

## Examination of Numerical Results from Tangent Linear and Adjoint of Discontinuous Nonlinear Models

S. ZHANG

*Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey*

X. ZOU AND JON E. AHLQUIST

*Meteorology Department, The Florida State University, Tallahassee, Florida*

(Manuscript received 18 August 2000, in final form 7 May 2001)

### ABSTRACT

The forward model solution and its functional (e.g., the cost function in 4DVAR) are discontinuous with respect to the model's control variables if the model contains discontinuous physical processes that occur during the assimilation window. In such a case, the tangent linear model (the first-order approximation of a finite perturbation) is unable to represent the sharp jumps of the nonlinear model solution. Also, the first-order approximation provided by the adjoint model is unable to represent a finite perturbation of the cost function when the introduced perturbation in the control variables crosses discontinuous points. Using an idealized simple model and the Arakawa–Schubert cumulus parameterization scheme, the authors examined the behavior of a cost function and its gradient obtained by the adjoint model with discontinuous model physics. Numerical results show that a cost function involving discontinuous physical processes is *zeroth-order* discontinuous, but piecewise differentiable. The maximum possible number of involved discontinuity points of a cost function increases exponentially as  $2^{kn}$ , where  $k$  is the total number of thresholds associated with on–off switches, and  $n$  is the total number of time steps in the assimilation window. A backward adjoint model integration with the proper forcings added at various time steps, similar to the backward adjoint model integration that provides the gradient of the cost function at a continuous point, produces a one-sided gradient (called a subgradient and denoted as  $\nabla^s J$ ) at a discontinuous point. An accuracy check of the gradient shows that the adjoint-calculated gradient is computed exactly on either side of a discontinuous surface. While a cost function evaluated using a small interval in the control variable space oscillates, the distribution of the gradient calculated at the same resolution not only shows a rather smooth variation, but also is consistent with the general convexity of the original cost function. The gradients of discontinuous cost functions are observed roughly smooth since the adjoint integration correctly computes the one-sided gradient at either side of discontinuous surface. This implies that, although  $(\nabla^s J)^T \delta \mathbf{x}$  may not approximate  $\delta J = J(\mathbf{x} + \delta \mathbf{x}) - J(\mathbf{x})$  well near the discontinuous surface, the subgradient calculated by the adjoint of discontinuous physics may still provide useful information for finding the search directions in a minimization procedure. While not eliminating the possible need for the use of a nondifferentiable optimization algorithm for 4DVAR with discontinuous physics, consistency between the computed gradient by adjoints and the convexity of the cost function may explain why a differentiable limited-memory quasi-Newton algorithm still worked well in many 4DVAR experiments that use a diabatic assimilation model with discontinuous physics.

### 1. Introduction

The goal of numerical weather prediction is to produce the most accurate forecasts possible. Once the equations are chosen and written in a form appropriate for computer solution, a global forecast depends on the initial conditions and model parameters. Four-dimensional variational data assimilation (4DVAR) chooses some or all of these input variables objectively. With 4DVAR, one begins by choosing a non-

negative, typically quadratic in terms of the model state, forecast error function (also called a cost function, usually chosen as a typically quadratic form of the model state). Then one applies variations of calculus to minimize the cost function by varying the input variables (also called control variables). The forecast model (called the assimilation model) serves as a strong constraint within the four-dimensional verification space of time and three-dimensional physical or Fourier-transformed space.

The cost function that is minimized in 4DVAR typically contains an observational term of the following form:

---

*Corresponding author address:* Dr. S. Zhang, GFDL/NOAA, Princeton University, P.O. Box 308, Princeton, NJ 08542.  
E-mail: snz@gfdl.noaa.gov

$$\begin{aligned}
 J_o(\mathbf{x}_0) = & \frac{1}{2} \sum_{i=0}^R \underbrace{\{\mathcal{H}[\mathcal{M}(\mathbf{x}_0)] - \mathbf{y}_{\text{obs}}\}^T}_{\mathbf{x}(t_i)} \\
 & \times \mathbf{R}^{-1} \underbrace{\{\mathcal{H}[\mathcal{M}(\mathbf{x}_0)] - \mathbf{y}_{\text{obs}}\}}_{\mathbf{x}(t_i)}. \quad (1)
 \end{aligned}$$

This is a weighted sum of the squared differences between the forecast model solutions  $\mathbf{x}(t_i) = \mathcal{M}(\mathbf{x}_0)$  and observations  $\mathbf{y}_{\text{obs}}$  in the time window between  $t_0$  and  $t_0 + t_R$ , where  $\mathbf{x}_0$  includes the model state variable at the initial time  $t_0$  and model parameters,  $\mathcal{M}$  is the operator representing the operations performed in the nonlinear model to obtain the model forecast at time  $t$  from an initial condition at time  $t_0$  ( $t > t_0$ ),  $\mathbf{y}_{\text{obs}}$  is a vector of observations,  $\mathbf{R}$  is an observational error covariance matrix, and  $\mathcal{H}$  is an observation operator that converts the model variables on model grids to the observation quantities and locations. When the assimilation model contains no discontinuous process,  $J_o$  is a continuously differentiable function of the control variables. Then a perturbation analysis of  $J_o$  yields

$$\delta J_o \approx (\nabla J_o)^T \delta \mathbf{x}_0, \quad (2)$$

where  $\nabla J_o$  is a column vector containing the gradient of the observational term of the the cost function with respect to the control variables and  $\delta \mathbf{x}_0$  is a perturbation of the control variables. Determination of the sensitivity of a forecast response to variations in initial conditions, or the iterative determination of initial conditions that gradually reduces the values of a cost function, requires knowledge of the gradient of the forecast response or the cost function with respect to the control variables.

By differentiation of (1) and a little manipulation, Le Dimet and Talagrand (1986) showed how the gradient of  $J_o$  can be found through a backward integration from  $t_R$  and  $t_0$  of the adjoint model, that is,

$$\begin{aligned}
 \nabla J_o = & \sum_{i=0}^R \underbrace{\left(\frac{\partial \mathcal{M}}{\partial \mathbf{x}_0}\right)^T}_{\text{(adjoint model)}} \underbrace{\left(\frac{\partial \mathcal{H}}{\partial \mathbf{x}(t_i)}\right)^T}_{\text{(adjoint observation operator)}} \underbrace{\mathbf{R}^{-1}[\mathcal{H}(\mathbf{x}) - \mathbf{y}_{\text{obs}}]}_{\text{(error weighted innovation)}}, \quad (3)
 \end{aligned}$$

where  $\mathbf{M} \equiv \partial \mathcal{M} / \partial \mathbf{x}_0$  is the tangent linear model (TLM) operator of the original forecast model (called forward nonlinear model), and  $\mathbf{M}^T$  is the adjoint model operator. Based on the same assumption of differentiability of the model solution and  $J_o$ , the perturbation analysis provides two procedures to check the correctness of the tangent linear model and the gradient of  $J_o$  (Courtier and Talagrand 1987; Navon et al. 1992). Once the values of  $J_o$  and the gradient of  $J_o$ ,  $\nabla J_o$ , with respect to the control variables, can be calculated, as well as the total cost function  $J$  and its gradient  $\nabla J$ , an unconstrained minimization algorithm (Liu and Nocedal 1989; Zou et al.

1993a) can be employed to find the ‘‘optimal’’ values of the control variables that minimize  $J$ .

When the assimilation model is diabatic, containing parameterized physics to account for the effects of the subgrid-scale processes (e.g., turbulence and cumulus convection) on the grid-resolvable scale state of atmosphere,  $J_o$  is discontinuous and nondifferentiable. These subgrid-scale physical processes in the atmosphere modify the local budget of mass, momentum, and energy and appear in the governing equations as source/sink terms. Usually, these processes depend strongly on local atmospheric conditions, that is, on the state in an individual grid-box column. These local conditions control ‘‘on-off’’ switches of the physical parameterizations. The on-off switches are governed by values of certain control variables or quantities derived from the control variables. If such a variable reaches a threshold value, these source/sink terms are turned on or off at a particular time step at a grid point.

Two examples are shown in Fig. 1, where  $J_o$  (denoted as  $J_{\text{model}}$ ) is defined as a weighted sum of the 6-h forecast errors [see Eq. (21)]. For simplicity, the time filter coefficient ( $\alpha_A$ ) and the horizontal diffusion coefficient ( $\alpha_{\text{HD}}$ ) are chosen as the input control variables of  $J_{\text{model}}$  using an adiabatic (upper panels) and a diabatic (lower panels) National Centers for Environmental Prediction (NCEP) global spectral model, respectively. The values of  $J_{\text{model}}$  are calculated at an interval of 0.001 for  $\alpha_A$  and  $0.001 \times 10^{16}$  for  $\alpha_{\text{HD}}$ . The variation of  $J_{\text{model}}$  in the adiabatic case (Figs. 1a,c) shows a very smooth behavior in the control variable space, while the variation of  $J_{\text{model}}$  in the diabatic case is obviously nonsmooth.

Should we use the diabatic assimilation model in 4DVAR? A diabatic forecast model including parameterized physics usually simulates the evolution of the atmosphere better than an adiabatic one does. Theoretically, a more realistic assimilation model will produce better 4DVAR results than an adiabatic model if real observations are assimilated, not to mention the necessity of including model physics to assimilate some physical quantities such as clouds and precipitation. Therefore, efforts have been made to incorporate parameterized physics into 4DVAR experiments using primitive equation models (Zupanski 1993; Xiao et al. 2000; Zou et al. 1993b; Zou and Kuo 1996; Zhu and Navon 1999; Mahfouf and Rabier 2000). Numerical results in these studies showed not only reasonably convergences of minimization procedures in which differentiable minimizations are used, but also various levels of improvement in analyses and model forecasts. In these experiments, some researchers retain the on-off switches of parameterizations in the tangent linear and adjoint models, as in the forward nonlinear model (Zou et al. 1993b; Zou and Kuo 1996; Zou 1997; Xiao et al. 1999; Mahfouf and Rabier 2000, hereafter referred to as ‘‘classical’’ adjoint method). Others (Zupanski 1993; Zhu and Navon 1999) attempted to remove the on-off switches using a transitional smooth function to make the model

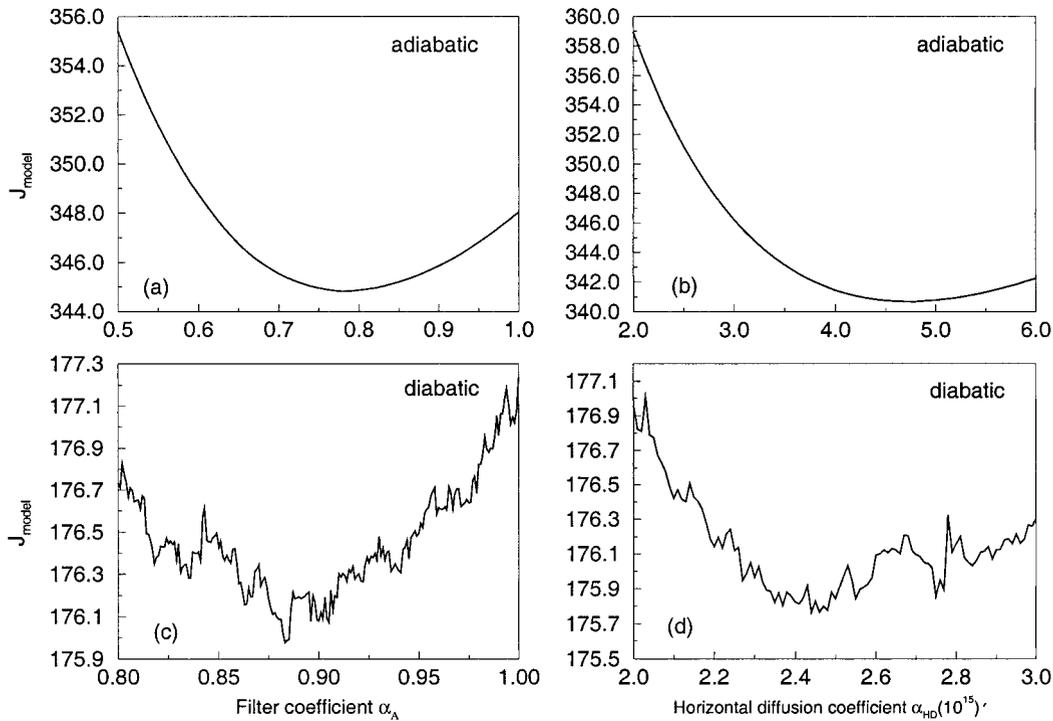


FIG. 1. Variation of the cost function  $J_{\text{model}}$  defined as a weighted sum of the 6-h squared forecast errors [see (21)] using the NCEP adiabatic (top) and diabatic (bottom) global spectral models with a time filter coefficient  $\alpha_A$  (left) and a horizontal diffusion coefficient  $\alpha_{\text{HD}}$  (right) (see section 4). The cost function is evaluated through integrating the nonlinear models using 0.001 and  $0.001 \times 10^{16}$  as the intervals of  $\alpha_A$  and  $\alpha_{\text{HD}}$ .

solution and the cost function smooth. In contrast, Xu (1996a) demonstrated that the classical adjoint method fails to correctly evaluate the gradient of a cost function that contains on-off switches using a perturbation analysis starting from a simple analytic model with a step-function source term. He proposed a generalized adjoint formalism that accounts for the effect of on-off switches in parameterized physical processes (Xu 1996a,b). The generalized adjoint was then extended to a vector system (Xu 1996b), as well as to multiple threshold conditions (Xu 1997a) and the time discretization situation (Xu 1997b). Furthermore, a ‘‘coarse-grain’’ tangent linearization and adjoint was proposed to deal with on-off switches triggered at discrete time levels (Xu et al. 1998). Numerical results on minimization are presented using the Lorenz-63 model (Xu et al. 1999). However, implementing their approach in data assimilation using a primitive equation model including complicated physical parameterization schemes seems difficult and requires further study.

In this study, we indicate that the cost function containing discontinuous parameterizations is *zereth-order* discontinuous (e.g., the function itself is discontinuous due to on-off switches in the forward model), and classical adjoint method provides the one-sided gradient of a cost function containing discontinuous parameterizations. This one-sided gradient, the limit of the ratio of vanishingly small per-

turbations of the cost function and control variable, can be used to construct a subgradient in a nondifferentiable optimization algorithm such as the bundle method (Lemarechal and Sagastizabal 1997; Zhang et al. 2000) for improving the minimization of discontinuous functions. We examine the behavior of such a discontinuous cost function in phase space. With the theoretical and numerical results from an adjoint model integration using the classical adjoint method, insights into the following questions are gained.

- 1) What does the tangent linear model solution fail to represent in the presence of on-off switches?
- 2) What does the result of an adjoint model integration represent, or fail to represent, in the presence of on-off switches?
- 3) How can the correctness for both the tangent linear and adjoint models be checked in the presence of on-off switches?
- 4) Why did a differential minimization algorithm still work well in many 4DVAR experiments that used a diabatic assimilation model with discontinuous physics?

This paper is arranged as follows. Section 2 uses an idealized simple model with a typical discontinuous source term to examine the behavior of the tangent linear and adjoint model solutions. A chain rule method is used to derive the gradient of a cost function at a con-

tinuous point, or a one-sided gradient at a discontinuous point, through an adjoint model integration. In section 3, a one-type cloud Arakawa–Schubert cumulus parameterization scheme employed in the NCEP global spectral model is used to investigate features of the forward model, cost function, TLM, and the adjoint model. Numerical results of the gradient, calculated using the adjoint of the NCEP diabatic model, are examined in section 4. Section 5 discusses the correctness check of the adjoint of the forward operator containing discontinuous physical processes. Summary and discussions are given in section 6.

**2. Zeroth-order discontinuous cost function and its one-sided gradient calculation using classical adjoint**

*a. Zeroth-order discontinuous cost function*

The governing equation of any numerical weather prediction (NWP) model can be written symbolically as

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{F}(\mathbf{x}, \boldsymbol{\beta}), \tag{4}$$

where  $\boldsymbol{\beta}$  represents a parameter vector that consists of physical parameters (such as the moistening parameter  $b$  in the Kuo scheme) and/or parameters introduced for computational stability (such as the filter coefficient). Numerical prediction of the state variable  $\mathbf{x}$  is produced by numerical integration of a forecast system, which results from expanding the governing equations (4) onto a mesh system. When  $\mathbf{F}(\mathbf{x}, \boldsymbol{\beta})$  involves no discontinuous process (e.g., an adiabatic model), both the model solution and the cost function defined by the model solution are differentiable with respect to control variables  $\mathbf{x}$  and/or  $\boldsymbol{\beta}$ .

Consistent with the previous work (Xu 1996a,b), when the source/sink terms defined by parameterized physics are included in the assimilation model,  $\mathbf{F}(\mathbf{x}, \boldsymbol{\beta})$  of (4) contains on–off switches and is differentiable only between two neighboring switches. The model operator  $\mathbf{F}(\mathbf{x}, \boldsymbol{\beta})$  is thus piecewise differentiable. In this section, we first ensure that the model solution, which is a result of many timestep integrations and the cost function defined on it are zeroth-order discontinuous, that is, it is the function itself that is discontinuous. In other situations, the cost function may be continuous, but with first- or high-order discontinuous derivatives. For sim-

plicity, without changing the essence of the problem, we use a simple single-variable piecewise differentiable model to examine the behavior of the model solution in the time domain and a cost function defined by this model in the phase space.

Consider a forecast model with the following form

$$\frac{\partial x}{\partial t} = \begin{cases} f_1(x), & x < x_c \\ f_2(x), & x \geq x_c, \end{cases} \tag{5}$$

where  $f_1(x)$  and  $f_2(x)$  are different (i.e.,  $f_1(x) \neq f_2(x)$  when  $x \geq x_c$ ) and both are differentiable functions. The similar stepfunction was used in the “generalized” adjoint study (Xu 1996a). The (classic) tangent linearized version of (5) takes the form of

$$\frac{\partial \delta x}{\partial t} = \begin{cases} f'_1(x)\delta x, & x < x_c \\ f'_2(x)\delta x, & x \geq x_c, \end{cases} \tag{6}$$

where  $f'_1$  and  $f'_2$  represent the differentiations of  $f_1$  and  $f_2$  with respect to  $x$ , respectively. Starting from an initial value  $x_0 < x_c$  and a perturbation  $\Delta x_0$  that crosses  $x_c$  (i.e.,  $x_0 + \Delta x_0 \geq x_c$ ) with a forward time integration scheme, the tangent linear and nonlinear changes of the perturbation through one time step integration are, respectively,

$$\Delta x_1^{\text{TL}} = [f'_1(x_0)\Delta t + 1]\Delta x_0, \tag{7}$$

$$\begin{aligned} \Delta x_1^{\text{NL}} &= [f_2(x_0 + \Delta x_0) - f_1(x_0)]\Delta t + \Delta x_0 \\ &= [f_2(x_0 + \Delta x_0) - f_1(x_0 + \Delta x_0) + f'_1(x_0)]\Delta t \\ &\quad + \Delta x_0. \end{aligned} \tag{8}$$

where superscripts TL and NL denote a tangent linear and nonlinear growth of a small perturbation, respectively. Apparently, the tangent linear change  $\Delta x_1^{\text{TL}}$  cannot approximate the nonlinear growth  $\Delta x_1^{\text{NL}}$  when the small perturbation  $\Delta x_0$  crosses a threshold at the first time step, no matter how small it is. Does this mean that the classical adjoint method fails to evaluate the gradient of the cost function containing discontinuous physics? In order to answer this question, we use the chain rule method to illustrate that an adjoint model integration provides a one-sided gradient at a discontinuous point of the cost function.

Using the same simple model, and considering the whole domain of the model state  $x$  after two time step integrations, one has four possible values for the model state at time  $t_0 + 2\Delta t$ :

$$x_2 = \begin{cases} f_1(x_1)\Delta t + x_1, & x_0 < x_c \text{ and } x_1 = f_1(x_0)\Delta t + x_0 < x_c \\ f_2(x_1)\Delta t + x_1, & x_0 < x_c \text{ and } x_1 = f_1(x_0)\Delta t + x_0 \geq x_c \\ f_2(x_1)\Delta t + x_1, & x_0 \geq x_c \text{ and } x_1 = f_2(x_0)\Delta t + x_0 \geq x_c \\ f_1(x_1)\Delta t + x_1, & x_0 \geq x_c \text{ and } x_1 = f_2(x_0)\Delta t + x_0 < x_c. \end{cases} \tag{9}$$

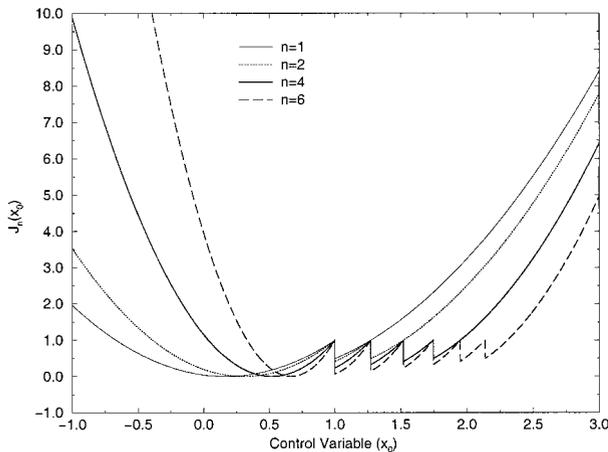


FIG. 2. Variation of the cost function  $J_n$  of a single-variable discontinuous model [see (10) and (11)]. The number “ $n$ ” represents the total time steps in the assimilation window.

As the  $n$ -step time integration proceeds for this one threshold situation, the number of possible switches that could occur increases exponentially by the law of  $2^n$ . If the number of thresholds increases (say,  $k$ ), the number of possible switches that were turned on and off will increase by the law of  $2^{kn}$ . The model state between two neighboring switches is differentiable. We indicate that in many situations, however, states at nearby grid points will be mutually compatible, and the effective number of possible different states will be smaller than  $2^{kn}$ .

A cost function defined as a continuous transformation of the model state in various times [usually a quadratic form such as the one in (11)] is also piecewise differentiable.

Let us consider a specific numerical example given by a forecast model

$$\frac{dx}{dt} = \begin{cases} 2x - 2, & x < 1 \\ x - 4, & x \geq 1 \end{cases} \quad (10)$$

and a series of cost functions

$$J_n(x_0) = x^2(t_n), \quad (11)$$

where  $t_n = t_0 + n\Delta t$ . Using  $t_0 = 0$  and  $\Delta t = 0.1$ , we obtain the distributions of  $J_n$  in the control variable ( $x_0$ ) space (Fig. 2). It is found that the cost functions consist of differentiable pieces between the discontinuity points. The longer the assimilation window (the more time integration steps), the more discontinuous points the cost function contains.

Without losing the generality, the on-off switches in parameterizations can be expressed as the form of “if” statements:

If ( $a > a_c$ ) then  
 ...  
 else  
 ...

where  $a$  is a scalar function (which may also be a function of the model input) of the model input and  $a_c$  is a threshold value of  $a$ .

Here,  $a = a_c$  defines the locus of the discontinuities in input space. That locus will normally be a “hypersurface” with dimension  $m - 1$ , where  $m$  is the dimension of the input space (this picture may be complicated by various facts, and in particular by the fact that the functions  $a$  and  $a_c$ , because of prior on-off processes, may themselves be discontinuous functions of the input). But if we accept that the locus of the discontinuities resulting from each individual on-off process is an  $(m - 1)$ -dimensional hypersurface, the hypersurfaces defined by all such processes in the model will partition the input space into  $m$ -dimensional cells, within each of which the cost function will be continuously differentiable. Therefore, we say, the zeroth-order discontinuous cost function defined on parameterized physics is piecewise differentiable.

*b. Calculation of one-sided gradient of a zeroth-order discontinuous cost function by classical adjoint approach*

The analysis of section 2a showed that the cost function defined on parameterized physics with on-off switches is piecewise differentiable. The classical adjoint approach individually conducts the linearization and transposition in each cell for coding the adjoint. The approach evaluates the gradient of the cost function at some given point in control variable space, say,  $\mathbf{x}_0$ , which is located within the certain cell, by the chain rule as

$$\nabla_{|\mathbf{x}_0} J = \begin{pmatrix} \frac{\partial J}{\partial x_{01}} \\ \frac{\partial J}{\partial x_{02}} \\ \vdots \\ \frac{\partial J}{\partial x_{0n}} \end{pmatrix} = \begin{pmatrix} \frac{\partial J}{\partial x_1} \frac{\partial x_1}{\partial x_{01}} + \frac{\partial J}{\partial x_2} \frac{\partial x_2}{\partial x_{01}} + \dots + \frac{\partial J}{\partial x_n} \frac{\partial x_n}{\partial x_{01}} \\ \frac{\partial J}{\partial x_1} \frac{\partial x_1}{\partial x_{02}} + \frac{\partial J}{\partial x_2} \frac{\partial x_2}{\partial x_{02}} + \dots + \frac{\partial J}{\partial x_n} \frac{\partial x_n}{\partial x_{02}} \\ \vdots \\ \frac{\partial J}{\partial x_1} \frac{\partial x_1}{\partial x_{0n}} + \frac{\partial J}{\partial x_2} \frac{\partial x_2}{\partial x_{0n}} + \dots + \frac{\partial J}{\partial x_n} \frac{\partial x_n}{\partial x_{0n}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\partial x_1}{\partial x_{01}} & \frac{\partial x_1}{\partial x_{02}} & \dots & \frac{\partial x_1}{\partial x_{0n}} \\ \frac{\partial x_2}{\partial x_{01}} & \frac{\partial x_2}{\partial x_{02}} & \dots & \frac{\partial x_2}{\partial x_{0n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial x_{01}} & \frac{\partial x_n}{\partial x_{02}} & \dots & \frac{\partial x_n}{\partial x_{0n}} \end{pmatrix}^T \begin{pmatrix} \frac{\partial J}{\partial x_1} \\ \frac{\partial J}{\partial x_2} \\ \vdots \\ \frac{\partial J}{\partial x_n} \end{pmatrix} = \mathbf{M}^T \begin{pmatrix} \frac{\partial J}{\partial x_1} \\ \frac{\partial J}{\partial x_2} \\ \vdots \\ \frac{\partial J}{\partial x_n} \end{pmatrix}, \quad (12)$$

where  $n$  is the dimension of the model state, and  $\mathbf{x}_0 = (x_{01}, x_{02}, \dots, x_{0n})^T$  and  $\mathbf{x} = (x_1^t, x_2^t, \dots, x_n^t)^T$  represent the initial condition and the model state at time  $t$ , respectively. Here  $\mathbf{M}$  is a linear propagator derived from differentiating  $\mathbf{F}(\mathbf{x}, \boldsymbol{\beta})$ , which consists of a product of

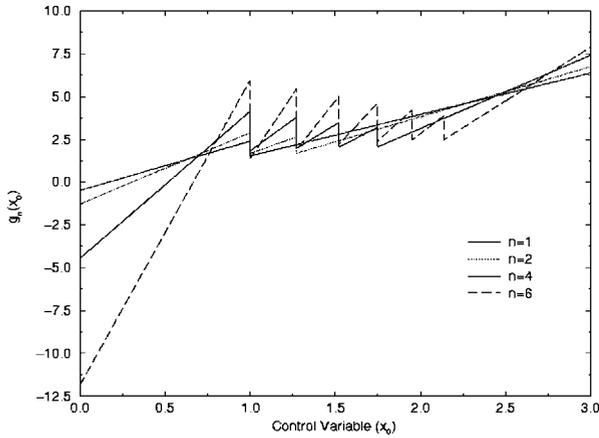


FIG. 3. Same as Fig. 2 except for gradient.

$t_R$  differentiable subpropagators represented by each step time integration where  $t_R$  denotes the total number of time step integrations, that is,

$$\mathbf{M} = \mathbf{M}_{t_R} \mathbf{M}_{t_R-1} \cdots \mathbf{M}_1. \quad (13)$$

Here

$$\mathbf{M}_t = \begin{pmatrix} \frac{\partial x_1^t}{\partial x_1^{t-1}} & \frac{\partial x_1^t}{\partial x_2^{t-1}} & \cdots & \frac{\partial x_1^t}{\partial x_n^{t-1}} \\ \frac{\partial x_2^t}{\partial x_1^{t-1}} & \frac{\partial x_2^t}{\partial x_2^{t-1}} & \cdots & \frac{\partial x_2^t}{\partial x_n^{t-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n^t}{\partial x_1^{t-1}} & \frac{\partial x_n^t}{\partial x_2^{t-1}} & \cdots & \frac{\partial x_n^t}{\partial x_n^{t-1}} \end{pmatrix}, \quad t = 1, \dots, t_R$$

where each element in  $\mathbf{M}_t$  can be evaluated in individual cell. Therefore, we have

$$\nabla_{|x_0} J = \mathbf{M}_1^T \mathbf{M}_2^T \cdots \mathbf{M}_{t_R}^T \frac{\partial J}{\partial \mathbf{x}_{t_R}}, \quad (14)$$

that is, an integration of the adjoint model containing discontinuous physical parameterizations is able to evaluate the one-sided gradient of the cost function at a discontinuity point as accurately as it does at continuous points.

Figure 3 displays the variations of the gradients of the cost functions defined in (11) using the model (10) (the cost functions were shown in Fig. 2). These gradients,  $dJ_n/dx_0$ , are calculated by the chain rule as in (12) for a one-dimensional control variable situation, in which the adjoint model of (10) is developed using the classical adjoint method, that is, keeping the on-off switches the same as in the nonlinear model. The gradients evaluated from the classical adjoint integration are found to be identical to the direct analytical solution.

The derivations and numerical examinations shown above suggest that the adjoint accuracy to evaluate the gradient of the cost function is identical to the accuracy of linearization that is conducted on the forward model

for building the tangent linear model. In other words, wherever the tangent linearization is conducted, the adjoint is unambiguously defined. At either side of a discontinuity point, the adjoint approach numerically determines the one-sided gradient of the cost function as accurately as the linearization is conducted at either side. This is consistent with some theoretical understanding (Errico 1997). In practice, for a continuous function, one often uses a linear change along the tangent direction to approximately describe a finite nonlinear perturbation, that is, the tangent linear approximation. The analysis above shows that the accuracy of the calculated one-sided gradients has nothing to do with the invalidity of the tangent linear approximation for a finite nonlinear perturbation that crosses the discontinuity point, that is,  $\Delta J \neq \mathbf{M} \Delta \mathbf{x}$ , and is irrelevant to the magnitude of the finite perturbations  $\Delta \mathbf{x}$ . The adjoint approach deals with vanishingly small perturbations, while the tangent linear approximation deals with finite perturbations. The degree of accuracy of the tangent linear approximation says something about the applications for which a numerically computed gradient can be useful, but that is a distinctly different question.

Therefore, the classical adjoint approach does not work for finite perturbations in a nonlinear and discontinuous system, which is for computing finite difference ratios of the form  $\Delta J/\Delta x_i$ , for a given  $J$  defined by a discontinuous model solution. In order to numerically determine such finite difference ratios in a discontinuous nonlinear system, one explicitly computes the perturbations  $\Delta J$  for each input perturbation  $\Delta x_i$ . However, to deal with the spike of  $\Delta J/\Delta x_i$ , some new approach such as the generalized adjoint (Xu 1996a,b; 1997a,b) has to be considered.

### 3. Numerical results using a one-type cloud Arakawa–Schubert cumulus parameterization

Sensitivity examinations of various parameterizations in the NCEP model showed that a simplified (one-type cloud) Arakawa–Schubert (hereafter referred to as OTC-AS) cumulus parameterization scheme (Pan and Wu 1995) causes the largest discontinuity of the model response (Zhang 2000). We therefore choose the OTC-AS cumulus parameterization scheme as a discontinuous model operator to investigate the features of the TLM and adjoint model solutions.

#### a. Typical discontinuities

The original Arakawa–Schubert cumulus parameterization scheme (Arakawa and Schubert 1974) describes an ensemble of all-type clouds. In the implementation of the NCEP global model, the scheme was simplified as a one-type cloud (Pan and Wu 1995). In the simplified scheme, a simple cloud model is used to describe the thermodynamical properties, as illustrated schematically by Fig. 4. The one-type cloud definition includes the

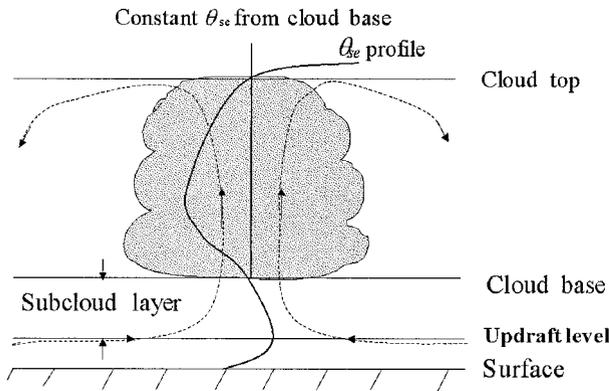


FIG. 4. A schematic diagram of the OTC-AS cumulus parameterization scheme.

thickness of the cloud/subcloud (updraft) layer that is greater/less than 150 hPa. This cloud/subcloud layer is determined by the vertical profiles of moist static energy and/or saturated moist static energy. Conditional instability defines the originating height of updraft (entrainment), where moist static energy reaches its maximum in the lower troposphere. At or above this height, a disturbed moist air parcel ascends adiabatically, due to buoyancy, until saturated where the cloud base is defined by lifting condensation level. Afterward, the moist air parcel will continuously rise along a constant moist static energy line (constant  $\theta_{se}$  line), assuming there is no heat exchange between the air parcel and the environment. The cloud top is defined when the vapor in the air parcel condenses completely.

Fillion and Errico (1997) studied the accuracy of linearized convection operator and the resulting impact on minimization using a column model relaxed Arakawa-Schubert scheme. Our study examines the features of the solutions of the tangent linear model and the adjoint model of the discontinuous OTC-AS parameterization. Generally speaking, any “if” statement in a parameterization scheme is able to cause discontinuities in model response. In the OTC-AS parameterization, obvious discontinuities occur when the following physical considerations are implemented in the computer code.

- 1) The convection process is turned on/off as the restriction of the thickness of the cloud layer and/or subcloud layer is reached/broken.
- 2) The convection process is turned on/off as the cloud work functions cross their critical values.
- 3) The level indices representing the cloud/subcloud base/top heights are shifted up/down if a small change occurs in temperature and moisture profiles.

In order to thoroughly investigate the impact of these discontinuities on the solutions of the TLM and the adjoint, we will examine the features of the nonlinear, tangent linear, and adjoint solutions of the OTC-AS parameterization in three cases. First, the input temperature at a single model-level of a sounding is changed

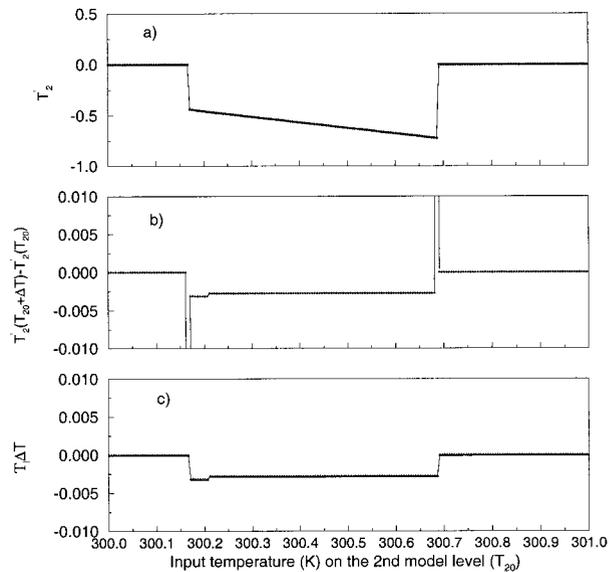


FIG. 5. Variation of (a) the temperature adjustment ( $T_2$ ) on the second model level using the OTC-AS parameterization scheme due to a perturbation of  $\Delta T = 0.005$  K added to the input temperature on the same level, (b) The nonlinear perturbation solution, and (c) the tangent linear approximation introduced by an input perturbation of temperature 0.005 K at the second model level.

(case 1). Second, the input temperature and moisture at all levels of a single sounding are varied (case 2). Finally, the input temperature and moisture are varied on multicolumns (case 3). The data used are extracted from the NCEP reanalysis dataset at 0000 UTC 1 November 1995.

*b. Features of the nonlinear and tangent linear solutions*

Figure 5 exhibits the numerical results of case 1 in which only the temperature at the second model level from the ground ( $\sigma = 0.9821$ ) of a single column located at and latitude 1.875°N and longitude 87°E [corresponding to model grid (43, 47) in a 62-wavenumber triangle truncation spectral model] is varied using an interval of 0.005 K. The other input variables (specific humidity, wind speed, vertical velocity, and the surface pressure) are kept unperturbed. The model solution, which is the adjusted temperature increment resulting from the OTC-AS parameterization at this model level, has two discontinuous points around  $T = 300.15$  K and  $T = 300.7$  K, where the OTC-AS cumulus convection process is turned on and off, respectively (Fig. 5a). Using 0.005 K as a small perturbation of the input temperature, we examine the features of the nonlinear and tangent linear perturbation solutions (Figs. 5b, 5c) resulting from a small perturbation made to each input temperature shown in Fig. 5a. The perturbations in Figs. 5b and 5c are identical, except at the discontinuity points. Figure 5a shows that the adjusted temperature varies linearly

in effect between the discontinuity points, so that the tangent linear model solution is nearly constant. At discontinuity points, Fig. 5b (nonlinear model solution) shows peaks in perturbations that are absent from Fig. 5c (TLM solution). These peaks correspond to cases where the unperturbed and perturbed input temperatures lie on different sides of the discontinuity. The tangent model is, of course, unable to produce the corresponding jumps in the adjusted temperature. However, except at the discontinuous points, the tangent linear model is as good for computing perturbations as the full model (with the first-order accuracy).

If more than one input variable is varied (cases 2 and 3, for instance), the following perturbing method and the response functions of the model (nonlinear and tangent linear) output are defined so that the relation of the nonlinear and tangent linear solutions and the input variable can still be presented graphically. The input temperature and specific humidity are varied by

$$\begin{aligned} \mathbf{T} &= \mathbf{T}_u + \alpha\Delta\mathbf{T} \\ \mathbf{q} &= \mathbf{q}_u + \alpha\Delta\mathbf{q}, \end{aligned} \tag{15}$$

where  $\Delta\mathbf{T} = \mathbf{T}_r - \mathbf{T}_u$  and  $\Delta\mathbf{q} = \mathbf{q}_r - \mathbf{q}_u$ . The subscripts  $u$  and  $r$ , respectively, denote the ‘‘unperturbed’’ state and ‘‘reference’’ state of temperature and specific humidity profiles. Again, the other input variables (wind speed, vertical velocity, and the surface pressure) are kept unperturbed. Once the unperturbed state and reference states are chosen, the parameter  $\alpha$  is varied to control the magnitude of the input perturbations  $\alpha\Delta\mathbf{T}$  and  $\alpha\Delta\mathbf{q}$  so that a multidimensional problem is converted to a one-dimensional problem that can easily be examined graphically.

The following three scalars,  $R_1$ ,  $R_2$ , and  $R_3$ , are defined using  $l_2$  Euclidean norm to represent the temperature and moisture increments due to cumulus convection ( $R_1$ ), the nonlinear perturbation of the output temperature and moisture variables caused by a small change in the input variables ( $R_2$ ), and the tangent linear solution ( $R_3$ ), respectively:

$$\begin{aligned} R_1(\alpha) &= 10^{-4}(\mathbf{T}^{\text{output}} - \mathbf{T}^{\text{input}})^T(\mathbf{T}^{\text{output}} - \mathbf{T}^{\text{input}}) \\ &\quad + 10^6(\mathbf{q}^{\text{output}} - \mathbf{q}^{\text{input}})^T(\mathbf{q}^{\text{output}} - \mathbf{q}^{\text{input}}) \\ &\equiv \left\| \begin{pmatrix} \mathbf{T}^{\text{output}} - \mathbf{T}^{\text{input}} \\ \mathbf{q}^{\text{output}} - \mathbf{q}^{\text{input}} \end{pmatrix} \right\|^2 \\ &= \|\mathbf{C}(\mathbf{T}_u + \alpha\Delta\mathbf{T}, \mathbf{q}_u + \alpha\Delta\mathbf{q}_r) \\ &\quad - \mathbf{l}(\mathbf{T}_u + \alpha\Delta\mathbf{T}, \mathbf{q}_u + \alpha\Delta\mathbf{q}_r)\|^2, \end{aligned} \tag{16}$$

where  $C$  represents the nonlinear operator of the OTC-AS parameterization,  $\mathbf{l}$  is a unity matrix, and the weighting coefficients used for the norm of temperature and specific humidity variables are  $10^{-4} \text{ K}^{-2}$  and  $10^6 \text{ kg}^2 \text{ kg}^{-2}$ , corresponding to typical orders of simulated temperature and specific humidity errors of  $1^\circ\text{K}$  and  $1 \text{ g kg}^{-1}$ ;

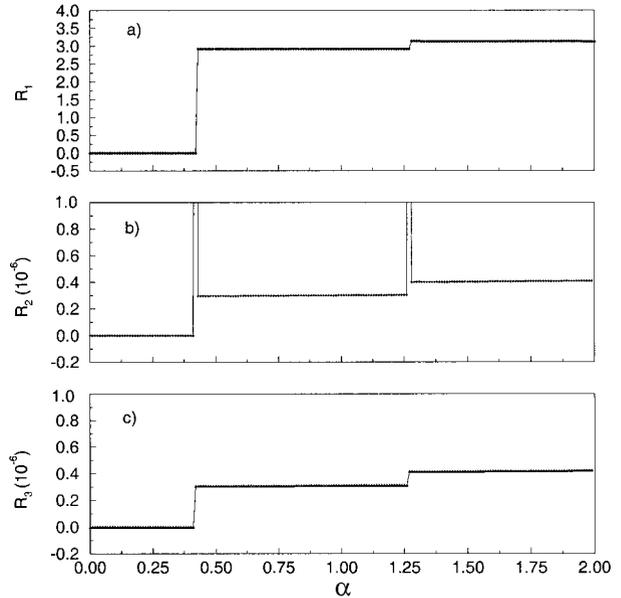


FIG. 6. Variation of the  $l_2$  Euclidean norm of (a) the adjusted temperature and specific humidity [ $R_1$ , see (16)] of the OTC-AS parameterization at a column around  $1.875^\circ\text{N}$ ,  $87^\circ\text{E}$  calculated at an interval of  $\Delta\alpha = 0.01$ , (b) the nonlinear perturbation solution produced by an input perturbation of  $\Delta\alpha = 0.01$  [ $R_2$ , see (17)], and (c) the corresponding tangent linear approximation to the nonlinear perturbation [ $R_3$ , see (18)].

$$R_2(\alpha) = \|C(\alpha) - C(0)\|^2 \tag{17}$$

where

$$C(\alpha) \equiv C(\mathbf{T}_u + \alpha\Delta\mathbf{T}, \mathbf{q}_u + \alpha\Delta\mathbf{q}), \tag{18}$$

$$R_3(\alpha) = \|\alpha\mathbf{C}(\alpha)\Delta\mathbf{x}\|^2, \tag{19}$$

where  $\mathbf{C}(\alpha) = \partial C(\alpha)/\partial\mathbf{x}$  is the tangent linear operator of the OTC-AS parameterization and  $\Delta\mathbf{x} = (\Delta\mathbf{T}, \Delta\mathbf{q})^T$ .

The features of various model solutions with respect to the multidimensional control variable can now be illustrated in a single variable space of  $\alpha$ .

For the numerical test conducted below, we choose a vertical column where the convection process is inactive [the model grid (44, 47) at  $1.875^\circ\text{N}$ ,  $87^\circ\text{E}$ ] as a basic state ( $\mathbf{T}_u$  and  $\mathbf{q}_u$ ), and another vertical column where the convection process is active [the model grid (43, 47) at  $1.875^\circ\text{N}$ ,  $85^\circ\text{E}$ ] as a reference state ( $\mathbf{T}_r$  and  $\mathbf{q}_r$ ). We could predict that the nonlinear model response has at least one discontinuous point between  $\alpha = 0$  and  $\alpha = 1$  when the convection process is turned on.

Figure 6a shows the distribution of  $R_1$  with respect to  $\alpha$  calculated at an interval of 0.01 (equivalent to a temperature interval of about  $0.005 \text{ K}$  and specific humidity interval of about  $0.05 \text{ g kg}^{-1}$ ). For each of these  $\alpha$  values, we then add a perturbation of  $\Delta\alpha = 0.01$  to the input variables of the OTC-AS cumulus parameterization scheme. Results of the nonlinear perturbation and tangent linear perturbation are shown in Figs. 6b and 6c, respectively. From Fig. 6a, we found that when

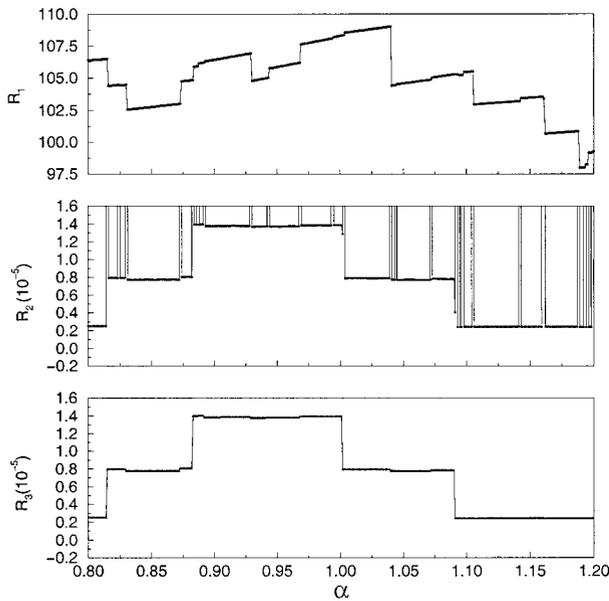


FIG. 7. Same as Fig. 6 except for all the 384 columns around latitude 3.75°N(S) and  $\Delta\alpha = 0.001$ .

$\alpha$  varies from 0 to 1 the convective process is activated at  $\alpha = 0.43$ , indicating the occurrence of discontinuity of the forward model solution in the phase space of  $\alpha$ . Another discontinuity occurs near  $\alpha = 1.28$  when the cumulus cloud top moved from the 12th to 13th model level. The results are qualitatively similar to those shown in Fig. 5. Outside and between discontinuous points,  $R_1$  varies in effect linearly, so that the tangent linear model represents perturbations as accurately as the nonlinear model (Figs. 6c and 6b, respectively). At discontinuity points, the tangent linear model is unable to represent the sharp jumps in  $R_1$ .

The above experiments were repeated by including all the columns (384 model grids) on two selected latitude belts, for example, a pair of Gaussian latitude circles symmetric with respect to the equator. The temperature and moisture profiles derived from the NCEP reanalysis on the pair of Gaussian latitude circles on 3.75°N(S) were used as  $\mathbf{T}_u$  and  $\mathbf{q}_u$  and those on 1.875°N(S) as  $\mathbf{T}_r$  and  $\mathbf{q}_r$  in (15). Numerical results of  $R_1$ ,  $R_2$ , and  $R_3$  are shown in Fig. 7. The nonlinear model solutions (Fig. 7a) contain many discontinuous points. The results are, again, very similar to those shown in Figs. 5 and 6. The tangent linear model accurately represents the (quasi linear) variations of  $R_1$  between discontinuous points, but does not represent the jumps at those points. These results are consistent with the examination using idealized examples (Xu 1996a,b; 1997a,b).

*c. Features of the adjoint model solutions*

In order to examine the features of the adjoint model solution of the OTC-AS scheme, we first define a cost function

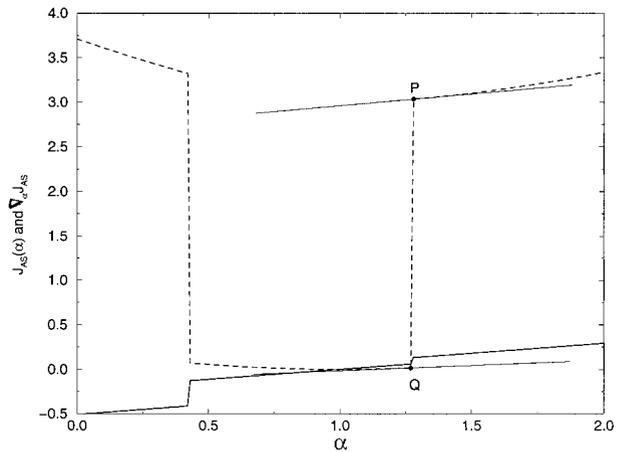


FIG. 8. Variation of the cost function ( $J_{AS}$ ) defined by (20) (thick dashed line), and the gradient ( $\nabla_{\alpha} J_{AS}$ ) (thick solid line) calculated from the adjoint of OTC-AS parameterization. A straight line crossing point  $P$  represents a local tangential slope derived from the adjoint. For graphing, the gradient is multiplied by a factor of 0.5. Both  $J_{AS}$  and  $\nabla_{\alpha} J_{AS}$  are calculated at an interval of  $\alpha = 0.01$ .

$$J_{AS}(\alpha) = \|C(\mathbf{T}_u + \alpha(\mathbf{T}_r - \mathbf{T}_u), \mathbf{q}_u + \alpha(\mathbf{q}_r - \mathbf{q}_u)) - \mathbf{l}(\mathbf{T}^{obs}, \mathbf{q}^{obs})\|^2 \tag{20}$$

where the norm  $\| \cdot \|$  is defined the same as in (16). We assume that the observations of the temperature and specific humidity ( $\mathbf{T}^{obs}$ ,  $\mathbf{q}^{obs}$ ) are the output of the parameterized convection process operating on the reference state. The cost function and its gradient are evaluated for different  $\alpha$  values with an interval of  $\Delta\alpha = 0.01$ . The corresponding numerical results are shown in Fig. 8 for both  $J_{AS}$  (dashed line) and  $\nabla J_{AS}$  (solid line). Within the range of  $\alpha$  from 0 to 2,  $J_{AS}$  has two discontinuity points (as described in Fig. 6). The gradient, calculated by the adjoint model integrations,  $\nabla J_{AS}$ , is also discontinuous at these two points. However, the slopes represented by the values of the adjoint-calculated gradient, for example, the thin solid lines crossing points  $P$  and  $Q$  represent the one-sided tangent directions of  $J_{AS}$  on both sides at a discontinuity point. Results in Figs. 8 and 9 are consistent with the theoretical results of the adjoint model integration obtained in section 2 using the chain rule method. They may also explain why the minimization with active discontinuous physics still performed reasonably well in the work of Zou and Kuo (1996), Fillion and Errico (1997), and Fillion and Mahfouf (2000).

Using the same data as Fig. 7 (case 3), we compute  $J_{AS}$  defined on 384 columns by (20) (thick-dashed line) and its gradient  $\nabla_{\alpha} J_{AS}$  (solid line) with different values of  $\alpha$  using the interval  $\Delta\alpha = 0.001$  to generate Fig. 9. In this case, the “observations” ( $\mathbf{T}^{obs}$  and  $\mathbf{q}^{obs}$ ) are the temperature and moisture profiles after the convective adjustment on 384 columns on the latitudes 1.875°N(S). Similar to Fig. 7a,  $J_{AS}$  contains many discontinuity points. Between two neighboring discontinuity points,

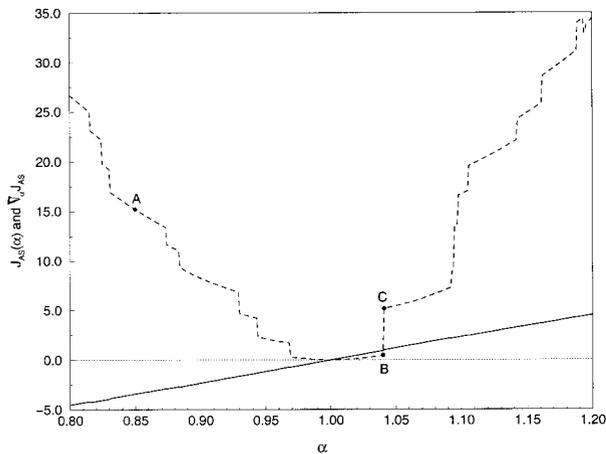


FIG. 9. Same as Fig. 8 except that the OTC-AS parameterization scheme is used on 384 columns around latitude  $3.75^{\circ}\text{N(S)}$  where the input state is varied through (15) using  $\Delta\alpha = 0.001$ ; observations are defined as the model solution of the OTC-AS parameterization on those columns around latitude  $1.875^{\circ}\text{N(S)}$  and a factor of  $10^{-2}$  used for graphing the gradient. A zero-gradient line (dotted line) is marked as a reference.

however,  $J_{AS}$  is continuous and differentiable. Therefore  $J_{AS}$  is piecewise differentiable on the whole  $\alpha$  domain. When the number of the discontinuity points increases, it becomes difficult to identify these differentiable pieces, and the cost function may appear to be oscillating with respect to a control parameter (see Figs. 1c,d). Compared to the variation of  $J_{AS}$ , the distribution of gradient derived from the adjoint integration is rather smooth, with only one stationary point. Obviously, the adjoint model integration of a discontinuous physical parameterization still provides either a tangential slope at a continuity point (point A, for instance) or a one-sided tangential slope at a discontinuity point (point B or C, for instance). It is also remarkable to notice that the variation of the gradient with respect to  $\alpha$  is in fact linear, which means that the variations of the cost function, in between its obvious discontinuous points, is quadratic, a nice feature that will make a minimization work well.

#### 4. Examining the gradients calculated using the adjoint of the NCEP diabatic global spectral model

##### a. Description of discontinuous physical processes and selection of control parameters

The NCEP global spectral model consists of a set of prognostic equations, including the tendency equations of vorticity, divergence, virtual temperature, specific humidity, and the surface pressure (Sela 1982, 1987). The model physics include large-scale precipitation, cumulus convection, shallow convection, vertical diffusion, gravity wave drag, and surface and boundary processes. The large-scale (grid scale) condensation consists of

condensation in supersaturated layers and evaporation in unsaturated layers. It adjusts both the temperature and specific humidity. The cumulus parameterization scheme is that described in section 3. The shallow convection is the Betts scheme (Betts 1986). A nonlocal vertical diffusion scheme is adapted (Hong and Pan 1996). The gravity wave drag parameterization is the Pierrehumbert (1986) scheme. The surface processes include a two-layer soil model. In the related processes the Monin–Obukhov similarity theory is applied to simulate the change of the surface (air–ground interface) temperature and humidity, and to calculate exchange coefficients and turbulence fluxes in the planetary boundary layer. These physical parameterization schemes are included in both the nonlinear and the adjoint models. Longwave and shortwave radiation are kept constant during the 6-h assimilation window.

The model dynamical core includes horizontal diffusion and time filtering in which the filter coefficient ( $\alpha_A$ ) is taken as  $1 - 2\epsilon$ , where  $\epsilon$  is the classical Asselin filter coefficient. The model carries out the time filter and horizontal diffusion at each time step integration. Although both the time filter coefficient  $\alpha_A$  and the horizontal diffusion coefficient  $\alpha_{HD}$  are not explicitly related to the model physical processes, any change in the values of these coefficients can affect model state and thus on–off switches in the parameterization schemes during the time integration. Therefore, these coefficients ( $\alpha_A$  or  $\alpha_{HD}$ ) play equivalent roles to  $\alpha$  in (15), where  $\alpha$  controls the magnitude of perturbation of a multidimensional control variable such as initial condition. Then, any cost function ( $J$ ) defined by the model solution and its gradient ( $\nabla J$ ) can be evaluated using a series of  $\alpha_A$  or  $\alpha_{HD}$  values rather than perturbing initial conditions so that the features of the  $J$  and  $\nabla J$  can be illustrated in  $\alpha_A$  or  $\alpha_{HD}$  space. Therefore, this study selects  $\alpha_A$  and  $\alpha_{HD}$  as the control variables to represent some numerical features of discontinuity in 4DVAR.

##### b. Examinations of the calculated gradient through the NCEP diabatic adjoint

In order to examine the gradient evaluated by the adjoint of a diabatic version of the NCEP model, we define the cost function using the model solution and the NCEP analysis as

$$J_{\text{model}}(\boldsymbol{\alpha}) = \frac{1}{2} \langle \mathbf{W}(\mathbf{x}^{6\text{-h forecast}} - \mathbf{x}^{\text{analysis}}), \mathbf{x}^{6\text{-h forecast}} - \mathbf{x}^{\text{analysis}} \rangle, \quad (21)$$

where  $\langle \cdot, \cdot \rangle$  represents the dot product, and the  $l_2$  Euclidean norm, of two vectors. Here  $\boldsymbol{\alpha}$  represents the control variable vector and we choose  $\boldsymbol{\alpha} = (\alpha_A, \alpha_{HD})^T$  in this case. The  $\mathbf{x}^{\text{analysis}}$  is the analysis state at 0600 UTC on 1 November 1995, representing observations  $\mathbf{y}_{\text{obs}}$  in (1), and  $\mathbf{x}^{6\text{-h forecast}}$  is the 6-h model forecast state  $[\mathcal{M}(\mathbf{x}_0)]$  starting from the initial state ( $\mathbf{x}_0$ ) at 0000 UTC on 1 November 1995. The model resolution is 62 waves (not

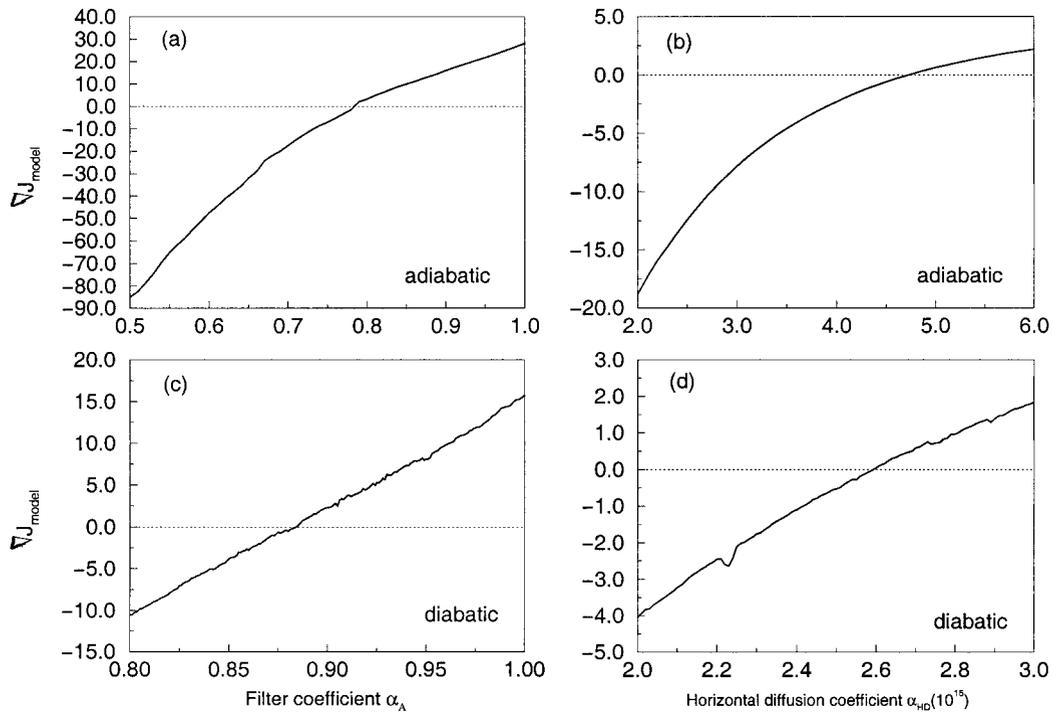


FIG. 10. Same as Fig. 1 except for gradient. A zero-gradient line is marked as a reference (dotted line) for each panel.

including the zonal mean) in horizontal domain and 28 vertical levels. The state vector  $\mathbf{x}$  consists of divergence ( $\mathbf{D}$ ), vorticity ( $\zeta$ ), virtual temperature ( $\mathbf{T}$ ), specific humidity ( $\mathbf{q}$ ), and the surface pressure ( $\mathbf{P}_s$ ), that is,  $\mathbf{x} = (\mathbf{D}, \zeta, \mathbf{T}, \mathbf{q}, \mathbf{P}_s)^T$ . Here the observational operator  $\mathbf{H}$  is a unity matrix in this case. Finally,  $\mathbf{W}$  is a diagonal weighting matrix approximated by the inverse of the absolute value of maximal 6-h differences of state variables. In this case, the diagonal values of  $\mathbf{W}$  are  $1.2 \times 10^{-6} \text{ s}^{-1}$  for divergence,  $2.4 \times 10^{-6} \text{ s}^{-1}$  for vorticity, 1 K for temperature, and 10 mb for surface pressure.

Variations of the cost function  $J_{\text{model}}$  defined by (21) and its gradient evaluated using the adjoint of the NCEP adiabatic (top panel) and diabatic (bottom panel) nonlinear forward models with respect to  $\alpha_A$  (left) and  $\alpha_{\text{HD}}$  (right) are shown in Figs. 1 and 10, respectively. The standard values of  $\alpha_A$  and  $\alpha_{\text{HD}}$  in the NCEP operational model are 0.92 and  $0.3 \times 10^{16}$ , respectively. Values of both  $J_{\text{model}}$  and  $\nabla J_{\text{model}}$  are computed at an interval of 0.001 for  $\alpha_A$  and  $0.001 \times 10^{16}$  for  $\alpha_{\text{HD}}$ . Comparing Figs. 1c,d with Figs. 10c,d, we found that, although the diabatic  $J_{\text{model}}$  appears jagged with respect to both coefficients due to the jumps in nonlinear model solution, the variation of the calculated gradients using the diabatic adjoint is not following the jagged aspect, but is consistent with the general convexity of  $J_{\text{model}}$ . There is only one point in the phase space at which the gradient is zero, a necessary condition for a minimum. These characteristics of the calculated gradient, on the one hand, confirm that an adjoint integration evaluates the one-

sided gradient at a discontinuity point as correctly as it does the gradient at a differentiable point. In fact,  $J_{\text{model}}$  in Figs. 1c,d can be viewed as an extension of the  $J_{\text{AS}}$  in Fig. 9, which is piecewise differentiable with many more discontinuities when all the discontinuous parameterization schemes are used in a global domain within a 6-h assimilation window (a total of 12 time steps using 30 min as the time integration step size). When the one-dimensional input space contains too many discontinuities, the cost function curve may appear “oscillating” with a finite interval of  $\alpha_A$  ( $\alpha_{\text{HD}}$ ) while the corresponding gradient curve appears relatively smooth. The phenomenon may be case dependent because there is no guarantee that the cost function, which is produced by the diabatic model, is only with zero-order discontinuity. In addition, we notice that the variations of the gradient are more linear (and the cost function must be more quadratic) in the diabatic case than in the adiabatic case, implying that the Hessian matrix is nearly constant in the diabatic case. Further research is of course needed to examine the nature of a diabatic cost function and clarify these issues, especially when the model resolution is increased.

### 5. Correctness check of the TLM and adjoint over multiphysical regimes in the presence of discontinuities

Traditionally, the correctness of the TLM and adjoint model is checked by examining the first-order approx-

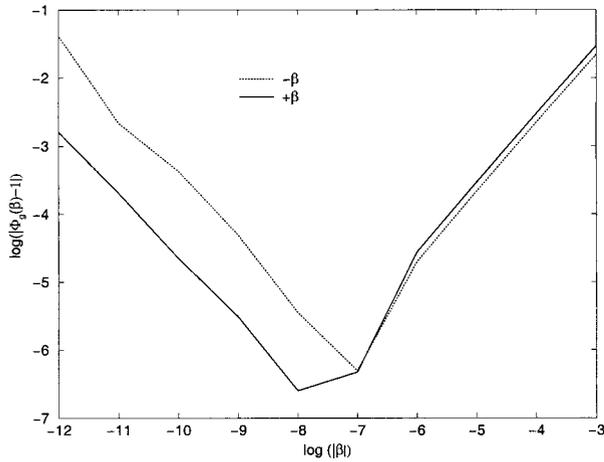


FIG. 11. The variation of  $\log_{10}(|\Phi_g - 1|)$  for the OTC-AS parameterization column model with  $-\beta$  (testing the left derivative) and  $+\beta$  (testing the right derivative) at a discontinuity point (point P in Fig. 8). The definition of the cost function and data used are same as Fig. 8. The basic perturbation is taken as  $\Delta\mathbf{T}$  and  $\Delta\mathbf{q}$  [see Eq. (15)].

iminations to a nonlinear perturbation of both the model solution (TLM test) and the cost function (gradient test). Typically, a ratio ( $\Phi_l$  for TLM test and  $\Phi_g$  for gradient test) is defined to measure the discrepancy between the nonlinear perturbation and its first-order approximation as

$$\begin{aligned} \Phi_l(\beta) &= \frac{\|\mathcal{M}(\alpha + \beta\delta\alpha) - \mathcal{M}(\alpha)\|}{\|\mathbf{M}\beta\delta\alpha\|} \\ &\times \left( \Phi_g(\beta) = \frac{J(\alpha + \beta\delta\alpha) - J(\alpha)}{(\nabla\alpha J)^T \beta\delta\alpha} \right) \\ &= 1 + O(\beta) \end{aligned} \tag{22}$$

(for a differentiable  $\mathcal{M}$  or  $J$ ),

where  $\|\cdot\|$  represents a  $l_2$  Euclidean norm,  $\mathcal{M}$  the nonlinear forward operator,  $\mathbf{M}$  ( $=\partial\mathcal{M}/\partial\alpha$ ) the tangent linear operator of  $\mathcal{M}$ ,  $J$  the defined cost function such as in (20) or (21),  $\nabla_\alpha J$  the gradient derived from an adjoint integration, and  $\beta$  is a scalar controlling the size of input perturbation. For a differentiable  $\mathcal{M}$  or  $J$ , once the basic state  $\alpha$  (usually an initial analysis) and the perturbation  $\delta\alpha$  (usually taken as  $\nabla J/\|\nabla J\|$ ) are chosen, the ratio is expected to approach 1 linearly, with increasing accuracy as  $\beta$  becomes progressively smaller. Assuming  $\Phi_l(\beta) = 1 + A\beta$  and  $\beta = 10^{-i}$ , the function  $\log(\Phi_l - 1)$  ( $=\log A - i$ ) is a linear function of  $i$ .

In section 2, we proved that the adjoint integration provides the one-sided gradient for the cost function defined by the parameterized physics at a discontinuity point. In order to examine the conclusion numerically, we use the gradient of  $J_{AS}$  [see (20)] evaluated by the adjoint of the OTC-AS parameterization column model at either side of a discontinuity point (points P and Q in Fig. 8) to conduct the gradient test. Figure 11 presents

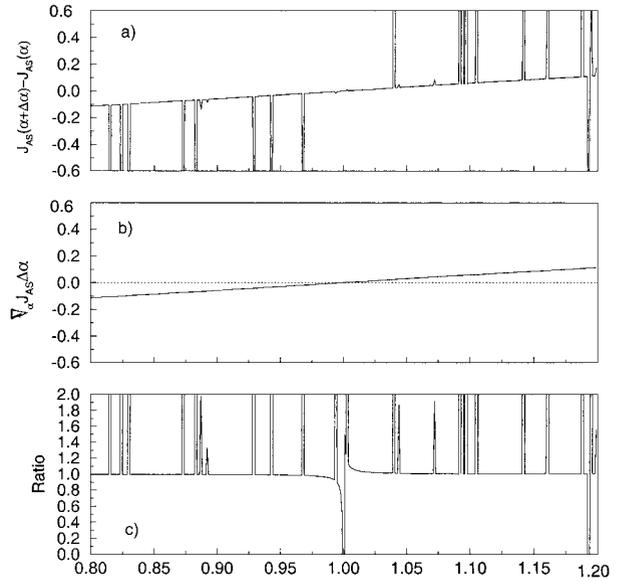


FIG. 12. Variation of (a) the nonlinear change of the perturbed cost function ( $J_{AS}$ ) produced by  $\Delta\alpha = 0.001$  [see (15)], (b) the leading order (linear) approximation derived from the adjoint of the perturbation of  $J_{AS}$ , and (c) the ratio of the nonlinear change and the linear approximation, with  $\alpha$ .

the variation of  $\log(|\Phi_g - 1|)$  with respect to  $\log|\beta|$  for a left ( $-\beta$ ) and right ( $+\beta$ ) perturbation [the basic perturbation is taken as  $\Delta\mathbf{T}$  and  $\Delta\mathbf{q}$ , see (15) for case 2]. It is found that, at the discontinuous point, the gradients computed by the adjoint of the OTC-AS parameterization are correct from either side, although the two gradients (at points P and Q in Fig. 8) are different.

Considering that the on-off switches in model parameterizations may produce lots of discontinuities in practice, and one does not always know where these discontinuities are, a test that conducts the computation of  $\Phi_g$  over a range of basic states may provide more information to verify the usefulness of an adjoint. With the same technique used in Fig. 7, we can check the feature of the adjoint solution with respect to a set of basic states defined by the different  $\alpha$  values. Figure 12 exhibits the features of the nonlinear change (the numerator of  $\Phi_g$  expression; Fig. 12a) and the first-order approximation (the denominator of  $\Phi_g$  expression; Fig. 12b) of a perturbed cost function using the OTC-AS parameterization and its adjoint, and their ratio ( $\Phi_g$ ; Fig. 12c). The cost function is defined by (20) and the basic state by (15) where the states on the latitudes of  $1.875^\circ\text{N(S)}$  and  $3.75^\circ\text{N(S)}$  are used as  $\mathbf{T}_r, \mathbf{q}_r$  and  $\mathbf{T}_u, \mathbf{q}_u$  through a 0.001 interval of  $\alpha$ .

Figure 12 shows that the ratio of nonlinear to linear perturbations derived from the adjoint (Figs. 12a and 12b, respectively) leads to conclusions similar to those obtained from Figs. 5–7. Between discontinuous points, the ratio is close to 1, meaning that the linear approximation derived from the adjoint accurately describes the nonlinear perturbations of the cost function. This is

true except in the vicinity of the point  $\alpha = 1$ , where the gradient is equal to 0 and the ratio becomes infinite. And, at discontinuous points, the linear approximation does not capture the sharp variations of the cost function  $J_{AS}$ .

## 6. Summary and conclusions

A diabatic forecast model including parameterized physics simulates the evolution of the atmosphere better than an adiabatic one does. Efforts have been made to incorporate parameterized physical processes into 4DVAR experiments to include satellite and radar measurements of rain rates and cloud properties and thus to enhance forecast skills. However, physical parameterizations include on-off switches that come from regime changes in the physical processes of the real atmosphere due to the numerical discretization. Many of them are hard to remove. Examples include the triggering of moist processes and different regimes in the surface processes. These on-off switches make the nonlinear model solution and the cost function defined in 4DVAR discontinuous.

By examining the numerical results from the tangent linear and adjoint operators of a simple idealized model, a simplified Arakawa-Schubert cumulus parameterization scheme, and the NCEP diabatic model, along with an analytical explanation for these calculated results, we illustrated that (i) the tangent linear approximation of a diabatic model is valid only in the differentiable interior of cells bounded in input spaces by discontinuity surface and that it could not represent the nonlinear perturbation growth when a perturbation crosses a discontinuity; (ii) the gradient calculated using the adjoint of a parameterization or a diabatic model correctly reflects a local tangential (or one-sided tangential) slope of the cost function, since the adjoint model integration represents a numerical implementation of the chain rule, using a complex NWP model as a compound transformation; and (iii) the test of the tangent linear and adjoint codes only using a basic state may be inadequate to ensure the correctness of codes in all physical regimes when the model is discontinuous and other checks based on different basic states may be recommended.

The calculated gradient using an adjoint model mainly serves as a descent direction in minimization (Navon and Legler 1987; Liu and Nosedal 1989; Zou et al. 1993a). A commonly used minimization algorithm, such as the limited-memory quasi-Newton algorithm, utilizes this information about local tangential slopes to define a search direction at each iteration. Therefore, the cost function in minimization always can be viewed as a smoothed version using a set of gradients. This smoothing process depends on the step size used in the line search and produces, somehow, a general convexity of the zigzag jumping cost function curve. The general convexity may mainly be governed by the adiabatic dynamics of the model. Therefore, the subgrid small-scale

processes described by the parameterized physics may not hurt the general convexity of the cost function, although they make the cost function jagged around. This explains, to some degree, why the minimization algorithm that was developed for minimizing differentiable cost functions may still perform well for most 4DVAR problems using a diabatic assimilation model. Examples include assimilation experiments using precipitation (Zou and Kuo 1996) and precipitable water (Kuo et al. 1996) using a diabatic mesoscale model and its adjoint, and 1DVAR minimization using the relaxed Arakawa-Schubert scheme (Fillion and Mahfouf 2000). However, under certain specific situations, the limited-memory quasi-Newton algorithm may fail to minimize a piecewise differentiable cost function (Zhang et al. 2000). In the case, a nonsmooth optimization algorithm, such as the bundle method (Lemaréchal 1978), may improve the minimization (Zhang et al. 2000). Even a nonsmooth optimization algorithm using subgradient is still not free of problems (Zhang et al. 2000). In such a situation, further test and thought on generalized adjoint (Xu 1996a,b; 1997a,b; Xu et al. 1998; Xu and Gao 1999) may be one of potentials to solve the problem.

This study only discussed the case of the zeroth-order discontinuities in which the cost function under consideration is itself discontinuous. Higher-order discontinuities of situations when the cost function is continuous, but one of its derivatives is discontinuous, may also be present in numerical models of the atmospheric flow. The investigation of the impact of higher-order discontinuities on variational data assimilation, especially for meso- and convection-scale analysis, requires more research work to be done.

This study only examined some of the numerical results from tangent linear and adjoint models of either simple models or a coarse-resolution global prediction model. The investigation of the behavior of the cost function and its minimization with discontinuities for meso- and convection-scale analysis is required and planned for future work.

*Acknowledgments.* The research is supported by NOAA Grant NA77WA0571 and NSF Grant ATM-9812729. The authors would like to thank Drs. E. Kalnay and J. Sela for their persistent support and encouragement, and Dr. Olivier Talagrand and Dr. Qin Xu for their suggestions that were useful for improving the original manuscript. Thanks go to two anonymous reviewers for thorough and helpful comments and suggestions.

## REFERENCES

- Arakawa, A., and W. H. Schubert, 1974: Interaction of a cumulus ensemble with the large-scale environment. Part I. *J. Atmos. Sci.*, **31**, 674–701.
- Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.*, **112**, 677–691.

- Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint equation—Part I. Numerical results. *Quart. J. Roy. Meteor. Soc.*, **113**, 1329–1347.
- Errico, R. M., 1997: What is an adjoint model? *Bull. Amer. Meteor. Soc.*, **78**, 2577–2591.
- Fillion, L., and R. Errico, 1997: Variational assimilation of precipitation data using moist convective parameterization schemes: A 1D-Var study. *Mon. Wea. Rev.*, **125**, 2917–2942.
- , and J.-F. Mahfouf, 2000: Coupling of moist-convective and stratiform precipitation processes for variational data assimilation. *Mon. Wea. Rev.*, **128**, 109–124.
- Hong, S. Y., and H. L. Pan, 1996: Nonlocal boundary-layer vertical diffusion in a medium-range model. *Mon. Wea. Rev.*, **124**, 2322–2339.
- Kuo, Y.-H., X. Zou, and Y.-R. Guo, 1996: Variational assimilation of precipitable water using a nonhydrostatic mesoscale adjoint model. Part I: Moisture retrieval and sensitivity experiments. *Mon. Wea. Rev.*, **124**, 122–147.
- Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
- Lemaréchal, C., 1978: Nonsmooth optimization and descent methods. International Institute for Applied System Analysis, Laxenburg, Austria, 25 pp.
- , and C. Sagastizabal, 1997: Variable metric bundle methods: From conceptual to implementable forms. *Math. Programm.*, **76**, 393–410.
- Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Math. Programm.*, **45**, 503–528.
- Mahfouf, J.-F., and F. Rabier, 2000: The ECMWF operational implementation of four-dimensional variational assimilation Part II: Experimental results with improved physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1171–1190.
- 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.
- Pan, H. L., and W. S. Wu, 1995: Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. NMC/NOAA/NWS Tech. Rep., 409, 40 pp.
- Pierrehumbert, R. T., 1986: An essay on the parameterization of orographic gravity wave drag. [Available from GFDL/NOAA, Princeton University, Princeton, NJ 08542.]
- Sela, J., 1982: The NMC spectral model. NOAA Tech. Rep. NWS 30, 36 pp.
- , 1987: The new T80 NMC operational spectral model. Preprints, *Eighth Conf. Numerical Weather Prediction*, Baltimore, MD, Amer. Meteor. Soc., 312–313.
- Xiao, Q., X. Zou, and Y.-H. Kuo, 2000: Incorporating the SSM/I-derived precipitable water and rainfall rate into a numerical model: A case study for the ERICA IOP-4 cyclone. *Mon. Wea. Rev.*, **128**, 87–108.
- Xu, Q., 1996a: Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.*, **53**, 1123–1142.
- , 1996b: Generalized adjoint for physical processes with parameterized discontinuities. Part II: Vector formulations and matching conditions. *J. Atmos. Sci.*, **53**, 1143–1155.
- , 1997a: Generalized adjoint for physical processes with parameterized discontinuities. Part III: Multiple threshold conditions. *J. Atmos. Sci.*, **54**, 2713–2721.
- , 1997b: Generalized adjoint for physical processes with parameterized discontinuities. Part IV: Problems in time discretization. *J. Atmos. Sci.*, **54**, 2722–2728.
- , and J. Gao, 1999: Generalized adjoint for physical processes with parameterized discontinuities. Part VI: Minimization problems in multidimensional space. *J. Atmos. Sci.*, **56**, 994–1002.
- , J. Gao, and W. Gu, 1998: Generalized adjoint for physical processes with parameterized discontinuities. Part V: Coarse-grain adjoint and problems in gradient check. *J. Atmos. Sci.*, **55**, 2130–2135.
- Zhang, S., 2000: Use of adjoint physics in 4D VAR with the NCEP Global Spectral Model. Ph.D. dissertation, The Florida State University, 185 pp.
- , X. Zou, J. Ahlquist, I. M. Navon, and J. G. Sela, 2000: Use of differentiable and nondifferentiable optimization algorithms for variational data assimilation with discontinuous cost functions. *Mon. Wea. Rev.*, **128**, 4031–4044.
- Zhu, Y., and I. M. Navon, 1999: Impact of key parameters estimation on the performance of the FSU spectral model using the full physics adjoint. *Mon. Wea. Rev.*, **127**, 1497–1517.
- Zou, X., 1997: Tangent linear and adjoint of “on-off” processes and their feasibility for use in 4-dimensional variational data assimilation. *Tellus*, **49A**, 3–31.
- , and Y.-H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.*, **124**, 2859–2882.
- , I. M. Navon, M. Berger, P. K. H. Phua, T. Schlick, and F. X. LeDimet, 1993a: Numerical experience with limited-memory quasi-Newton and truncated-Newton methods. *SIAM J. Optimization*, **3**, 582–608.
- , —, and J. G. Sela, 1993b: Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus*, **45A**, 370–387.
- Zupanski, D., 1993: The effects of discontinuities in the Betts–Miller cumulus convection scheme on four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Tellus*, **45A**, 511–524.