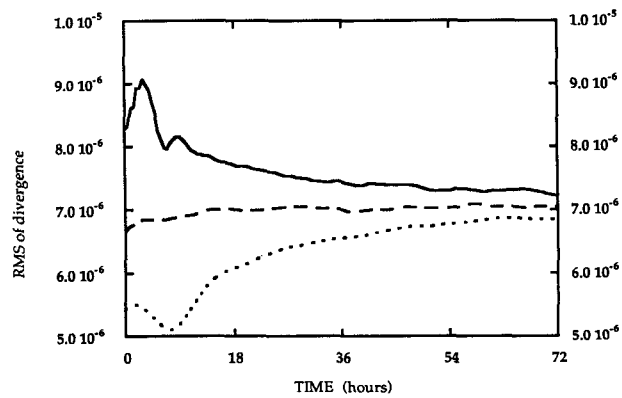

FIG. 10. Time variation of the total rms of the divergence field. Solid line: no penalty term. Dotted line: with penalty terms $J_{p1} + J_{p2}$. Dashed line: NNMI.

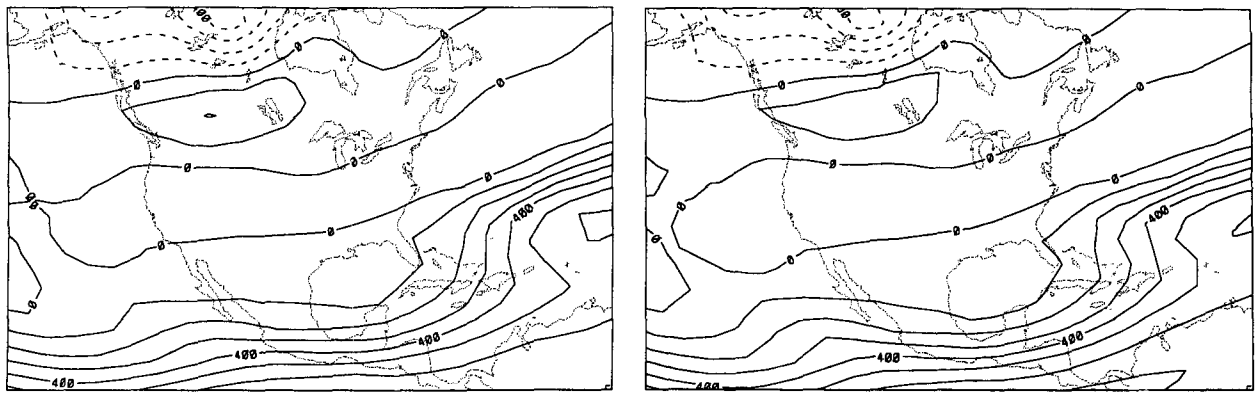

FIG. 11. Distribution of streamfunction at model level 9. (a) no penalty term; (b) with penalty terms $J_{p1} + J_{p2}$. Contour interval is 1.0×10^7 m^2 s^{-1} and values on the isolines are scaled by 10^{-5}.

tendency of the surface pressure is smaller than the uninitialized corresponding norm, since a considerable number of gravity waves were eliminated by the NNMI scheme. During the minimization, the amount of gravity-wave activity, as measured by J_{p1}/r_1, increases due to the application of the adjoint of the NNMI initialization algorithm (see Fig. 12). After 20 iterations, the process of minimization failed due to the presence of a large amount of high-frequency gravity noise. The explicit Machenhauer NNMI employed in the minimization does not seem to be able to damp out the gravity noise reintroduced in the gradient calculation due to the inclusion of the adjoint of NNMI scheme.

We present in Fig. 13 the variation of the cost function when no NNMI is applied before the forward integration of the model (solid line) and when the integration of the model is preceded by 1 (dash–dot line), 2 (dotted line), 3 and 4 (dashed line) iterations of the Machenhauer process. We note that when the NNMI process is not performed, the convergence rate is better. The solution matches the analysis at initial time extremely well, but contains a large number of gravity waves since the analyzed data being assimilated are uninitialized analysis. If we carry out the NNMI before the integration of the model, only the Rossby part of the initial field is reconstructed, and the minimum of the cost function becomes larger.

Another measure of the performance of the NNMI is the variation of the term $\| d\mathbf{z}(t_0)/dt \|_2^2$ as a function of the iteration number (see Fig. 14). This term represents the energy of the gravity-wave tendency at time t_0 for the first four vertical modes of the model's initial state. Henceforth, this energy will be denoted by BAL. Starting the minimization from the analysis at time t_0, BAL contains the same amount of gravity-wave energy at the beginning of the assimilation.

For two iterations of the NNMI (dash–dot line), BAL decreases during the first 4 minimization steps, then increases and reaches its initial value after 12 iterations. Slightly increasing the number of Machenhauer iterations seems to delay the step at which the

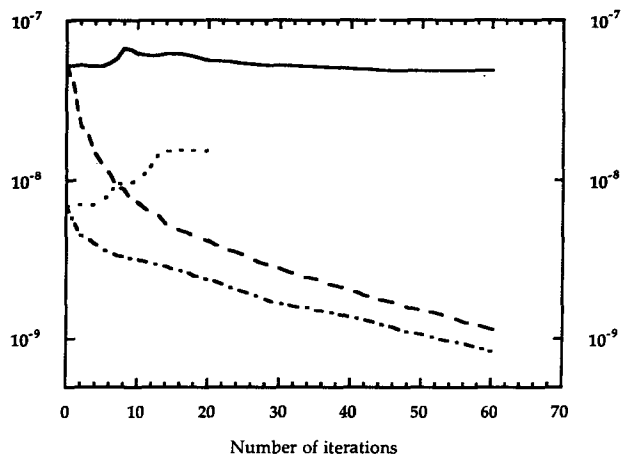

FIG. 12. The value of J_{p1}/r_1 as a function of the iteration number. Solid line: no penalty term. Dashed line: with only the penalty term J_{p1} ($r_1 = 10^{15}$). Dotted line: with only the adjoint of NNMI. Dash–dot line: with both the adjoint of NNMI and the penalty term J_{p1} ($r_1 = 10^{15}$).

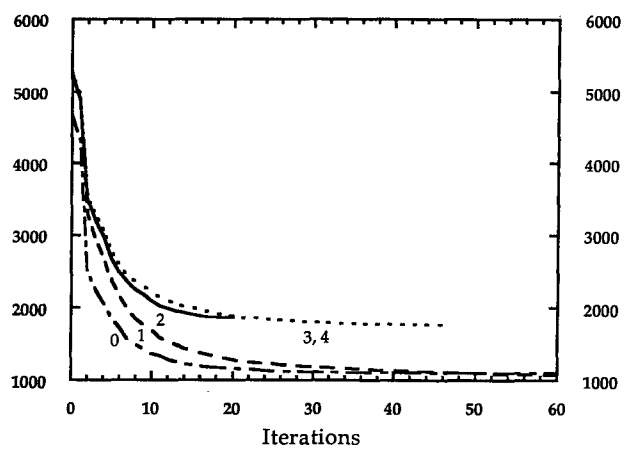

FIG. 13. Variations of the cost function J with the number of iterations when the Machenhauer NNMI is included in the minimization process. The numbers under the line represent the numbers of iterations of NNMI.

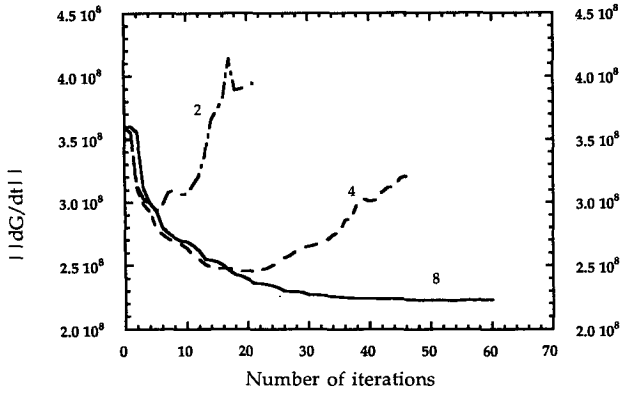

FIG. 14. Variation of $\|dz(t_0)/dt\|_2^2$ during the minimization process. The numbers above the line represent the numbers of NNMI iterations.

minimization process builds excessive gravity-wave oscillations. In these cases, the NNMI is practically invertible. Only when 8 iterations of NNMI are applied is the invertibility property lost.

If we combined the penalty term J_{p1} with the NNMI scheme in the variational process, the growth of the gravity noise in the case of NNMI application was eliminated by the penalty control (figures omitted).

Next, we carried out an experiment where the penalty term J_{p4} with the $\mathcal{L}_2$ norm definition (3.11) (constraint on time tendency of the gravity components only at the initial time) was added to the cost function J. A measure of the efficiency of the constraint imposed by J_{p4} is given by the corresponding variation of $\|dz(t_0)/dt\|_2^2$, which was found to decrease from 1.45 $\times 10^9$ to 8.77 $\times 10^7$ m^4 s^6 during the course of the minimization. The retrieved initial state, however, still contains a large number of gravity-wave oscillations. Therefore, we carried out another experiment in which the penalty term J_{p4m} with the $\mathcal{L}_2$ norm definition (3.11) (constraint on time tendency of the gravity components at every time step) was added to the cost function J. The value of $\sum_{i=0}^{R} \|dz(t_i)/dt\|_2^2$ was found to decrease from 2.63 $\times 10^{10}$ to 1.35 $\times 10^7$ m^4 s^6 after 60 iterations. As for the elimination of gravity-wave oscillations of the minimizing solution, we also carried out a 24-h forecast of the surface pressure. We see that the minimization process converged to a solution that produced a very smooth variation of the surface pressure (Fig. 15) even after 20–30 iterations.

To better exhibit the difference resulting from different definitions of the norm for J_{p4} and J_{p4m} (i.e., different weighting for the different variables of the gravity component), two additional experiments were carried out using the energy-norm definition. The value of $\|dz(t_0)/dt\|_E^2$ was found to decrease from 2.56 to 2.18 $\times 10^{-1}$ m^4 s^6 after 60 iterations and to 9.58 $\times 10^{-2}$ m^4 s^6 after 120 iterations during the minimization of $J + J_{p4}$. The value of $\sum_{i=0}^{R} \|dz(t_i)/dt\|_E^2$ decreased from 3.40 $\times 10^1$ to 5.98 m^4 s^6 after 60 iterations during the minimization of $J + J_{p4m}$. Unfortunately, there were still some gravity oscillations present in the subsequent 24-h forecast of the surface pressure for both experiments. We conclude that only the penalty term with a constraint on the gravity component applied at all time steps in the window of assimilation with the $\mathcal{L}_2$ norm yields a satisfactory elimination of the gravity-wave oscillations and does not slow the convergence rate of the minimization process.

In order to obtain a qualitative estimate about the elimination of high-frequency gravity-wave components after the minimization, we present in Table 1 and Fig. 16 the values of the norm of $dz(t_0)/dt$, that is, $\|dz(t_0)/dt\|_E^2$, contained in different retrieved initial states and the same norm for different vertical modes, respectively. As a reference, we also calculated $\|dz(t_0)/dt\|_E^2$ for the data at the initial time before and after the application of NNMI in Table 1. We see that after 30 iterations of the minimization with the penalty term (3.14) of the $\mathcal{L}_2$ norm, the solution contains an acceptable amount of gravity-wave components. The minimization of $J + J_{p4}$ with the energy norm for J_{p4} yields an acceptable amount of the gravity-

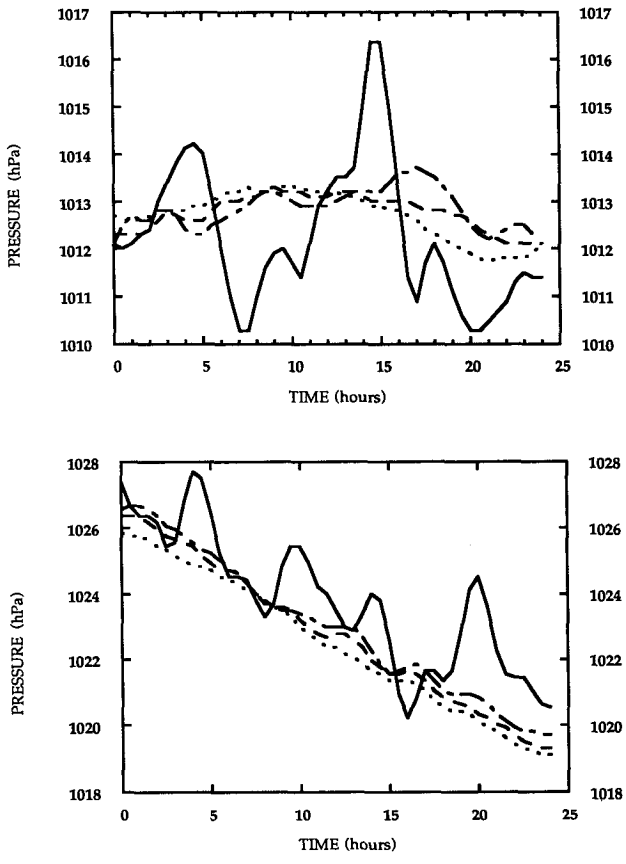

FIG. 15. Time traces of surface pressure at (a) point 1, and (b) point 2. Solid line: no penalty term. Other lines: with the penalty term J_{p4m} ($r_3 = 10^{-3}$) after 20 (dash–dot line), 30 (dashed line), and 60 (dotted line) iterations.

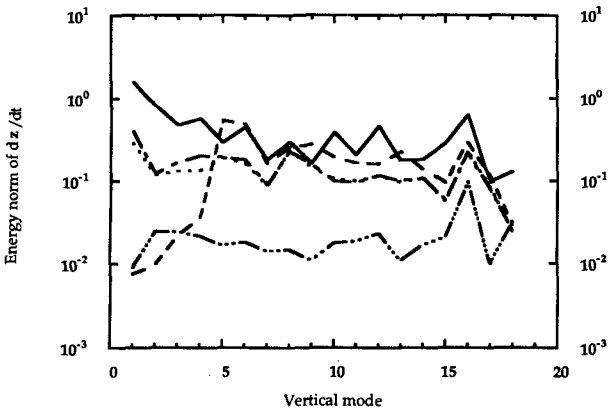

FIG. 16. The value of $\|dz(t_0)/dt\|_E^2$ as a function of the vertical mode. Solid line: no penalty. Dash–dot line: J_{p4} with the energy norm. Dotted line: J_{p4m} with the energy norm. Dashed line: J_{p4m} with the $\mathcal{L}_2$ norm. Dash–dot line: $J_{p1} + J_{p2}$.

wave component. The minimization of the cost function penalized by either (3.3) with the $\mathcal{L}_2$ norm or (3.14) with the energy norm does not result in a sufficient decrease in the value of $\|dz(t_0)/dt\|_E^2$. That is why the subsequent forecasts from these retrievals still exhibit the presence of gravity-wave oscillations, as mentioned previously. These results confirm that the application of penalty method where the constraint on the gravity modes' tendency is applied only at the initial time does not ensure the vanishing of the gravity modes' tendencies in the subsequent tendency computation due to the nonlinear nature of the model. Also, different weightings for the different variables of the gravity component play a very important role in ensuring the success of the minimization process.

From Table 1 we also observe that the constraints on the time tendency of the surface pressure and the divergence fields reduced the amount of the initial gravity-wave tendency to a value very near the smallest value in Table 1, which was obtained by the minimization of $J + J_{p4m}$ with the $\mathcal{L}_2$ norm for J_{p4m}. If, however, we take into account the large difference in the total CPU time spent in the minimization, we conclude that the penalization of the terms (2.4) and (2.6) results in a much cheaper computational procedure.

Figure 16 shows that the penalization with the terms given by (2.4) and (2.6) results in a reduction of $\|dz(t_0)/dt\|_E^2$ for all the vertical modes. The minimization of $J + J_{p4m}$ reduces the value of $\|dz(t_0)/dt\|_E^2$ only for the first four vertical modes as was to be expected (since only the first four vertical modes were included in the calculation of J_{p4m}). Minimization of J with the penalty terms given by (2.4) and (2.6), however, yields less damping of the gravity component contained in the second vertical mode. This might account for the fact that the subsequent forecast started from the solution of the minimization of the cost function $J + J_{p1} + J_{p2}$ still exhibits the presence of some half-day oscillations (see Fig. 6).

c. *Second-order time-tendency constraint*

In this experiment, the penalty term requiring the vanishing of the second-order time tendency of the surface pressure was included in the cost function ($r_3 = 10^{22}$). The value of J_{p3}/r_3 decreased two orders of magnitude during the minimization process (Fig. 17). Moreover, it removes the high-frequency oscillations with damping the low frequencies (see Fig. 18). If we combine both the first- and second-order time-tendency penalty to the surface pressure, J_{p1} and J_{p3}, using penalty parameters $r_1 = 10^{15}$ and $r^3 = 10^{22}$, we obtained a lower amplitude of gravity-wave oscillations than when only the third penalty term is applied. The increase of one order of magnitude in the penalty parameter r_1 resulted in a retrieval that was just as good as the one resulting from the application of the first penalty control. If, however, we implement the two penalty terms sequentially at the same cost, that is, the assimilated initial fields after 30 iterations of the minimization with the first penalty term included in the cost function were used as the starting point for a second minimization in which the second-order penalty term was added to the cost function, some small improvement may be observed (Fig. 19). Combining the second-order time-tendency penalty term with the first time-tendency penalty term might be important in situations where the first penalty term is not adequate to damp a sufficient amount of gravity-wave oscillations. The results obtained here point to the fact that there are situations where it may be beneficial to implement more than one constraint simultaneously in 4D variational data assimilation.

5. Summary and conclusions

The results presented here are obtained by augmenting the cost function with various penalty terms in order to control gravity oscillations in variational data assimilation using an adiabatic version of the NMC spectral model. While use of a weak constraint

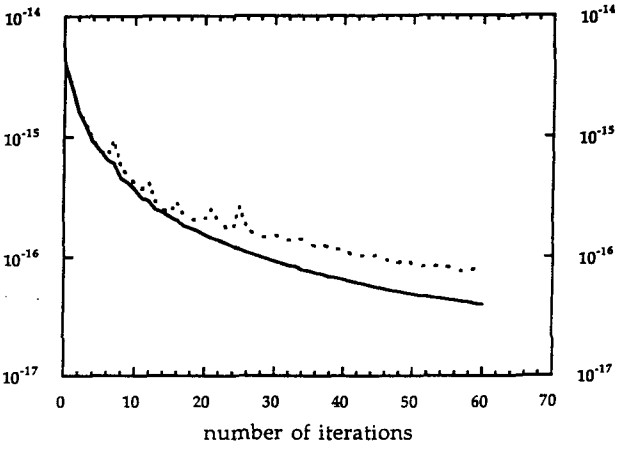

FIG. 17. The value of J_{p3}/r_3 as a function of the iteration number. Solid line: $r_3 = 10^{18}$. Dotted line: $r_3 = 10^{22}$.

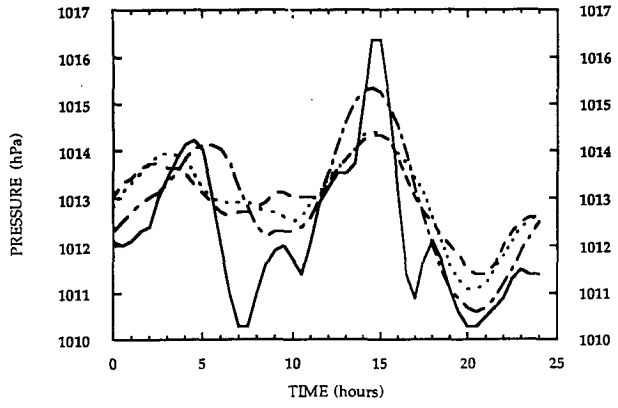

FIG. 18. Time traces of surface pressure at point 1. Solid line: no penalty term. Dash–dot line: with the penalty term J_{p3} ($r_3 = 10^{22}$). Dotted line: with the penalty terms $J_{p1} + J_{p3}$ ($r_1 = 10^{15}$ and $r_3 = 10^{22}$). Dashed line: same as dotted line but with $r_1 = 10^{16}$.

on the time tendency of the surface pressure controls the external gravity waves, a combination of both this constraint with the weak constraint on the time tendency of the divergence yielded satisfactory control of high-frequency gravity-wave noise present in the data.

Another penalty method with constraint applied only on the time tendency of the gravity components, similar to the one used by Thépaut and Courtier (1991), is also implemented.

We showed in this paper that the penalty constraints should be imposed at all the time steps in the window of assimilation to efficiently damp the high-frequency gravity-wave oscillations. The additional computational effort involved in adding the penalty terms to the cost function is very modest. The gradient of the penalized cost function can still be obtained by one single integration of the adjoint model. The additional computation involves adding more forcing terms to the right-hand side of the adjoint equations model. Therefore, the coding work is rather simple.

The positive numerical results on damping the gravity oscillations are reconfirmed by inspecting 24- and 72-h forecasts of the surface pressure, the retrieved initial surface pressure and divergence fields, and the value of rms of the divergence field.

In principle, the approach presented here can perform the function of initialization in the framework of variational data assimilation. The convergence rate of the penalized cost function does not appear to be affected by ill conditioning. The penalty method with constraints on the surface pressure and the divergence fields does not require any knowledge of the model's normal modes. Use of NNMI combined with the variational data assimilation does not yield good results due to the excess gravity-wave oscillations introduced in the gradient calculation. The variational data-assimilation procedure itself (without the inclusion of any penalty term), however, remains computationally expensive since the dimension is large ($\approx 10^6$) and the cost function and gradient evaluation is very costly.

One issue of major importance in the computational efficiency of variational 4D data assimilation is the efficiency of the optimization process. Two aspects should be mentioned. First, in order to achieve a reduction in the cost function, we need to carry out many iterations of the minimization procedure. Each iteration requires at least 2 model integrations over the span of the window for data assimilation. Second, in order to study the quality of the solution retrieved by the variational data-assimilation technique, it is necessary to compute some of the eigenvalues and eigenvectors of the Hessian matrix of the cost function with respect to the control variables. The Hessian matrix is a huge symmetric positive-definite matrix. If the Hessian has a large condition number, it means that there are model parameters not well determined by data. A spectral analysis of the Hessian matrix will reveal which directions in parameter space are well determined and which are poorly determined. One also needs to obtain crude estimates of the inverse Hessian matrix at the solution, since it can provide information about the uncertainty of the model parameters and even about the correlation that may exist between them (Thacker 1989).

Therefore, the need to develop parallel minimization algorithms becomes apparent. Variational data assim-

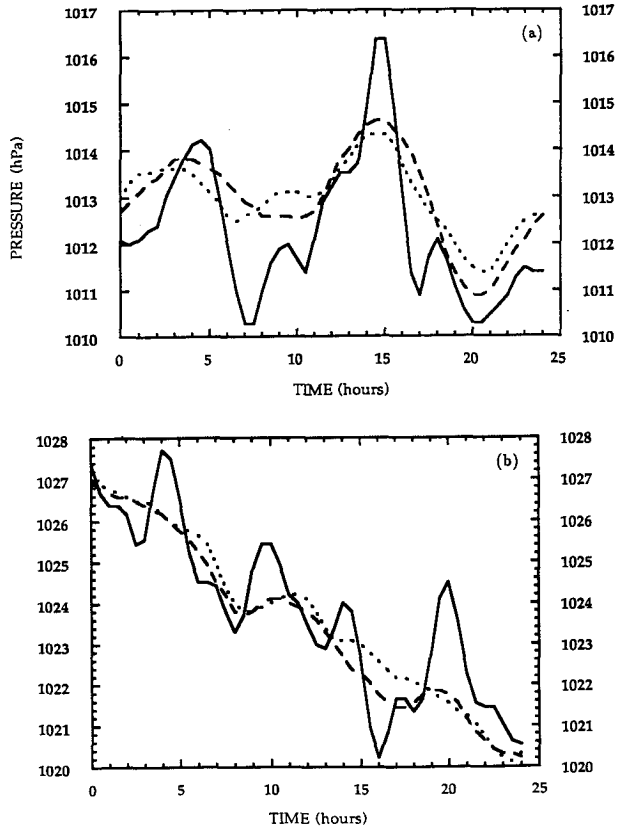

FIG. 19. Time traces of surface pressure at (a) point 1; (b) point 2. Solid line: no penalty term. Dotted line: with the penalty term J_{p1} ($r_1 = 10^{15}$). Dashed line: with both the penalty terms J_{p1} and J_{p3} carried out sequentially.

ilation will obtain speedups from the use of massively parallel processors, as well as from the efficient use of each processor's resources. After modifying both the direct and adjoint models as well as the unconstrained minimization algorithm, so as to take advantage of available parallel-processing architectures, variational data assimilation will become a more promising tool within reach of operational applications.

Acknowledgments. This research was funded by NSF Grant ATM-9102851. Additional support was provided by the National Meteorological Center and the Supercomputer Computations Research Institute at Florida State University, which is partially funded by the Department of Energy through Contract DE-FC0583ER250000. The authors would like to acknowledge the contribution of Professor John Lewis of NSSL and an anonymous reviewer whose insightful comments improved the presentation of the paper. The helpful comments of the internal reviewers at NMC are appreciated.

APPENDIX A

Derivation of the Evaluation of ∇J_{p1} or J_{p2} by Integrating the Adjoint Model

The derivation of the gradient calculation when the cost function J is penalized by either (2.4) or (2.5) is, of course, identical. Here, we only present the derivation of ∇J_{p1}.

Taking the finite-difference approximation for the time derivative, Eq. (2.4) can be written as

$$J_{p1} = \frac{r_1}{(\Delta t)^2}\mathbf{Y}^{\mathrm{T}}(t_0)\mathbf{Y}(t_0) + \frac{r_1}{(2\Delta t)^2}\sum_{i=1}^{R-1}\mathbf{Y}^{\mathrm{T}}(t_i)\mathbf{Y}(t_i), \quad \text{(A.1)}$$

where

$$\mathbf{Y}(t_0) = \ln p_s(t_0 + \Delta t) - \ln p_s(t_0)$$
$$\mathbf{Y}(t_i) = \ln p_s(t_i + \Delta t) - \ln p_s(t_i - \Delta t), \quad i = 1, \cdots, R-1, \quad \text{(A.2)}$$

where $\ln p_s(t_i + \Delta t)$, $i = 1, \cdots, R-1$ represents the value of the variable field after the application of a Robert time filter that is included in the model.

We define J'_{p1} to be the change in the cost function resulting from a small perturbation $\mathbf{X}'(t_0)$ about the initial condition of the control variable $\mathbf{X}(t_0)$.

For the definition of the gradient, we have

$$J'_{p1} = (\nabla J_{p1})^{\mathrm{T}}\mathbf{X}'(t_0). \quad \text{(A.3)}$$

Here, $\mathbf{X}(t_0)$ is the state vector of dimension $M(4K + 1)$ and contains the values of divergence, vorticity, temperature, surface pressure, and moisture over all the Gaussian grid points at the initial time. Using (A.1), we may write

$$J'_{p1} = \frac{2r_1}{(\Delta t)^2}\mathbf{Y}^{\mathrm{T}}(t_0)\mathbf{Y}'(t_0) + \frac{r}{(2\Delta t)^2}\sum_{i=1}^{R-1}\mathbf{Y}^{\mathrm{T}}(t_i)\mathbf{Y}'(t_i). \quad \text{(A.4)}$$

We now express $\ln p'_s(t_i)$, $i = 0, 1, 2, \cdots, R$, where $t_i = t_0 + i\Delta t$, in terms of $\mathbf{X}'(t_0)$ as

$$\ln p'_s(t_i) = \mathbf{Q}_{t_i}\mathbf{X}'(t_0). \quad \text{(A.5)}$$

Here, $\mathbf{Q}_{t_i}$ represents the product of all the operator matrices in the tangent linear model that produce $\ln p'_s(t_i)$ from $\mathbf{X}'(t_0)$, and $\mathbf{Q}_{t_0} = (\mathbf{I}\,\mathbf{0})$ is an $M \times (3M*K + M)$ matrix, and $\mathbf{I}$ is a unit matrix of dimension $M \times M$.

Substituting (A.5) into (A.4) and using the definition (A.2), we obtain

$$J'_{p1} = \frac{2r_1}{(\Delta t)^2}[(\mathbf{Q}^{\mathrm{T}}_{t_0+\Delta t} - \mathbf{Q}^{\mathrm{T}}_{t_0})\mathbf{Y}(t_0)]^{\mathrm{T}}\mathbf{X}'(t_0) + \frac{2r_1}{(2\Delta t)^2}\sum_{i=1}^{R-1}[(\mathbf{Q}^{\mathrm{T}}_{t_i+\Delta t} - \mathbf{Q}^{\mathrm{T}}_{t_i-\Delta t})\mathbf{Y}(t_i)]^{\mathrm{T}}\mathbf{X}'(t_0)$$
$$= \frac{2r_1}{(\Delta t)^2}\langle\{\mathbf{Q}^{\mathrm{T}}_{t_0}[-4\mathbf{Y}(t_0) - \mathbf{Y}(t_1)]\}^{\mathrm{T}}\mathbf{X}'(t_0) + \{\mathbf{Q}^{\mathrm{T}}_{t_1}[4\mathbf{Y}(t_0) - \mathbf{Y}(t_2)]\}^{\mathrm{T}}\mathbf{X}'(t_0)$$
$$+ 2r_1\sum_{i=2}^{R-2}\{\mathbf{Q}^{\mathrm{T}}_{t_i}[\mathbf{Y}(t_{i-1}) - \mathbf{Y}(t_{i+1})]\}^{\mathrm{T}}\mathbf{X}'(t_0) + 2r_1\sum_{i=R-1}^{R}[\mathbf{Q}^{\mathrm{T}}_{t_i}\mathbf{Y}(t_{i-1})]^{\mathrm{T}}\mathbf{X}'(t_0)\rangle, \quad \text{(A.6)}$$

where the matrix $\mathbf{Q}^{\mathrm{T}}_{t_i}$ represents the adjoint of $\mathbf{Q}_{t_i}$.

By comparing Eqs. (A.3) and (A.6), we note that $\nabla J'_{p1}$ can be obtained by integrating the adjoint model from the final time t_R to the initial time t_0 while inserting as forcing terms the penalized differences

$$2r_1\frac{\mathbf{Y}(t_{i-1})}{(2\Delta t)^2} \quad \text{(A.7)}$$

at times t_i, $i = R, R-1$; the terms

$$2r_1\left[\frac{\mathbf{Y}(t_{i-1})}{(2\Delta t)^2} - \frac{\mathbf{Y}(t_{i+1})}{(2\Delta t)^2}\right] \quad \text{(A.8)}$$

at times t_i, $i = R-2, \cdots, 2$, the terms

$$2r_1\left[\frac{\mathbf{Y}(t_0)}{(\Delta t)^2} - \frac{\mathbf{Y}(t_2)}{(2\Delta t)^2}\right] \quad \text{(A.9)}$$

at time t_1, and, finally, the term

$$-2r_1\left[\frac{\mathbf{Y}(t_0)}{(\Delta t)^2}+\frac{\mathbf{Y}(t_1)}{(2\Delta t)^2}\right] \tag{A.10}$$

at time t_0.

Therefore, the gradient of the cost function augmented by the penalty term J_{p1} defined in (2.4) can be obtained by integrating the adjoint model from time t_R to t_0 and inserting into the adjoint model at each time step the penalized terms defined in Eqs. (2.9)–(2.12), plus the weighted differences

$$2W(t_r)[\mathbf{X}(t_r)-\mathbf{X}^{\text{obs}}(t_r)] \tag{A.11}$$

whenever an analysis time t_r ($t_r \in [t_0, t_R]$) is encountered. Thus, a single integration of the adjoint model can still yield the value of the gradient of the penalized cost function with respect to the initial conditions.

If both penalty terms J_{p1} and J_{p2} are added to the cost function J, one integration of the adjoint model is still sufficient for the calculation of the gradient of the penalized cost function; under this condition, forcing terms resulting from both penalty terms have to be added to the backward integration of the adjoint model.

APPENDIX B

Derivation of the Evaluation of ΔJ_{p3} by Integrating the Adjoint Model

By defining J'_{p3} as the change in the cost function resulting from a small perturbation $\mathbf{X}'(t_0)$ about the initial condition $\mathbf{X}(t_0)$, we have

$$J'_{p3}=\frac{2r_3}{(\Delta t)^4}\sum_{i=1}^{R-1}[P(t_i)]^{\mathrm{T}}P'(t_i). \tag{B.1}$$

Using the definition of (2.12), (2.21) can be written as

$$J'_{p3}=\frac{2r_3}{(\Delta t)^4}\sum_{i=1}^{R-1}\{[\mathbf{Q}^{\mathrm{T}}(t_i+\Delta t)+\mathbf{Q}^{\mathrm{T}}(t_i-\Delta t)-2\mathbf{Q}^{\mathrm{T}}(t_i)]P(t_i)\}^{\mathrm{T}}\mathbf{X}'(t_0)=\frac{2r_3}{(\Delta t)^4}\big\langle[\mathbf{Q}^{\mathrm{T}}(t_0)P(t_1)]^{\mathrm{T}}\mathbf{X}'(t_0)$$
$$+\{\mathbf{Q}^{\mathrm{T}}(t_1)[P(t_2)+2P(t_1)]\}^{\mathrm{T}}\mathbf{X}'(t_0)+\sum_{i=2}^{R-2}\{\mathbf{Q}^{\mathrm{T}}(t_i)[P(t_{i+1})+P(t_{i-1})-2P(t_i)]\}^{\mathrm{T}}\mathbf{X}'(t_0)$$
$$+\{\mathbf{Q}^{\mathrm{T}}(t_{R-1})[P(t_{R-2})-2P(t_{R-1})]\}^{\mathrm{T}}\mathbf{X}'(t_0)+[\mathbf{Q}^{\mathrm{T}}(t_R)P(t_{R-1})]^{\mathrm{T}}\mathbf{X}'(t_0)\big\rangle. \tag{B.2}$$

On the other hand, the perturbation J'_{p3} can be expressed as

$$J'_{p1}=(\nabla J_{p1})^{\mathrm{T}}\mathbf{X}'(t_0). \tag{B.3}$$

Comparing (B.2) with (B.3), we immediately obtain that ∇J_{p3} can be calculated by integrating the adjoint model backward in time while adding at each time step the following penalty terms:

$$\frac{2r_3}{(\Delta t)^4}P(t_{R-1}) \tag{B.4}$$

at time t_R, the term

$$\frac{2r_3}{(\Delta t)^4}[P(t_{R-2})-2P(t_{R-1})] \tag{B.5}$$

at time t_{R-1}, the term

$$\frac{2r_3}{(\Delta t)^4}[P(t_{i+1})+P(t_{i-1})-2P(t_i)] \tag{B.6}$$

at times t_i, $i=R-2, \cdots, 2$, the terms

$$\frac{2r_3}{(\Delta t)^4}[P(t_2)-2P(t_1)] \tag{B.7}$$

at time t_1, and finally the term

$$\frac{2r_3}{(\Delta t)^4}P(t_1) \tag{B.8}$$

at initial time t_0.

Due to the linearity of the adjoint model, the gradient of the penalized cost function, $J+J_{p3}$, can be obtained by a single integration of the adjoint model with both the forcing terms (B.4)–(B.8) and (A.11) added to the right-hand side of the adjoint equations model.

APPENDIX C

Derivation of the Evaluation of ∇J_{p4} or J_{p4m} by Integrating the Adjoint Model

After the normal-mode projection, one has

$$\frac{d\mathbf{y}}{dt}=-i\Lambda_y\mathbf{y}+R_y(\mathbf{y},\mathbf{z}) \tag{C.1a}$$

$$\frac{d\mathbf{z}}{dt}=-i\Lambda_z\mathbf{z}+R_z(\mathbf{y},\mathbf{z}), \tag{C.1b}$$

where R_y and R_z are nonlinear terms for slow and fast mode equations, and Λ_y and Λ_z are the true slow and fast frequencies, respectively.

Assuming R_z is independent of time, we obtain

$$\mathbf{z}(t)=\frac{R_z[\mathbf{y}(t_0),\mathbf{z}(t_0)]}{i\Lambda_z}+\left\{\mathbf{z}(0)-\frac{R_z[\mathbf{y}(t_0),\mathbf{z}(t_0)]}{i\Lambda_z}\right\}e^{-i\Lambda_z t}. \tag{C.2}$$

Equation (C.2) consists of a constant term and a term oscillating at the gravity-wave frequency Λ_z. To elim-

inate high-frequency oscillations, one simply requires the second oscillating term to vanish initially, that is,

$$\mathbf{z}(t_0) = \frac{R_z[\mathbf{y}(t_0), \mathbf{z}(t_0)]}{i\Lambda_z}, \quad \text{or} \quad \dot{\mathbf{z}}|_{t=t_0} = 0, \quad \text{(C.3)}$$

resulting in the Machenhauer balance condition.

Here, however, R_z is time dependent. Machenhauer (1977) used a Picard-type iteration

$$\mathbf{z}_k(t_0) = \frac{R_z[\mathbf{y}(t_0), \mathbf{z}_{k-1}(t_0)]}{i\Lambda_z}, \quad k = 1, 2, \cdots. \quad \text{(C.4)}$$

Combining (C.4) with (3.1), we have

$$\mathbf{z}_{k-1}(t_0) - \mathbf{N}\mathbf{x}_{k-1}(t_0) = \frac{R_z[\mathbf{y}(t_0), \mathbf{z}_{k-1}(t_0)]}{i\Lambda_z}, \quad \text{(C.5)}$$

that is,

$$\mathbf{N}\mathbf{x}(t_0) = z(t_0) - \frac{R_z[\mathbf{y}(t_0), \mathbf{z}(t_0)]}{i\Lambda_z}, \quad \text{(C.6)}$$

where the subscript $k-1$ for variables $\mathbf{x}$ and $\mathbf{z}$ has been omitted.

Substituting (C.6) into (C.1b), we finally obtain

$$\frac{d\mathbf{z}(t_0)}{dt} = \mathbf{M}\mathbf{x}(t_0), \quad \text{(C.7)}$$

where $\mathbf{M} = -i\Lambda\mathbf{N}$.

Using (C7), (3.5) and $\mathbf{x} = \mathbf{S}^{-1}\mathbf{X}$, where $\mathbf{S}$ represents the transformation from the spectral space to the grid-point space, J_{p4} may be written as

$$J_{p4} = \frac{1}{2} r_4[\mathbf{WSHMS}^{-1}\mathbf{X}(t_0)]^{\mathrm{T}}[\mathbf{SHMS}^{-1}\mathbf{X}(t_0)]. \quad \text{(C.8)}$$

For the $\mathcal{L}_2$ norm (2.11), J_{p4} assumes the same form as (C.8) with both operators $\mathbf{H}$ and $\mathbf{W}$ being replaced by the identity operator.

A perturbation of J_{p4} resulting from a small perturbation of the initial state may be expressed as

$$J'_{p4} = r_4[\mathbf{S}^{-\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{WHMS}^{-1}\mathbf{X}(t_0)]^{\mathrm{T}}\mathbf{X}'(t_0). \quad \text{(C.9)}$$

Therefore, the gradient assumes the form

$$\nabla J_{p4} = r_4\mathbf{S}^{-\mathrm{T}}\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{S}^{\mathrm{T}}\mathbf{WHMS}^{-1}\mathbf{X}(t_0). \quad \text{(C.10)}$$

Thus, the process of calculating the gradient of J_{p4} consists mainly of a projection of the time tendency of the initial state on the gravity manifold, represented by $\mathbf{M}$, and the adjoint operation of $\mathbf{M}$.

Now, we will derive the calculation of ∇J_{p4m} defined in (3.6).

Substituting (C.7) and (3.5) into (3.6) and using $\mathbf{x} = \mathbf{S}^{-1}\mathbf{X}$, we obtain

$$J_{p4m} = \frac{1}{2} r_4 \sum_{i=0}^{R} [\mathbf{AX}(t_i)]^{\mathrm{T}}\mathbf{AX}(t_i), \quad \text{(C.11)}$$

where $\mathbf{A} = \mathbf{SHMS}^{-\mathrm{T}}$.

In order to derive the expression of the gradient of J_{p4m}, we again employ the perturbation method and obtain

$$J'_{p4m} = r_4 \sum_{i=0}^{R} [\mathbf{WAX}(t_i)]^{\mathrm{T}}\mathbf{AX}'(t_i). \quad \text{(C.12)}$$

We now express $\mathbf{X}'(t_i)$, $i = 1, 2, \cdots, R$, in terms of $\mathbf{X}'(t_0)$ as

$$\mathbf{X}'(t_i) = \mathbf{P}_i\mathbf{X}'(t_0), \quad \text{(C.13)}$$

where $\mathbf{P}_i$ represents the result of applying all the operator matrices in the linear tangent model to obtain $\mathbf{X}'(t_i)$ from $\mathbf{X}'(t_0)$.

Substituting (C.13) into (C.12), we obtain

$$J'_{p4m} = r_4 \sum_{i=0}^{R} [\mathbf{WAX}(t_i)]^{\mathrm{T}}\mathbf{AP}_i\mathbf{X}'(t_0). \quad \text{(C.14)}$$

On the other hand, in the limit as $\|\mathbf{X}'\| \to 0$, J'_{p4m} is the directional derivative in the $\mathbf{X}'(t_0)$ direction and is given by

$$J'_{p4m} = \{\nabla J_{p4m}[\mathbf{X}(t_0)]\}^{\mathrm{T}}\mathbf{X}'(t_0). \quad \text{(C.15)}$$

Comparison between (C.14) with (C.15) yields the gradient of the cost function J_{p4m}, ∇J_{p4m}, as

$$\nabla J_{p4m} = \sum_{i=0}^{R} \mathbf{P}_i^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}[r_4\mathbf{WAX}(t_i)]. \quad \text{(C.16)}$$

Since the operators $\mathbf{P}_i$ are linear, ∇J_{p4m} may be obtained by integrating the adjoint model from t_R to t_0 while the weighted differences

$$\mathbf{A}^{\mathrm{T}}[r_4\mathbf{WAX}(t_i)] \quad \text{(C.17)}$$

are inserted at every time step. Thus, a single integration of the adjoint model can yield the value of the gradient of the cost function J_{p4m} with respect to the initial conditions.

REFERENCES

Baer, F., 1977: Adjustments of initial conditions required to suppress gravity oscillations in nonlinear flows. *Contrib. Atmos. Phys.*, **50**, 350–366.

——, and J. Tribbia, 1977: On complete filtering of gravity waves through non-linear initialization. *Mon. Wea. Rev.*, **105**, 1536–1539.

Barker, E. H., G. J. Haltiner, and Y. K. Sasaki, 1977: Three-dimensional initialization using variational analysis. Preprints, *Third Conf. on Numerical Weather Prediction*, Omaha, Amer. Meteor. Soc., 169–181.

Bennett, A. F., and R. N. Miller, 1991: Weighting initial conditions in variational assimilation schemes. *Mon. Wea. Rev.*, **119**, 1098–1102.

Charney, J. G., 1955: The use of the primitive equations of motion in numerical prediction. *Tellus*, **7**, 22–26.

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

Courtier, P., 1984: Présentation d'une méthode variationelle d'assimilation dynamique de données meteorologique réparties dans l'espace et le temps. *Note E.E.R.M.*, **101**, 21 pp.

——, and O. Talagrand, 1990: Variational assimilation of meteo-

rological observations with the direct and adjoint shallow-water equations. *Tellus,* **42A,** 531–549.

Errico, R. M., 1989a: Theory and application of nonlinear normal mode initialization. NCAR Tech. Note, NCAR/TN-344+1A, 137 pp.

——, 1989b: The degree of Machenhauer balance in a climate model. *Mon. Wea. Rev.,* **117,** 2723–2733.

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

Hinkelmann, K., 1951: Der mechanismus des meteorologischen larmes. *Tellus,* **3,** 285–296. [Available in English as NCAR Tech. Note TN-203+STR from the National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307.]

LeDimet, 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 constraints. *Tellus,* **37A,** 309–322.

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

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with applications to normal mode initialization. *Contrib. Atmos. Phys.,* **50,** 253–271.

——, 1982: Fundamentals of nonlinear normal mode initialization. NCAR Tech. Note, NCAR/TN-204+PROC, 193 pp.

Miyakoda, K., and R. W. Moyer, 1968: A method of initialization for dynamical weather forecasting. *Tellus,* **20,** 115–128.

——, R. F. Strickler, and J. Chludzinski, 1978: Initialization with the data assimilation method. *Tellus,* **30,** 32–54.

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

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

Phillips, N. A., 1960: On the problem of initial data for the primitive equations. *Tellus,* **12,** 121–126.

Sasaki, Y., 1970: Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.,* **98,** 875–883.

——, P. S. Ray, J. S. Goerss, and P. Soliz, 1979: Inconsistent finite differencing errors in the variational adjustment of horizontal wind components. *J. Meteor. Soc. Japan,* **57,** 88–92.

Sela, J. G., 1980: Spectral modeling at the National Meteorological Center. *Mon. Wea. Rev.,* **108,** 1279–1292.

——, 1982: The NMC Spectral Model. NOAA Tech. Rep. NWS 30.

Thacker, W. C., 1989: The role of the Hessian matrix in fitting models to measurements. *J. Geophy. Res.,* **94,** 6177–6196.

——, 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.

Tribbia, J. J., 1984: A simple scheme for high-order nonlinear normal mode initialization. *Mon. Wea. Rev.,* **112,** 278–284.