Tellus (1997), 49A, 3–31
Printed in UK – all rights reserved

TELLUS
ISSN 0280–6495

Tangent linear and adjoint of "on-off" processes and their feasibility for use in 4-dimensional variational data assimilation

By X. ZOU, *MPG/MMM/NCAR, P.O. Box 3000, Boulder, CO 80307-3000, USA*

(Manuscript received 21 July 1995; in final form 11 March 1996)

ABSTRACT

Problems in using the adjoint of moist physical processes that include "on-off" switches are analyzed via using both analytical examples and a numerical model. The analytic examples allow us to access such problems accurately since there exists the truth for both the model perturbation solution and the gradient of the cost function. The numerical model, which is the Penn State/NCAR nonhydrostatic mesoscale model version 5 (MM5), is used to test the "on-off" problems in a real-data environment. The SESAME 10–11 April 1979 storm case is chosen for the MM5 simulation since the forecast is strongly affected by moist diabatic processes. Analysis based on idealized continuous examples shows that if (i) the moist physics contains "on-off" switches, (ii) these switches are retained in the assimilation model, (iii) the tangent linear model (TLM) and adjoint model follow the same route as in the NLM for the basic state around which the linearization is carried out (a way of constructing the TLM and adjoint model of a discretized numerical model), a sufficient condition for such developed TLM and adjoint model is that the perturbation solution of the NLM or the tendency equation is continuous at switching points. Otherwise, the accuracy of the TLM and adjoint model depends on the relative magnitude of the jump caused by "on-off" processes with respect to other terms in the tendency equations. However, numerical experiments using a discretized model including cumulus parameterization show that such errors, if any, do not seem to impact the validity of the TLM and the adjoint model for use in rainfall assimilation. The presence of "on-off" switches increased the degree of the nonlinearity more than that of the loss of accuracy in TLM and adjoint model, the former makes the TLM solution different from the true perturbation solution after a few hours of model integration. Three 4DVAR experiments are carried out in order to test the performance of the minimization procedure using numerical models which possess different "on-off" properties. The 4DVAR experiments are carried out over a 3-h window in which the initial precipitation associated with the Wichita Falls tornado outbreak took place. Assimilation of the 3-h rainfall observation produces a close fit to the observed 3-h rainfall during the assimilation window. Combination of the 3-h rainfall with the wind, temperature, surface moisture, and precipitable water produces a significant improvement on the short-range rainfall prediction. The presence of the "bad behavior" of the Kuo scheme with convection being turned on and off every other time step does not cause the minimization to fail, and similar results are obtained as far as the convergence rate and improvement to the model forecast are concerned.

1. Introduction

Over the past few years, it has become apparent that variational assimilation schemes could be made practical. The National Meteorological Center (NMC) has implemented a three-dimensional variational data-assimilation system (3DVAR) (Parrish and Derber, 1992), and is currently testing a four-dimensional variational data assimilation (4DVAR) system (Navon et al., 1992;

Zupanski and Zupanski, 1995). The European Centre for Medium-Range Weather Forecasts (ECMWF) has put forth tremendous effort in developing the 4DVAR system, and is now testing both the 3DVAR and 4DVAR systems (Andersson et al., 1994). The Met Office in England is expected to finish the development of the basic 3DVAR system by mid-1996 and have a feasible 4DVAR by 1997 (Lorenc, 1995). The HIRLAM project, which is a joint effort among the several weather services in the Nordic countries, the Netherlands and Ireland for research and development in short-range mesoscale numerical weather prediction, has started to develop a 4DVAR system in order to better utilize the new sources of data that are becoming available via satellites and radars (Gustafsson and Lönnberg, 1995). The main trend of data-assimilation schemes in numerical weather prediction (NWP) is the 4DVAR, which makes the analysis and forecast algorithms more consistent than other schemes, such as multivariate optimal interpolation, nudging, and 3DVAR.

The objective of 4DVAR is to find the solution to the constraining forecast model that will best fit a series of observational fields distributed over some space and time interval. One possible measure of the fit, the cost function J, consists of a weighted least-square fit of the model forecast to the observations. The cost function can be, in general, defined as

$$J(\boldsymbol{x}_0') = \tfrac{1}{2}\boldsymbol{x}_0'^{\mathrm{T}} \boldsymbol{E}^{-1} \boldsymbol{x}_0' + \tfrac{1}{2}(\mathrm{CHL}(\boldsymbol{x}_0' + \boldsymbol{x}_{\mathrm{g}}) - \boldsymbol{o})^{\mathrm{T}} \times \boldsymbol{F}^{-1}(\mathrm{CHL}(\boldsymbol{x}_0' + \boldsymbol{x}_{\mathrm{g}}) - \boldsymbol{o}), \qquad (1)$$

where $\boldsymbol{x}_{\mathrm{g}}$ is the background estimation of the model state (short-range forecast), $\boldsymbol{x}_0' = \boldsymbol{x}_0 - \boldsymbol{x}_{\mathrm{g}}$, $\boldsymbol{x}_0$ is the analyzed variable at the initial time, $\boldsymbol{o}$ is a vector of the observations, $\boldsymbol{L}$ is a transform from analyzed variables to model variables, $\boldsymbol{C}$ is a transform from model variables to observational variables, $\boldsymbol{H}$ represents the numerical model integration, $\boldsymbol{E}$ is the short-range forecast error covariance matrix, and $\boldsymbol{F}$ is the observational error covariance matrix.

In order to find the minimum $\boldsymbol{x}_0'^{*}$ of the cost function defined in (1), an unconstrained minimization scheme is used to iteratively find $\boldsymbol{x}_0'^{*}$, which allows the model to produce the best fit to all available observations in a least-square sense. In order to do that, the values of the cost function J and the gradient of the cost function ∇J with respect to the initial condition (IC) are required. The value of J can simply be calculated by integrating the numerical model forward. It is prohibitive to calculate ∇J using finite-difference approximation due to the large dimension of the control variable. The use of the adjoint model made such problems possible to solve. The ∇J can be found by integrating the adjoint model backward, and is expressed as (Navon et al., 1992):

$$\nabla J = \boldsymbol{E}^{-1}\boldsymbol{x}_0' + \boldsymbol{L}^{\mathrm{T}}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{F}^{-1}\mathrm{CH}(\boldsymbol{L}(\boldsymbol{x}_0' + \boldsymbol{x}_{\mathrm{g}}) - \boldsymbol{o}) \qquad (2)$$

where $\boldsymbol{H}^{\mathrm{T}}$ is the adjoint model operator and $\boldsymbol{C}^{\mathrm{T}}\boldsymbol{F}^{-1}(\mathrm{CHL}(\boldsymbol{x}_0' + \boldsymbol{x}_{\mathrm{g}}) - \boldsymbol{o})$ is the forcing term determined by J.

Adjoint techniques were successfully used in meteorological data assimilation with different numerical models ranging from simple models (Le Dimet and Talagrand, 1986), adiabatic primitive equation models (Thépaut and Courtier, 1991; Navon et al., 1992; Zou et al., 1993a), to high-resolution simple models (Sun et al., 1991; Qiu and Xu, 1992). However, none of these models contain physical parameterization processes. In 1993, moist physical processes were added to a few 4DVAR systems using primitive equation models and some preliminary data-assimilation experiments were conducted (Zou et al., 1993b; Zupanski, 1993).

It is also generally recognized since the "Second Workshop on Adjoint Applications in Dynamic Meteorology" (held on 2–6 May 1994 in Visegrad, Hungary) that physical parameterization processes should be included in the 4DVAR system, i.e., the best model should be used for the real-data 4DVAR experiments. Since moist processes play a significant role in simulating various large-scale and mesoscale phenomena, they should be the first to be incorporated into the adjoint model. If an adiabatic or a simplified physics version of the NLM is used for real-data assimilation, there exist two problems: (i) the misfit between model solution and data will remain large due to the imperfect nature of the numerical prediction model; (ii) many satellite observations, such as radiances, rainfall measurements, and cloud properties, cannot be directly assimilated into the model. Developing, testing, and including the adjoints of various physical parameterization schemes in the 4DVAR system thus become a major task in the near future.

Parameterized moist processes often contain "on-off" switches. Since the adjoint model is initially derived for differentiable systems of equations, (Le Dimet and Talagrand, 1986), the following problems should be addressed before a full-physics adjoint model is employed: How do we deal with these "on-off" switches in adjoint models? What do we obtain by integrating the adjoint model backward in time with the forcing terms $\partial J/\partial x(t_r)$ (Navon et al., 1992) added to its right-hand side? How does the minimization perform?

Several papers touched on the problems related to "on-off" processes for the TLM and adjoint model (Douady and Talagrand, 1990; Vukicevic and Errico, 1993; Zou et al., 1993b; Zupanski, 1993; Verlinde and Cotton, 1993; Bao and Warner, 1993; Zupanski and Mesinger, 1995; Bao and Kuo, 1995; Xu, 1996). Vukicevic and Errico (1993) tested the accuracy of the TLM of a mesoscale model, against the "true" perturbation obtained by direct nonlinear integration and concluded that significant errors may be expected in the regions where the moist diabatic processes are important for finite perturbations in IC. Zou et al. (1993b), Zupanski (1993), Verlinde and Cotton (1993), and Bao and Warner (1993) applied the adjoint of moist processes to 4DVAR using various numerical models ranging from a simple one-dimensional cloud model to complicated primitive-equation models. Zou et al. (1993b) showed how to develop adjoints of a Kuo–Anthes type cumulus parameterization scheme and a large-scale precipitation scheme containing "on-off" switches. The adjoint version of moist parameterization was constructed by linearizing the nonlinear model around the basic state at every time step while keeping the "on-off" switches the same as in the nonlinear model, and then transposing the TLM. They showed that the minimization procedure converged at a similar rate as the one with the adiabatic version of the NMC model without any difficulty. Zupanski (1993) and Zupanski and Mesinger (1995) eliminated some selected switches by applying a vertical smoothing to the layers of transition in the Betts–Miller cumulus convective scheme before it was used for the 4DVAR experiments, and the other "on-off" switches were treated same as in Zou et al. (1993b). Verlinde and Cotton (1993) completely eliminated the switches before developing the adjoint model to avoid problems caused by the switches. After these modifications and approximations applied to the moist processes, they found that the adjoint technique worked well with TLM and adjoint model linearized about the basic-state solution at every time step, although the physical process modeled might change rapidly when processes turned on and off. These studies demonstrated that minimization still converged with "on-off" switches appearing during the model integration.

In 4DVAR, the adjoint model is used to derive the sensitivity — values of gradient of the cost function to changes in IC, lateral boundary condition (LBC), and model parameters. The use of adjoint model for sensitivity analysis appeared as early as the 1940's. It was seeded in the earliest works on sensitivity analysis of reactor physics problem. Later on, Cacuci (1981) presented an adjoint sensitivity theory for a class of discretized nonlinear systems to treat problems involving discontinuous state function. It was shown that an adjoint-based procedure can be rigorously formulated, provided that G-differential of the nonlinear operators describing the physical problem (model equation, lateral boundary condition, cost function) exist. The necessary and sufficient condition in order that a nonlinear operator admit a G-differential that is linear in perturbation was also provided.

Therefore, having "on-off" switches in the assimilation model may not be a problem for a rigorous formulation of the adjoint-based sensitivity calculation. In the practical coding of an adjoint model of a discretized numerical weather prediction including physics, the "on-off" switches are kept the same as in the basic state (the reference state), if the switches in the original model are to be retained for data assimilation (Zou et al., 1993b; Zupanski, 1993; Vukicevic and Errico, 1993). Recently, both Bao and Kuo (1995) and Xu (1996) made a detailed study using idealized continuous examples with delta function mimicking the "on-off" switches in physical schemes. They indicated that ignoring the variation of the switch point due to the perturbation in IC, i.e., keeping the switching point in the TLM the same as in the basic state, could cause significant errors in TLM solution and gradient calculation. There is thus a need to re-examine the appropriateness of such treatment of "on-off" switches for adjoint development of a discretized model.

In this paper, both simplified examples similar to those used by Xu (1996), and realistic storm convection simulated by MM5 will be used to analyze the "on-off"-related problems in more details, and to examine the feasibility of using the tangent linear and adjoint of moist processes for 4DVAR. The simplified examples make such analysis very accurate since an analytic solution of the first order variation exists. The TLM and adjoint of MM5 including "on-off" switches allows us to examine such problems in a realistic context. Specifically, the following problems are considered: (i) What kind of information are the TLM and adjoint models able to provide if the "on-off" switches are kept the same as in the basic-state solution? (ii) What modifications are needed so that these "on-off" switches could contribute to the change of model prognostic variables in space and time? (iii) Does the "on-off" switch impact the accuracy of the TLM solution as the first-order approximation to the nonlinear perturbation, as opposed to how good the TLM solution approximates the nonlinear perturbation? (iv) How do we check the accuracy of the TLM and adjoint models in the presence of "on-off" switches? (v) Can we use a less frequent basic state-update frequency instead of every time step when the model contains "on-off" switches? (vi) Are the gradient values calculated by adjoint model integration able to provide the minimization process with a good descent direction?

In the simple examples, "on-off" switches are introduced through a Heaviside unit-step function, one with a constant increment and the other a variable (model-dependent) increment added to the model equation when the moist processes are turned on. In both examples, the switch time is assumed to be model-dependent. In MM5, the "on-off" switches are introduced through two moist processes: (i) the Kuo cumulus parameterization; and (ii) the resolvable-scale precipitation. The special data sets collected in Project SESAME 1979 are used to provide IC and lateral boundary conditions (LBC) for model simulation and observation data in 4DVAR experiments.

The following section describes the MM5 modeling system, the 4DVAR problems, and the minimization procedure. The MM5 adjoint model, the gradient calculation using the adjoint model, and a generalized definition of the derivative of the cost function are presented in Section 3. The analytical derivations using simple examples are provided in Section 4. Section 5 shows numerical results including MM5's TLM test and 4DVAR experiments, while the summary and conclusions are given in Section 6.

2. Numerical model

2.1. Dynamics of MM5

The TLM and adjoint model of the nonhydrostatic version of MM5 have recently been developed. The nonhydrostatic model eliminates the limitation on the resolution inherent in a hydrostatic model. It can thus be used to investigate meteorological phenomena with scales ranging from a few to several hundred kilometers. The model is based upon a set of equations for a fully compressible atmosphere in a rotating frame of reference. It uses a terrain-following σ vertical coordinate. For more detailed descriptions of the nonhydrostatic model, see Dudhia (1993).

The version of MM5, the TLM and adjoint model (Zou et al., 1995, Kuo et al., 1995), includes model dynamics, diffusion, bulk planetary-boundary-layer processes, surface friction, a semi-implicit time-split integration scheme, dry convective adjustment, a cumulus parameterization scheme, and resolvable-scale precipitation processes. Since MM5 is a limited-area model, the adiabatic MM5 adjoint is developed with the flexibility of controlling either IC or LBC, or both.

For the convenience of the follow-up discussions, we will describe the Kuo cumulus parameterization scheme first. The purpose is for the ease of clarifying some problems related to this scheme and to supply a base for a critical evaluation of its suitability for use under various circumstances.

2.2. Kuo cumulus parameterization scheme

Cumulus parameterization is an attempt to formulate the collective effect of cumulus (subgrid-scale) clouds in terms of the prognostic variables of grid scale. The large-scale potential-temperature and water-vapor budget equations may be written, in a general form, as:

$$\frac{\partial T}{\partial t} = \left(\frac{\partial T}{\partial t}\right)_{\mathrm{LS}} + \frac{1}{c_{\mathrm{p}}} Q_1 \qquad (2.1)$$

$$\frac{\partial q}{\partial t} = \left(\frac{\partial q}{\partial t}\right)_{\mathrm{LS}} - \frac{1}{L} Q_2, \qquad (2.2)$$

where the subscript LS denotes the contribution to the time derivative by large-scale advective processes, and Q_1 and Q_2 are the apparent heat source and apparent moisture sink, respectively. All the other symbols are standard. The assumption of the Kuo scheme, which is widely used in NWP, is based on the coupling of the Q_1 and Q_2 terms with the $(\partial T/\partial t)_{\mathrm{LS}}$ and $(\partial q/\partial t)_{\mathrm{LS}}$ terms.

Three checks are performed to impose realistic limits on the types of columns that are capable of supporting convection. The first criterion is defined through moisture convergence:

Criterion 1

$$M_{\mathrm{t}} \equiv \left(\frac{m^2}{g}\right) \int_0^1 \left(\frac{\nabla \cdot (p^* v q)}{m}\right) \mathrm{d}\sigma$$
$$\geqslant 3 \times 10^{-7}\ (\mathrm{cb\ s^{-1}}), \tag{2.3}$$

where $\boldsymbol{v}$ is the horizontal wind vector, $\nabla \cdot$ represents the divergence calculation, p^* is defined as $p_{\mathrm{s}} - p_{\mathrm{t}}$, where p_{s} is the surface pressure, and p_{t} is the pressure at the top of the model (100-hpa). The value of $3.0 \times 10^{-7}\ \mathrm{cbs^{-1}}$ is the precipitation threshold for moisture convergence. The second criterion is the cloud thickness check:

Criterion 2

$$\sigma_{k_{\mathrm{Ctop}}} - \sigma_{k_{\mathrm{Cbase}}} \geqslant 0.3, \tag{2.4}$$

where k_{Ctop} and k_{Cbase} are cloud top and cloud base, respectively. The third criterion is related to the stability check.

Criterion 3

$$\sum_{k=k_{\mathrm{Cbase}}}^{k_{\mathrm{Ctop}}} ((\theta_{\mathrm{s}})_k - (\theta_{\mathrm{e}})_k)\Delta\sigma_k \geqslant 0, \tag{2.5}$$

i.e., there is a positive buoyant energy in the cloud.

The procedures for calculating the cloud base (k_{Cbase}) and cloud top (k_{Ctop}) are provided in Appendix A.

If all three criteria are satisfied, then convective precipitation occurs. The moisture convergence is then partitioned into a rain-producing portion $(1-b)M_{\mathrm{t}}$ and a humidity increasing portion bM_{t}, where b is calculated from the column-integrated relative humidity:

$$b = 1 - \frac{\bar{q}}{\bar{q}_{\mathrm{s}}}. \tag{2.6}$$

The cumulus clouds are assumed to exist only momentarily, and they dissipate by mixing with the environmental air at the same level. The heat and moisture carried out by the cloud air are imparted to the environmental air on the basis of the vertical distribution of an empirical normalized vertical profile of convective heating ($N_{\mathrm{h}}(\sigma)$) and moistening ($N_{\mathrm{m}}(\sigma)$), and a vertical eddy-flux divergence of water vapor associated with cumulus convection ($V_{\mathrm{qf}}(\sigma)$) by the following relations:

$$T_k^* = T_k + \frac{L_{\mathrm{v}}}{c_{\mathrm{p}}}(1-b)gM_{\mathrm{t}}N_{\mathrm{h}}(\sigma_k), \tag{2.7}$$

$$q_k^* = q_k + bgM_{\mathrm{t}}N_{\mathrm{m}}(\sigma_k) + V_{\mathrm{qf}}(\sigma_k), \tag{2.8}$$

$$P^* = (1-b)M_{\mathrm{t}}, \tag{2.9}$$

where T_k^* and q_k^* are the temperature and specific humidity after mixing.

For expressions of $N_{\mathrm{h}}(\sigma)$, $N_{\mathrm{m}}(\sigma)$, and $V_{\mathrm{qf}}(\sigma)$, see Grell et al. (1994).

Many case studies show that the Kuo scheme performs reasonably well, although it is not the best one among the seven cumulus parameterizations available in MM5. Since this scheme is simple and computationally inexpensive while possessing typical "on-off" switches as in most moist parameterization schemes, it is worthwhile to give a detailed analysis of its tangent linear and adjoint versions and their validity for use in 4DVAR experiments.

2.3. *Mathematical representation of the 4DVAR problem*

To be definite $\boldsymbol{x}(t)$ is considered to be the unique solution of the physical problem described by MM5. This requirement is usually fulfilled (or assumed to be fulfilled when rigorous existence and uniqueness proofs are lacking) in most NWP problems of practical interest, where a discretized numerical model is used. In 4DVAR, a minimization problem is solved. This problem arises in the course of attempting to estimate certain parameters (IC, LBC, or model parameters) in a functional relationship by means of observational data. Suppose, for example, that some quantity y is measured at various times, say, $t_1, \ldots, t_m$; it is desired to estimate the vector x_0 of the model state at the initial time t_0. If these measurements are exact and the model is perfect, then the parameter vector $\boldsymbol{x}_0$ would satisfy $\boldsymbol{C}\boldsymbol{x}(t_r) = \boldsymbol{y}_r$, $r = 1, \ldots, m$, which is a system of m equations in the n unknowns $\boldsymbol{x}_0 = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$, where $\boldsymbol{C}$ is an operator projecting the model forecast to observational quantities. However, in general, the $\boldsymbol{y}_r$ will be subject to error, and the model solution $x(t)$ will be subject to model error. Then a standard procedure is to take more measurements than the number of unknowns, so that $m > n$, and to find x_0 so as to minimize the sum of the weighted squares of the deviations $Cx(t_r) - y_r$; that is, we seek to minimize the functional J: $R^n \to R^1$ defined by

$$J(x_0) = \sum_{r=1}^{m} W(Cx(t_r) - y_r)^2, \tag{2.12}$$

where W is a weighting matrix proportional to the inverse of the error covariance of observations. If J is differentiable, then it is known from calculus that the partial derivative of J must all vanish at x_0^*; that is:

$$g_i \equiv \frac{\partial}{\partial x_i} J(x_0) = 0, \quad \text{when } x_0 = x_0^*, \qquad i = 1, \ldots, n. \tag{2.13}$$

Therefore, the problem of locating the minimizer leads naturally to the solution of systems of equations.

In meteorology, usually $n > m$. A background term (the first term in (1)) is thus included in the cost function, which possesses prior information. Having defined a cost function, the 4DVAR problem is solved by a minimization procedure. An essential concept in the minimization algorithm is that of a measure of progress. For an iterative method, it is important to have some reasonable ways of deciding whether a new IC is "better" than the old IC. A natural measure of progress is provided by the value of the cost function. It seems reasonable to require a decrease of J at every iteration; i.e., $J(x_0^{(k+1)}) < J(x_0^{(k)})$ for all $k \geqslant 0$. This requirement is imposed by all descent methods for unconstrained minimization problems, and has the following common structure:

(1) choose a guess IC $x_0^{(0)}$ and set $k = 0$;

(2) compute a search direction d_k;

(3) compute a step length α_k, for which it holds that $J(x_0^{(k)} + \alpha_k d_k) < J(x_0^{(k)})$;

(4) update the estimate of the minimum: set $x_0^{(k+1)} = x_0^{(k)} + \alpha_k d_k$, $k = k + 1$;

(5) test convergence. If the conditions for convergence are satisfied, the algorithm terminates with $x_0^{(k)}$ as the solution; otherwise, go to step (2).

Since J is bounded below, if $J(x_0^{(k+1)}) < J(x_0^{(k)})$, then clearly the sequence $\{J_k\}$ converges. However, the requirement that $J(x_0^{(k+1)}) < J(x_0^{(k)})$ is not sufficient in itself to ensure that the sequence $\{x_0^{(k)}\}$ converges to a minimum of J, x_0^*. There are two obvious situations in which the decreasing sequence $\{J_k\}$ might converge to a non-optimal point:

(i) whatever the choice of search direction d_k, the sequence of step lengths $\{\alpha_k\}$ could be chosen so that J is reduced by an ever-smaller amount at each iteration;

(ii) d_k is almost orthogonal to the gradient $g_k = \nabla J$ and J is almost constant along d_k.

Therefore, to avoid these two situations, α_k must be chosen so that J is "sufficiently decreased" at each iteration, and there must be a limit on the closeness of d_k to the negative gradient direction to avoid its orthogonality to the gradient. The limited-memory quasi-Newton method of Liu and Nocedal (1989), which is used in this paper, calculates the search direction by

$$d_k = -H_k g_k, \tag{2.14}$$

where the matrix H_k is an approximation of the inverse Hessian matrix obtained through an updating cycle using current and previous gradient information (for details, see Zou et al., 1993c). Such a choice of d_k prevents its orthogonality to g_k. The step lengths ($\{\alpha_k\}$, $k = 1, 2, \ldots$) are obtained by using the cubic interpolation method to satisfy the so-called Wolf conditions:

$$J(x_0^{(k)} + \alpha_k d_k) \leqslant J(x_0^{(k)}) + \beta_1 \alpha_k g_k^{\mathrm{T}} d_k \tag{2.15}$$

and

$$|g(x_0^{(k)} + \alpha_k d_k)^{\mathrm{T}} d_k| \leqslant -\beta_2 g_k^{\mathrm{T}} d_k, \tag{2.16}$$

where $\beta_1 = 0.0001$ and $\beta_2 = 0.9$. The first criterion prevents the possibility of J being reduced by an ever-smaller amount at each iteration, and the second criterion ensures that α_k will be a close approximation to a minimum of J along d_k.

Examining the entire procedure of finding the minimum of J, we find that both the values of J_k and the derivative g_k are required at each iteration to find the search direction and step length. J can be calculated by integrating the model forward, and the gradient g_k is obtained through adjoint-model backward integration with proper forcing terms added to its right-hand side (see Subsection 3.2). Therefore, the concern in 4DVAR with the moist physics possessing "on-off" switches in the adjoint model is really related to whether the adjoint technique can provide sufficient information for minimization-finding descent directions.

3. Adjoint model and gradient calculation

3.1. MM5 adjoint model

The discretized MM5 can be symbolically written as

$$\boldsymbol{x}(t_n) = \boldsymbol{Q}_n^0(\boldsymbol{x})\boldsymbol{x}(t_0), \tag{3.1}$$

where $\boldsymbol{x}(t_0) = (u, v, T, q, p', w)|_{t=t_0}$ represents a vector of IC for the model integration, $\boldsymbol{x}(t_n)$ represents the model forecast at time t_n, and $\boldsymbol{Q}_n^0$ represents all the operator matrices in MM5 to obtain $\boldsymbol{x}(t_n)$ from IC $\boldsymbol{x}(t_0)$. Here we assume that the LBC which specifies values of the model variables at the boundaries of the model domain will not change during the data-assimilation procedure.

The TLM was obtained by linearizing the MM5 around the NLM solution, and can be written as:

$$\boldsymbol{x}'(t_n) = \boldsymbol{P}_n^0(\boldsymbol{x})\boldsymbol{x}'_o, \tag{3.2}$$

where $\boldsymbol{x}'$ and $\boldsymbol{P}$ represent the perturbed model solution and the TLM's operator, respectively. For those "on-off" switches (if statement depending on model variables), we use the basic state values to determine whether they are switched on or off.

The adjoint model is simply the transpose of the TLM (assuming that an Euler norm is used for inner product and $\boldsymbol{P}_n^0$ is a real matrix), which can be written as

$$\hat{\boldsymbol{x}}^{(n)}(t_0) = (\boldsymbol{P}_n^0(\boldsymbol{x}))^{\mathrm{T}}\hat{\boldsymbol{x}}(t_n), \tag{3.3}$$

where $\hat{\boldsymbol{x}}$ and $(\boldsymbol{P}_n^0)^{\mathrm{T}}$ represents the adjoint variables and the adjoint model's operator, respectively. The term $\hat{\boldsymbol{x}}(t_n)$ on the right-hand side of the adjoint model (3.3) is the initial condition for the adjoint model, but is often called the forcing term in the practical application where adjoint model is used for the gradient calculation. The value of $\hat{\boldsymbol{x}}(t_n)$ depend on the specific problem one solves. However, the existence of the adjoint model does not depend on any specific application.

In 4DVAR, the NLM (3.1) is integrated forward in time to provide the value of the cost function, and the adjoint model (3.3) is integrated backward to provide the value of the gradient of the cost function with respect to IC. For different definitions of the cost function, one needs to supply proper forcing terms $\hat{\boldsymbol{x}}(t_n)$ for the adjoint-model backward integration. Therefore, the adjoint-model development involves the two stages: (i) differentiation to obtain the TLM, and (ii) transpose to obtain the adjoint model. The correctness of the TLM can be checked against the NLM, and the correctness of the adjoint model is verified against the TLM.

For a model without "on-off" switches, the linearization around the basic state — the NLM solution — can be carried out exactly without any approximations required except some modification required purely due to computational reasons. For example, for $x = \ln q$, $q > 0$, in the TLM it becomes $x' = q'/q$, $q > \varepsilon$, where ε is a small number equal to zero in the range of machine accuracy. However, when the forward model includes "on-off" switches, which depend on the model state, how to linearize the NLM at switching points becomes ambiguous. One common used method (Zou et al., 1993b, Zupanski, 1993; Vukicevic and Errico, 1993; Bao and Warner, 1993) is to keep the "on-off" route the same as in the NLM solution. This is carried out under the assumption that a very small perturbation in IC will not change when and where the "on-off" switches take place. However, the validity of such approximate treatment to the "on-off" switches have not been rigorously analyzed. That the adjoint model integration seems to provide an accurate enough first-order information of the cost function is experimentally verified by all the 4DVAR conducted in the papers by Zou et al. (1993b), Zupanski (1993), (1993), and Bao and Warner (1993). Recently, a detailed analysis using idealized examples is given first by Bao and Kuo (1995) and then by Xu (1996) and it was pointed out that such treatment of the "on-off" switches in the TLM and adjoint model could be a problem in expecting them to provide gradient information. A careful examination of the feasibility of using such developed adjoint model for 4DVAR experiment is thus required.

3.2. Gradient calculation through adjoint techniques

The gradient of the cost function defined in (2.12) with respect to IC can be obtained through the use of adjoint equations for the forecast model. As shown in Navon et al. (1992), the gradient of the cost function can be expressed as

$$\nabla J(\boldsymbol{x}(t_0)) = \sum_{r=0}^{R} \hat{\boldsymbol{x}}^r(t_0), \tag{3.4}$$

where the values of the adjoint variables $\hat{x}^r(t_0)$ $(r = 0, 1, ..., R)$ are obtained through adjoint model integration:

$$\hat{x}^r(t_0) = (P_r^0)^T C^T W(t_r)(x(t_r) - y^{obs}(t_r)). \tag{3.5}$$

The eqs. (3.4) and (3.5) were obtained under the assumption that the TLM provides the first-order approximation to the nonlinear perturbation. Conceptually, (3.4) and (3.5) can still be used for a model containing "on-off" switches. However, since the gradient of J is discontinuous at switching points, a more general definition of the derivative of the cost function is required to understand the result of adjoint model integration.

3.3. *A generalized definition of the derivative of J*

The concept of the derivative to functionals can be generalized in several ways, such as the Frechet differential, the Gateaux differential (G-differential) and the Gateaux variation (see Sagan, 1969). They are defined as follows.

Frechet differential $L_f[h]$

$L_f[h]$ is called the Frechet differential of $J(x_0)$ at $x_0^{(k)}$ if and only if there exists a $\delta > 0$ such that

$$J(x_0^{(k)} + h) - J(x_0^{(k)}) = L_f[h] + \varepsilon_1(h)$$
$$\text{for all } \|h\| < \delta, \tag{3.6}$$

where $L_f[h]$ is a linear function of h and where

$$\lim_{\|h\| \to 0} \frac{\varepsilon_1(h)}{\|h\|} = 0.$$

Here, h is a small perturbation to the initial condition $x_0^{(k)}$.

***G*-differential**

$L_g[h]$ is called the G-differential of $J(x_0)$ at $x_0^{(k)}$ if and only if there exists a $\delta > 0$ such that

$$J(x_0^{(k)} + h) - J(x_0^{(k)}) = L_g[h] + \varepsilon_2(h)$$
$$\text{for all } \|h\| < \delta, \tag{3.7}$$

where $L_g(h)$ is a linear function of h and where

$$\lim_{\alpha \to 0} \frac{\varepsilon_2(\alpha h)}{\alpha} = 0.$$

Gateaux variation

$VJ[h]$ is called the Gateaux variation of $J(x_0)$ at $x_0^{(k)}$ if and only if there exists a $\delta > 0$ such that

$$J(x_0^{(k)} + h) - J(x_0^{(k)}) = VJ[h] + \varepsilon_3(h)$$
$$\text{for all } \|h\| < \delta, \tag{3.8}$$

where

$$\lim_{\alpha \to 0} \frac{\varepsilon_3(\alpha h)}{\alpha} = 0.$$

We see that the concept of the G-differential is somewhat weaker than the concept of both the Frechet differential and the derivative (the conventional first-order derivative for differentiable function), since for the Frechet differential and the derivative, $\varepsilon(h)$ or $\varepsilon_1(h)$ has to tend to zero uniformly in h, while in the case of the G-differential or Gateaux variation, $\varepsilon_2[h]$ only has to tend to zero along each h. It is also important to note that J does not need to be continuous with respect to x_0 for the existence of the Gateaux variation, and the Gateaux variation is not even required to be linear in h.

By considering a typical nonlinear problem as a mapping defined on a space R^n, Cacuci (1981) indicated that the adjoint procedure (3.4) and (3.5) can still be formulated rigorously if the Gateaux variation exists (Notice that the G-differential in Cacuci (1981) is equivalent to the Gateaux variation defined here). This completely eliminates the need for the existence of the derivatives of the model state with respect to IC. Therefore, when models contain "on-off" switches introduced by moist physical parameterizations, the G-differential may still exist. Methods for unconstrained minimization of non-smooth functions may be used to solve the variational problem. However, variational approaches do not require (in principle) the existence of continuous derivatives of the cost function with respect to the state functions IC, although variation of a functional (i.e., the Frechet differential, the G-differential, or the Gateaux variations) is still needed. Gill et al. (1981) also indicated that if a function has just a "few" discontinuities in its first derivative and these discontinuities do not occur in the neighbourhood of the solution, the smooth methods are likely to be more efficient. The success of minimization will then depend critically on whether the model gives a close fit to the data.

4. Analytical study

When the Kuo cumulus convection is turned on at a grid point at time τ, the moisture convergence is partitioned into a rain-producing portion and a humidity-increasing portion; the latter is added to the specific humidity tendency at this particular point and time. We will use two simple examples to examine theoretically the feasibility of using the adjoint model constructed above, including "on-off" switches which depend on model variables for adjoint sensitivity studies and 4DVAR experiments. The first example assumes a constant increment to the tendency of model variables when moist processes are turned on, and the second assumes that the increment to the tendency term depends on model variables. In both cases, the "on-off" switches are turned on and off depending on the model variables.

4.1. *Constant increment*

The time evolution of the specific humidity at one grid point can be simplified into the following equation, as was used in Xu (1996):

$$\begin{cases} \dfrac{\mathrm{d}q}{\mathrm{d}t} = F + \beta H_+(q - q_c) \\ q|_{t=0} = q_0 \end{cases} \tag{4.1}$$

where β is the humidity-increasing portion resulting from cumulus convection, and we assumed that $\beta > 0$, and $H_+(\cdot)$ is the unsymmetrical Heaviside unit-step function defined as

$$H_+(x) = \begin{cases} 1, & \text{if } x \geqslant 0, \\ 0, & \text{if } x < 0. \end{cases}$$

To ensure that the switch is turned on once during the times t_0 and t_m, we assume $q_0 < q_c$ and $F > 0$. The switch time is dependent on the model state q, and q_c is the threshold moisture for the switch. This resembles both the large-scale precipitation and cumulus parameterization. We assume that the 4DVAR is to find the best estimate of q_0 that minimizes the following cost function:

$$J(q_0) = \frac{1}{2} \int_{t_0}^{t_m} (q - q^{\mathrm{obs}})^2 \, \mathrm{d}t, \tag{4.2}$$

where q is the solution of (4.1), and q^{obs} is the observed value of q during the assimilation period $[t_0, t_m]$. We will compare the TLM or adjoint-derived ΔJ with the true variation δJ, where we assume that the TLM and adjoint models are constructed by keeping all the switching points the same as in the NLM, which determines the basis-state solution.

The true solution of (4.1) is

$$q = Ft + q_0 + \left(t - \frac{q_c - q_0}{F}\right) \beta H_+(q - q_c). \tag{4.3}$$

For a perturbation of q_0' to IC q_0, the perturbed solution is

$$q^{\mathrm{ptb}} = Ft + q_0 + q_0' + \left(t - \frac{q_c - q_0 - q_0'}{F}\right) \beta H_+(q^{\mathrm{ptb}} - q_c). \tag{4.4}$$

Here $\tau = (q_c - q_0)/F$ and $\tau' = (q_c - q_0 - q_0')/F$ are switching times for the basic-state solution and the perturbed solution, respectively. Assuming $q_0' > 0$, the difference between the perturbed and basic state solution can be expressed as

$$\begin{aligned} q' \equiv q^{\mathrm{ptb}} - q &= q_0' + \beta(t - \tau')H_+(q^{\mathrm{ptb}} - q_c) \\ &\quad - \beta(t - \tau)H_+(q - q_c) \\ &= \begin{cases} q_0', & \text{if } t < \tau' \\ q_0' + \beta\left(t - \dfrac{q_c - q_0 - q_0'}{F}\right), & \text{if } \tau' \leqslant t < \tau. \\ q_0' + \dfrac{\beta}{F} q_0', & \text{if } t \geqslant \tau \end{cases} \end{aligned} \tag{4.5}$$

The perturbation in the cost function resulting from the perturbation $\delta q_0 = q_0'$ to IC q_0 is

$$\delta J = \int_{t_0}^{t_m} (q - q^{\mathrm{obs}}) q' \, \mathrm{d}t. \tag{4.6}$$

Substituting (4.5) into (4.6) gives

$$\begin{aligned} \delta J &= \int_{t_0}^{t_m} (q - q^{\mathrm{obs}}) q_0' \, \mathrm{d}t + \int_{\tau}^{t_m} (q - q^{\mathrm{obs}}) \frac{\beta}{F} q_0' \, \mathrm{d}t \\ &\quad + \int_{\tau'}^{\tau} (q - q^{\mathrm{obs}})(t - \tau')\beta \, \mathrm{d}t \\ &\equiv \int_{t_0}^{t_m} (q - q^{\mathrm{obs}}) \left(1 + \frac{\beta}{F} H_+(q - q_c)\right) q_0' \, \mathrm{d}t \\ &\quad + Z(q_0'). \end{aligned} \tag{4.7}$$

Since in the integral of Z, $|t - \tau'| \leqslant |\tau - \tau'| = |q_0'/F|$, we have

$$\lim_{q_0' \to 0} \frac{Z(q_0')}{q_0'} = 0. \tag{4.8}$$

We thus obtain the G-differential of J:

$$VJ(q_0; q_0') = \int_{t_0}^{t_m} (q - q^{\mathrm{obs}}) \times \left(1 + \frac{\beta}{F} H_+(q - q_c)\right) q_0' \, \mathrm{d}t. \quad (4.9)$$

Keeping the "on-off" switch the same as in the NLM (i.e., the Heaviside unit-step function will not depend on perturbation), and linearizing the rest of the terms in (4.3) we obtain

$$q_{\mathrm{approx}}^{\mathrm{ptb}} = Ft + q_0 + q_0' + \left(t - \frac{q_c - q_0 - q_0'}{F}\right) \beta H_+(q - q_c), \quad (4.10)$$

$$q_{\mathrm{approx}}' \equiv q_{\mathrm{approx}}^{\mathrm{ptb}} - q = q_0' + \frac{\beta}{F} q_0' H_+(q - q_c). \quad (4.11)$$

Substituting (4.11) into (4.6) we obtain the same result as (4.9). This proves that keeping the "on-off" switches fixed the resulted "first-order" perturbation solution is correct. However, in the practical TLM and adjoint model development, linearization is carried out for the tendency eq. (4.1) instead of its solution (4.3), result becomes different as shown below.

The TLM equation of (4.1) should take the form of

$$\begin{cases} \dfrac{\mathrm{d}q'}{\mathrm{d}t} = 0 \\ q'|_{t=0} = q_0' \end{cases} \quad \text{for } t < \tau, \quad (4.12a)$$

$$\begin{cases} \dfrac{\mathrm{d}q'}{\mathrm{d}t} = 0 \\ q'|_{t=\tau} = q_0' + \dfrac{\beta}{F} q_0' \end{cases} \quad \text{for } t \geqslant \tau. \quad (4.12b)$$

Notice that due to the presence of a discontinuity occurring at the "on-off" switch on time $t = \tau$, the TLM now split into two equations preserving the discontinuity at the switching point. Therefore, (4.12a) and (4.12b) together form the TLM of (4.1). The adjoint model can thus be written as (see Navon et al., 1992)

$$\begin{cases} -\dfrac{\mathrm{d}\hat{q}}{\mathrm{d}t} = q - q^{\mathrm{obs}} \\ \hat{q}|_{t=\tau} = 0 \end{cases} \quad \text{for } t < \tau, \quad (4.13a)$$

$$\begin{cases} -\dfrac{\mathrm{d}\hat{q}}{\mathrm{d}t} = q - q^{\mathrm{obs}} \\ \hat{q}|_{t=t_m} = 0 \end{cases} \quad \text{for } t \geqslant \tau, \quad (4.13b)$$

where $\hat{q}$ represents the adjoint variable of specific humidity, $\partial J / \partial q = q - q^{\mathrm{obs}}$ is forcing term. Integrating (4.13a) and (4.13b) backward in time, we obtain

$$\hat{q}_0 = \int_{t_0}^{\tau} (q - q^{\mathrm{obs}}) \, \mathrm{d}t,$$

$$\hat{q}_\tau = \int_{\tau}^{t_m} (q - q^{\mathrm{obs}}) \, \mathrm{d}t. \quad (4.14)$$

Following the gradient derivation in Navon et al. (1992) using adjoint technique, we obtain that the first-order variation can be expressed as

$$\Delta J = \hat{q}_0 q_0' + \hat{q}_\tau q_\tau' = \hat{q}_0 q_0' + \hat{q}_\tau \left(q_0' + \frac{\beta}{F} q_0'\right). \quad (4.15)$$

Substituting (4.14) into (4.15) we obtain the first-order variation of J, in the presence of the discontinuous "on-off" switch, as

$$\Delta J = \int_{t_0}^{\tau} (q - q^{\mathrm{obs}}) q_0' \, \mathrm{d}t + \int_{\tau}^{t_m} (q - q^{\mathrm{obs}}) \left(q_0' + \frac{\beta}{F} q_0'\right) \mathrm{d}t = \int_{t_0}^{t_m} (q - q^{\mathrm{obs}}) \left(1 + \frac{\beta}{F} H_+(q - q_c)\right) q_0' \, \mathrm{d}t, \quad (4.16)$$

which is exactly the G-differential of J in (4.9).

However, in the practical adjoint development with moist physics, eqs. (12b) and (13b) are not considered. The adjoint model is integrated backward continuously from t_m to t_0 and the gradient of the cost function is obtained as $\hat{q}_0 + \hat{q}_\tau$. Examining (4.15) we see that this is right if $\beta / F \ll 1$. Otherwise, the straight backward adjoint model integration will contain an error of magnitude $\int_\tau^{t_m} (\beta/F)(q - q^{\mathrm{obs}}) \, \mathrm{d}t$, where $t_m - \tau$ is the time duration of the moist process being turned on. Such error could be significant (Xu, 1996).

Analyzing the example given by (4.1), we find that the reason which causes such errors in the gradient calculation is two: (i) NLM solution is discontinuous at the switching point, and (ii) the switch on-off information is totally lost in the TLM due to the overly simplified assumption for a constant increment due to convection.

4.2. *Variable increments*

A more realistic assumption for a conceptual model would be that the contribution to the temperature and moisture time tendency introduced by the moist convection depends on model variables, which may be illustrated as

$$\begin{cases} \dfrac{\mathrm{d}q}{\mathrm{d}t} = F + H_+(q-q_c)\alpha q \\ q|_{t=0} = q_0 \end{cases} \tag{4.17}$$

where α is a constant. We will now use this case to examine the feasibility of using the "on-off" adjoint for 4DVAR. The same cost function as defined in (4.2) will be used for the following discussions. Again we assume $q_c > q_0$, $F > 0$, and that the "on-off" process will occur at least once between time t_0 and t_m.

The solution to (4.17) is

$$q = \begin{cases} Ft + q_0, & \text{if } t < \tau \\ \left(q_c + \dfrac{F}{\alpha}\right) \mathrm{e}^{\alpha(t-(q_c-q_0)/F)} - \dfrac{F}{\alpha}, & \text{if } t \geqslant \tau. \end{cases} \tag{4.18}$$

The perturbed solution can be written as

$$q^{\text{ptb}} = \begin{cases} Ft + q_0 + q_0', & \text{if } t < \tau' \\ \left(q_c + \dfrac{F}{\alpha}\right) \mathrm{e}^{\alpha(t-(q_c-q_0-q_0')/F)} - \dfrac{F}{\alpha}, & \text{if } t \geqslant \tau'. \end{cases} \tag{4.19}$$

Assuming $q_0' > 0$, we can obtain the perturbation solution as

$$q' \equiv q^{\text{ptb}} - q = \begin{cases} q_0', & \text{if } t < \tau' \\ \left(q_c + \dfrac{F}{\alpha}\right) \mathrm{e}^{\alpha(t-(q_c-q_0-q_0')/F)} - \dfrac{F}{\alpha} - (Ft + q_0), & \text{if } \tau' \leqslant t < \tau \\ \left(q_c + \dfrac{F}{\alpha}\right) (\mathrm{e}^{\alpha(t-(q_c-q_0-q_0')/F)}) - \mathrm{e}^{\alpha(t-(q_c-q_0)/F)}), & \text{if } t \geqslant \tau \end{cases}$$

$$= \begin{cases} q_0', & \text{if } t < \tau' \\ \left(q_c + \dfrac{F}{\alpha}\right) \mathrm{e}^{\alpha(t-(q_c-q_0)/F)} - \dfrac{F}{\alpha} - (Ft + q_0) \\ \quad + \dfrac{\alpha}{F}\left(q_c + \dfrac{F}{\alpha}\right) q_0' \, \mathrm{e}^{\alpha(t-(q_c-q_0-q_0')/F)} + \mathrm{O}(q_0'^2), & \text{if } \tau' \leqslant t < \tau \\ \left(q_c + \dfrac{F}{\alpha}\right) \dfrac{\alpha}{F} \mathrm{e}^{\alpha(t-(q_c-q_0)/F)} q_0' + \mathrm{O}(q_0'^2), & \text{if } t \geqslant \tau \end{cases} \tag{4.20}$$

Substituting (4.20) into (4.6) gives

$$\delta J = \int_{t_0}^{t_m} (q - q^{\text{obs}}) \left(1 + H_+(q-q_c)\left(q_c + \frac{F}{\alpha}\right) \frac{\alpha}{F} \mathrm{e}^{\alpha(t-(q_c-q_0)/F)}\right) q_0' \, \mathrm{d}t$$

$$+ \int_{\tau'}^{\tau} (q - q^{\text{obs}}) \left(\left(\left(q_c + \frac{F}{\alpha}\right) \mathrm{e}^{\alpha(t-(q_c-q_0)/F)} - \frac{F}{\alpha}\right) - (Ft + q_0)\right. \tag{4.21}$$

$$+\left(\left(q_c+\frac{F}{\alpha}\right)\frac{\alpha}{F}\,e^{\alpha(t-(q_c-q_0)/F)}-1\right)\right)q_0'\,dt+O(q_0'^2)$$

$$\equiv\int_{t_0}^{t_m}(q-q^{obs})\left(1+H_+(q-q_c)\left(q_c+\frac{F}{\alpha}\right)\frac{\alpha}{F}\,e^{\alpha(t-(q_c-q_0)/F)}\right)q_0'\,dt+Z(q_0')+O(q_0'^2).$$

Using the following Taylor expansion near $t=\tau(\tau=(q_c-q_0)/F)$:

$$e^{\alpha(t-(q_c-q_0)/F)}=1+\alpha\left(t-\frac{q_c-q_0}{F}\right)+O\left(\left(t-\frac{q_c-q_0}{F}\right)^2\right), \tag{4.22}$$

we obtain

$$\lim_{q_0'\to 0}\frac{Z(q_0')}{q_0'}=0. \tag{4.23}$$

Therefore, the G-differential of J at q_0 exists and is equal to

$$VJ(q_0;q_0')=\int_{t_0}^{t_m}(q-q^{obs})\left(1+H_+(q-q_c)\left(q_c+\frac{F}{\alpha}\right)\frac{\alpha}{F}\,e^{\alpha(t-(q_c-q_0)/F)}\right)q_0'\,dt. \tag{4.24}$$

If the switch on-off points are kept same as in the NLM solution, linearizing (4.18) we obtain

$$q_{approx}^{ptb}=\begin{cases}Ft+q_0+q_0', & \text{if } t<\tau\\ \left(q_c+\dfrac{F}{\alpha}\right)e^{\alpha(t-(q_c-q_0-q_0')/F)}-\dfrac{F}{\alpha}, & \text{if } t\geqslant\tau\end{cases} \tag{4.25}$$

and

$$q_{approx}'\equiv q_{approx}^{ptb}-q=\begin{cases}q_0', & \text{if } t<\tau\\ \left(q_c+\dfrac{F}{\alpha}\right)\dfrac{\alpha}{F}\,e^{\alpha(t-(q_c-q_0)/F)}q_0'+O(q_0'^2), & \text{if } t\geqslant\tau\end{cases} \tag{4.26}$$

Substituting (4.26) into (4.6) gives the same results as (4.24). This shows that the approximated perturbation solution (assuming that the switch on and off timing and location are not influenced by the perturbation) also provide correct G-differential of the cost function in the case of variable increment due to convection.

Now we examine what happens if the linearization is carried out on the tendency equation (4.17), not the NLM solution itself (4.18). Linearizing the model equation (4.17), assuming that the "on-off" switches are determined by q, we obtain the TLM

$$\begin{cases}\dfrac{dq'}{dt}=H_+(q-q_c)\alpha q'\\ q_0'|_{t=t_0}=q_0'.\end{cases} \tag{4.27}$$

The perturbation solution from TLM is

$$q'=q_0'+H_+(q-q_c)C\,e^{\alpha(t-(q_c-q_0)/F)}, \tag{4.28}$$

where coefficient C is the value of the perturbation solution at the switching time $t=\tau\equiv(q_c-q_0)/F$. Since the perturbation solution of (4.17) is discontinuous at $t=\tau$, there are two options in the choice of the C value (4.20):

$$C\equiv\lim_{t^+\to\tau}q'(t)=\left(q_c+\frac{F}{\alpha}\right)\frac{\alpha}{F}\,q_0', \tag{4.29}$$

or

$$C=\lim_{t^-\to\tau}q'(t)=q_0'. \tag{4.30}$$

Substituting (4.29) or (4.30) into (4.28) we obtain

$$q'=q_0'+H_+(q-q_c)\left(q_c+\frac{F}{\alpha}\right)\frac{\alpha}{F}\,q_0'\,e^{\alpha(t-(q_c-q_0)/F)}, \tag{4.31}$$

and

$$q'=q_0'+H_+(q-q_c)q_0'\,e^{\alpha(t-(q_c-q_0)/F)}, \tag{4.32}$$

respectively. Substituting (4.31) into (4.6) we obtain the same expression as the right-hand side of (4.24). However, if the perturbation solution of (4.32) is used we obtain

$$VJ^{\mathrm{tlm}}(q_0; q_0') = \int_{t_0}^{t_m} (q - q^{\mathrm{obs}}) \times (1 + H_+(q - q_c)\, e^{\alpha(t - (q_c - q_0)/F)}) \times q_0'\, dt, \tag{4.33}$$

which is different from the true solution (4.24).

We see that the TLM solution will provide the G-differential of J if the perturbation solution q' at the switch-on point takes the value of $\lim_{t^+\to\tau} q'(t)$ as in (4.27) and (4.29). In the TLM and adjoint model, the perturbation solution q' at the switch-on point takes the value of $\lim_{t^-\to\tau} q'(t)$ as in (4.30). No additional information related to the perturbation values at switching points are used in the TLM and adjoint model integrations. Therefore, an error seems to occur in the gradient calculation using the TLM and adjoint model when

$$\lim_{t^+\to\tau} q'(t) \neq \lim_{t^-\to\tau} q'(t). \tag{4.34}$$

The magnitude of the relative error depends on the value of $q_c\alpha/F$. This is of course not desirable in the adjoint applications.

The reason that such error exists in the gradient calculation using conventional adjoint model is because of the existence of discontinuity at switching point in the perturbation solution, which is neglected in the TLM and the gradient calculation using adjoint model (see second term in eq. (4.15)). If the perturbation solution is continuous at the switching point, such error in the gradient calculation will not exist. This can be further illustrated as follows.

Modifying (4.17) into

$$\begin{cases} \dfrac{dq}{dt} = F + H_+(q - q_c)\alpha(q - q_c) \\ q|_{t=0} = q_0. \end{cases} \tag{4.35}$$

The analytical solution of (4.35) and the true G-differential of the cost function (4.2), corresponding to (4.18) and (4.24), are

$$q = \begin{cases} Ft + q_0, & \text{if } t < \tau \\ q_c + \dfrac{F}{\alpha} e^{\alpha(t - (q_c - q_0)/F)} - \dfrac{F}{\alpha}, & \text{if } t \geqslant \tau, \end{cases} \tag{4.36}$$

and

$$VJ(q_0; q_0') = \int_{t_0}^{t_m} (q - q^{\mathrm{obs}}) \times (1 + H_+(q - q_c)\, e^{\alpha(t - (q_c - q_0)/F)}) q_0'\, dt. \tag{4.37}$$

The TLM of (4.35) is the same as (4.37), and its solution (4.32) provides the correct G-differential as in (4.37). This proves that if the perturbation solution is continuous at the switching point, i.e., a gradual change (instead of an abrupt change) are brought into the model, the conventional TLM and adjoint model will still provide the correct first-order approximation to the NLM perturbation solution and the cost function, respectively. A sufficient condition for the correctness of approximated TLM and adjoint model solution is

$$\lim_{t^+\to\tau_i} q'(t) = \lim_{t^-\to\tau_i} q'(t), \tag{4.38}$$

which is equivalent that the right-hand-side of the forward model equation is continuous at the switching time $t = \tau$. If (4.38) does not hold, we may still expect a good approximation of the TLM and adjoint model solution to the first-order information if

$$\frac{q_c\alpha}{F} \ll 1, \tag{4.39}$$

which is obtained by comparing (4.33) with (4.24). Eq. (4.39) means that if the heating and moistening due to convection at the switch-on point is much smaller than all the other terms in the tendency equation, then the possible errors in the TLM solution and the adjoint calculated gradient resulted from keeping the "on-off" switches fixed will be small.

For a discretized numerical model, the discontinuity of the tendency equation becomes less obvious. The jump occurs over a Δt (time step) window, not at a single time level as the case in continuous model. For instance, we can always modify (4.17) into (4.35) over time window $\tau_i - \Delta t$ and τ_i, where τ_i is the switch-on time of the moist processes, which is not turned on at the previous time step $\tau - \Delta t$. In this case, we can assume that there exists threshold time t_{c_i}, where $\tau_i - \Delta t < t_{c_i} < \tau_i$, and $q(t_{c_i}) = q_c$. The q in the second term of (4.17) on the right-hand side may be modified into

$$q = \frac{q_{\tau_i} - q_c}{\tau_i - t_{c_i}} (t - \tau_i) + q_{\tau_i}, \tag{4.40}$$

near the time $t = \tau_i$. The right-hand-side of the tendency equation (4.17) thus becomes continuous at the switching point t_{c_i}. Notice that values of the NLM, TLM, and adjoint model solutions are not altered by (4.40) at every time step. From the previous derivation we know that the TLM and adjoint model developed assuming that all "on-off" switches are determined by the unperturbed NLM solution may still provide a correct first-order information.

The MM5 contains much more complicated "on-off" related functions than examples in (4.1), (4.17), and (4.35). It is impossible to derive their analytical expressions for the solutions. The MM5 TLM and adjoint model are developed at the coding level and the linearization is carried out around the instantaneous NLM solution. When the moist processes are turned on, the linearization is carried out for the moist processes. Constructing the adjoint model at the coding level may prove to be more advantageous than deriving the adjoint equations in analytic form followed by the discrete approximation. It avoids not only the inconsistency arised from different discretization in the forward and adjoint model, but also the above-mentioned "on-off" linearization problem.

In the following, we will examine how these conclusions are valid for a realistic model applied to a mesoscale storm case in which convection occurs.

5. Numerical results

5.1. *Nonlinear model solution with "on-off" switches*

In order to better understand what affect the "on-off" processes may bring to the model variables, we will first examine the NLM solution.

We choose the case of a strong convective storm from 1800 UTC 10 April to 0600 UTC 11 April 1979. This case was selected because the forecast was strongly affected by moist processes. The experiment was initialized with an analysis using all the available surface and upper-air data collected at 1800 UTC 10 April. In order to identify the regions where "on-off" switches occur, we plotted in Fig. 1 the observed (a-b) and model-predicted (c-d) 3-h rainfall ending at 2100 UTC and 2400 UTC 10 April 1979, respectively. The model-predicted precipitation took place near the border of Oklahoma and Texas during the first 3-h, to the southeast of the observed 3-h rainfall (Fig. 1a). This convective rain corresponded to the Wichita Falls tornado outbreak (Kuo et al., 1993). In the next three hours, the observed precipitation expanded into a line pattern, and by 2400 UTC 10 April, it covered western Oklahoma (Fig. 1b). The model precipitation from 2100 UTC to 2400 UTC April 10 (Fig. 1d) exhibited some skill in both the pattern and the magnitude of the 3-h rainfall as compared with observations (Fig. 1b), except that there existed a southeast bias errors of the rain over Oklahoma. The model overestimated precipitation over Colorado.

To examine the detailed behavior of the convective precipitation during the MM5 model integration, time series of convective rain produced at every time step are shown in Fig. 2. These time series are produced for the three model grid points that are shown in Fig. 1c. Point A is in the center of the first-hour rainfall, point B is located northeast of point A where there was no rain initially, and point C is located even further north. Convective rain was repeatedly turned on and off during the first three and a half hours ($\Delta t = 2$ min) at point A. We noticed that gradual delays of convective precipitation at points B and C, which corresponded to the northeastward movement of the convective system. We also noticed several occurrrences of $2\Delta t$ oscillations with convective rain being turned on and off at every other time step. This feature was also pointed out by Vukicevic and Errico (1993). In order to understand what was really causing such undesirable "discontinuities" in Fig. 3, we present the three criteria listed in section 2.3 for convection to occur at point A during the 3-h forecast. We found that these oscillations are caused by the violation to stability criterion 3. We also checked results at points B and C, and similar behaviors were found.

We can therefore remove such $2\Delta t$ oscillation

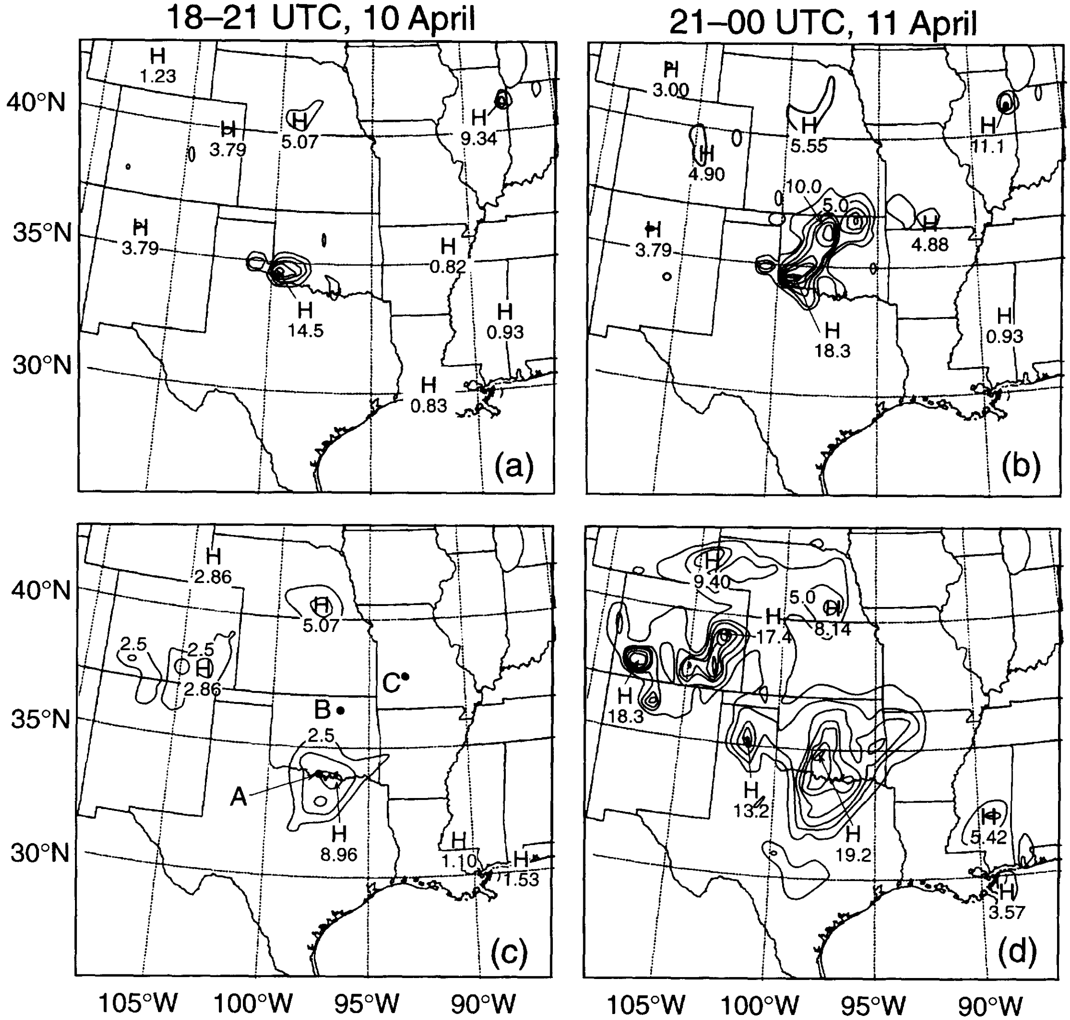


Fig. 1. The observed (a-b) and model-predicted (c-d) 3-h rainfall (mm) ending at 2100 UTC and 2400 UTC 10 April 1979, respectively. Contour interval is 2.5 mm.

simply by eliminating the third criteria in the Kuo cumulus parameterization scheme (we will call it as smoother Kuo scheme). Fig. 4 shows the variation of the convective rain at every time step during the 12-h forecast using the smoothed Kuo scheme. We observe that the amount of convective rain is slightly higher than the original run while the $2\Delta t$ oscillations disappear. This modification will be used for comparison study of the "on-off" switches under different circumstances.

To examine the contribution of these moisture processes to the evolution of the model's solution, the time variation of T and q at point A at $\sigma = 0.25$ and $\sigma = 0.85$ are given in Fig. 5, where the original Kuo scheme is used. On each panel, we can detect the signals of $2\Delta t$ oscillations, even though the amplitude of the jigs is very small compared to the total changes of temperature and moisture during the 12-h forecast ($\leqslant 10\%$). This also shows that condition of (4.39) is satisfied here for the TLM and adjoint model to produce a good approximation of the first-order information to the NLM solution and the gradient of the cost function, respectively.

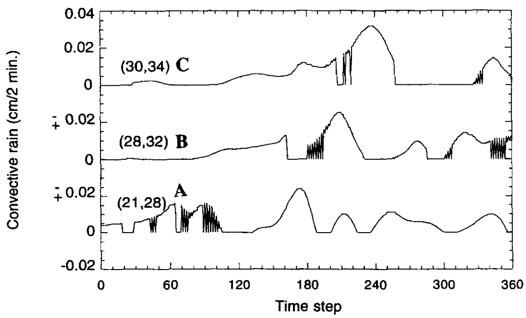


Fig. 2. The time series of convective rain produced every time step (cm/2 min) at 3 grid points: A(21,28), B(28,32), and C(30,34) shown in Fig. 1c.

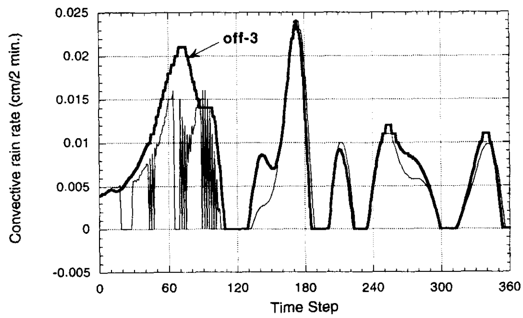


Fig. 4. The variation of the convective rain at every time step (cm/2 min) during the 12-h forecast with (thick solid line) and without (thin solid line) criterion 3 turned off.

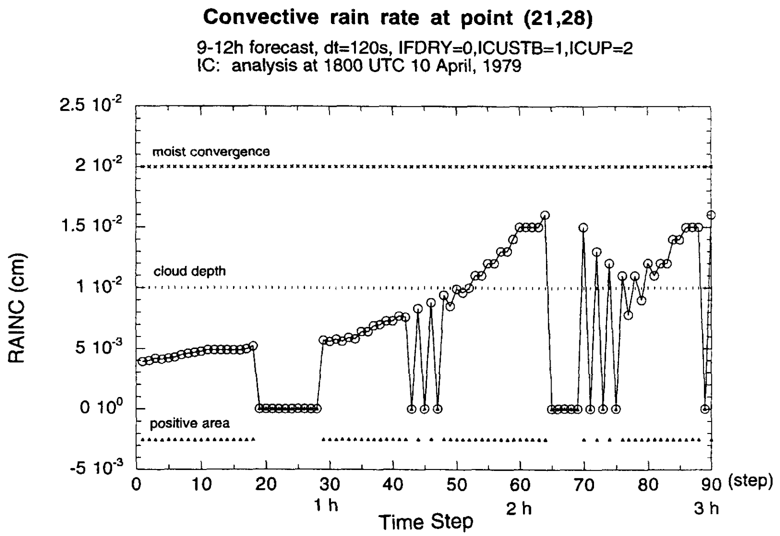


Fig. 3. Tracing of the 3 criteria for convection to occur at point A during the 3-h forecast.

5.2. Tangent linear approximation

Knowing IC and LBC, (3.1) determines a unique solution in time. We will restrict the discussion to IC here, leaving LBC and model parameters fixed.

The TLM (3.2) can be specifically written as

$$\begin{cases} \boldsymbol{x}'_1 = \boldsymbol{P}^0_1(\boldsymbol{x}_0)\boldsymbol{x}'_0, & \\ \boldsymbol{x}^1_n = \boldsymbol{P}^{n-2}_n(\boldsymbol{x}_{n-1}, \boldsymbol{x}_{n-2})\begin{pmatrix} \boldsymbol{x}'_{n-2} \\ \boldsymbol{x}'_{n-1} \end{pmatrix}, & n = 2, 3, \ldots \end{cases} \tag{5.1}$$

where a forward two-time-level integration at the initial step and three-time-level leapfrog time scheme afterward is assumed, and $\boldsymbol{x}'_n$ is the first-order approximation to

$$x'_n = \boldsymbol{x}_n(\boldsymbol{x}_0 + \Delta\boldsymbol{x}_0) - \boldsymbol{x}_n(\boldsymbol{x}_0). \tag{5.2}$$

Operator $\boldsymbol{P}$ is obtained through a linearization procedure linearized about the NLM solution $\boldsymbol{x}_{n-1}$ and $\boldsymbol{x}_{n-2}$ at every time step:

$$\boldsymbol{P}^{n-2}_n \equiv \boldsymbol{P}(\boldsymbol{x}_{n-1}, \boldsymbol{x}_{n-2}) = \left(\frac{\partial \boldsymbol{Q}(\boldsymbol{x}_{n-1}, \boldsymbol{x}_{n-2})}{\partial \boldsymbol{x}_{n-1}}, \frac{\partial \boldsymbol{Q}(\boldsymbol{x}_{n-1}, \boldsymbol{x}_{n-2})}{\partial \boldsymbol{x}_{n-2}}\right). \tag{5.3}$$

The TLM operator in (3.2) is now decomposed

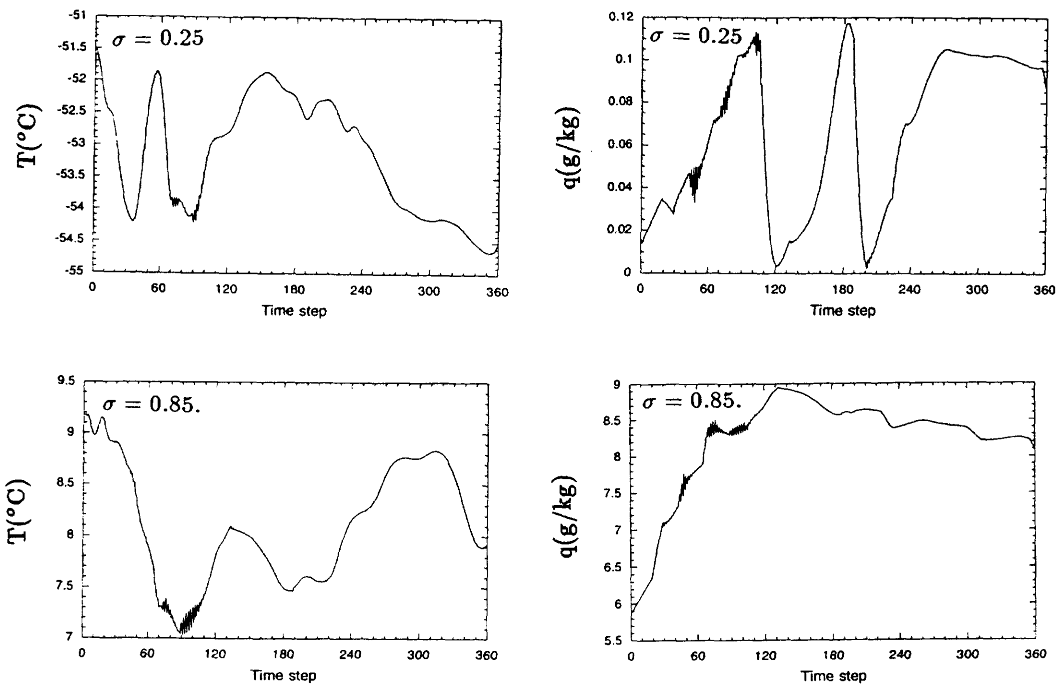


Fig. 5. The time variation of the temperature (°C) (left panel) and the specific humidity (right panel) (g/kg) at point A with $\sigma = 0.25$ and $\sigma = 0.85$.

into several operator matrices:

$$\boldsymbol{P}_n^0 = \boldsymbol{P}_n^{n-2} \boldsymbol{P}_{n-1}^{n-3} \cdots \boldsymbol{P}_2^0 \boldsymbol{P}_1^0. \tag{5.4}$$

If the NLM includes "on-off" switches, which might cause either discontinuities or non-differentiabilities, we should carefully examine the feasibility of applying the adjoints of such processes for gradient calculation. However, the accuracy of the TLM may mean either of two things:

(1) How well does the TLM solution approximate the NLM solution?

(2) How accurately does the TLM provide the first-order approximation to the NLM solution?

The answer to the first question depends on the degree of nonlinearity of the NLM solution. The TLM results may still be reasonably accurate even for significantly large perturbations if the NLM solution has a weak nonlinearity, such as during the initial short forecast period. However, when the NLM solution becomes highly nonlinear, using TLM to approximate such a NLM solution will, of course, not be accurate, especially for finite-amplitude perturbations. The higher-order terms become important. Including the "on-off" related physical parameterization schemes in the model may increase the model's nonlinearity. However, in 4DVAR application, the adjoint model is used no more than to provide the first-order derivative information for a quantitative aspect J of a numerical model:

$$\nabla J = \lim_{\Delta \boldsymbol{x}_0 \to 0} \frac{J(\boldsymbol{x}_0 + \Delta \boldsymbol{x}_0) - J(\boldsymbol{x}_0)}{\Delta \boldsymbol{x}_0}. \tag{5.5}$$

Since the derivative of a cost function J at $\boldsymbol{x}_0$ is derived for an infinitesimal perturbation $\Delta \boldsymbol{x}_0$ ($\Delta \boldsymbol{x}_0 \ll \varepsilon$, where ε is an arbitrary small number), the accuracy of the TLM and adjoint calculated derivative should be checked against very small perturbations in IC, not for significantly large perturbations; the same is true for the TLM test. "Very small" perturbations means small enough to be able to ignore the higher-order terms and "large" enough to avoid machine round-off errors. The TLM and adjoint model should not include any kinds of approximations for these tests, such as an incomplete physics in the TLM and adjoint model compared to the NLM, or a less frequent basic state updating procedures than at every time step. These are separate issues due to computational practice or other research interests.

One way of verifying the correctness of the TLM for a model without "on-off" switches is as follows: Define

$$\Phi(\alpha)=\frac{Q(\boldsymbol{x}_0+\alpha\boldsymbol{h})-Q(\boldsymbol{x}_0)}{\alpha\boldsymbol{P}_n\boldsymbol{h}}. \tag{5.6}$$

From (5.6) for values of α that are small but not too close to the machine zero, one should expect to obtain a value for $\Phi(\alpha)$ that is close to 1. (5.6) should be verified at any grid point in the model domain. However, when the denominator is close to zero, the function of $\Phi(\alpha)$ is not appropriate to use.

The values of $\Phi(\alpha)$ for the specific humidity and temperature at $\sigma=0.95$ and point A are shown in Fig. 6 for 1-h, 2-h, 3-h, and 4-h forecasts, in which the 64-bit real numbers were used (Cray). It is clearly seen that, for α between 10^{-3} and 10^{-12}, a unit value of $\Phi(\alpha)$ is obtained only for the 1-h to 3-h forecasts.

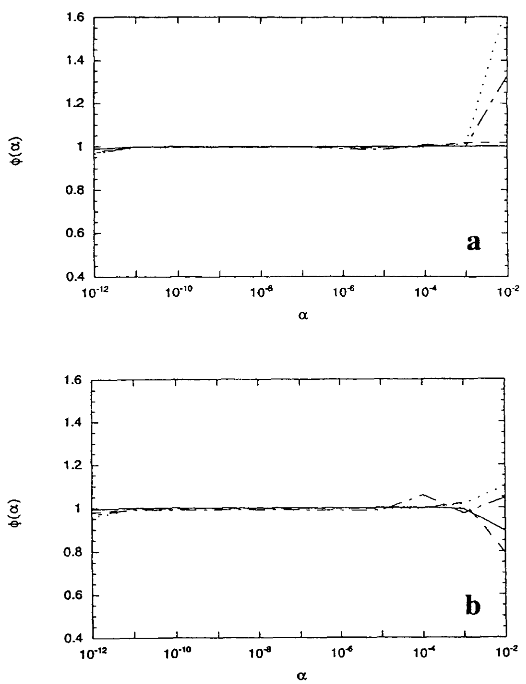


Fig. 6. The value of $\Phi(\alpha)$ for the temperature (a) and the specific humidity (b) at $\sigma=0.95$ for 1-h (solid line), 2-h (dashed line), 3-h (dotted line), and 4-h (dash-dotted line) forecasts.

From Fig. 2, we know that repeated on-off have occurred during the first 3-h forecast. But a unit value of $\Phi(\alpha)$ defined in (5.6) is still obtained. The non-unit value for the 4-h forecast may thus be attributed to other reasons, such as the gradual increase of the nonlinear effect which causes the perturbation solution at 4 h not to be small enough for the accuracy check of the TLM. This will be illustrated by the following discussions.

For a perturbed IC;

$$\boldsymbol{x}_0+\Delta\boldsymbol{x}_0, \tag{5.7}$$

we can obtain the perturbation solution: $\boldsymbol{x}'_n$, $n=1,2,3,\ldots$, through 2 ways: one is to integrate the NLM (3.1) twice from ICs of $\boldsymbol{x}_0$ and $\boldsymbol{x}_0+\Delta\boldsymbol{x}_0$, respectively; the other is to integrate the TLM (3.2) from $\boldsymbol{x}'_0$ at time t_0. We will denote the first one as $\boldsymbol{x}_n^{\prime\mathrm{ptb}}$ and the second as $\boldsymbol{x}_n^{\prime\mathrm{tlm}}$, which is the first-order approximation to $\boldsymbol{x}_n^{\prime\mathrm{ptb}}$.

In order to examine how well the TLM solution approximates the true perturbation, we show in Fig. 7 the true perturbation ($P^{\prime\mathrm{ptb}}$, solid line) and the TLM approximation ($P^{\prime\mathrm{tlm}}$, dotted line) of the convective rain rate at every time step using (a) the original Kuo scheme and (b) the smoothed Kuo scheme. We observe that the degree of discrepancy between the true perturbation and the TLM approximation is quite similar no matter whether or not a smoothed Kuo scheme is used. In order to see the match between $P^{\prime\mathrm{ptb}}$ and $P^{\prime\mathrm{tlm}}$ more clearly during periods when $2\Delta t$ oscillations occurred, we plot in Fig. 7c the initial 6-h predictions. $P^{\prime\mathrm{tlm}}$ approximates $P^{\prime\mathrm{ptb}}$ quite well when $2\Delta t$ oscillations occurred. This shows that the large discrepancy between the NLM and TLM solutions is not so much related to the presence of seemingly discontinuous "on-off" switches.

To better understand the impact of nonlinearity and discontinuity of the cumulus parameterization on the accuracy of TLM solution, we examine the time variation of the true perturbation for the specific humidity at point A from the NLM integration $\boldsymbol{q}_n^{\prime\mathrm{ptb}}$, and the corresponding TLM solution $\boldsymbol{q}_n^{\prime\mathrm{tlm}}$, using (a) an adiabatic version of MM5 without moist physics, (b) a diabatic version of MM5 with the original Kuo scheme, and (c) a diabatic version of MM5 with the smoothed Kuo scheme (Fig. 8), where $\Delta\boldsymbol{x}_0=10^{-9}\boldsymbol{x}_0$ is assumed. Switches not depending on initial perturbation

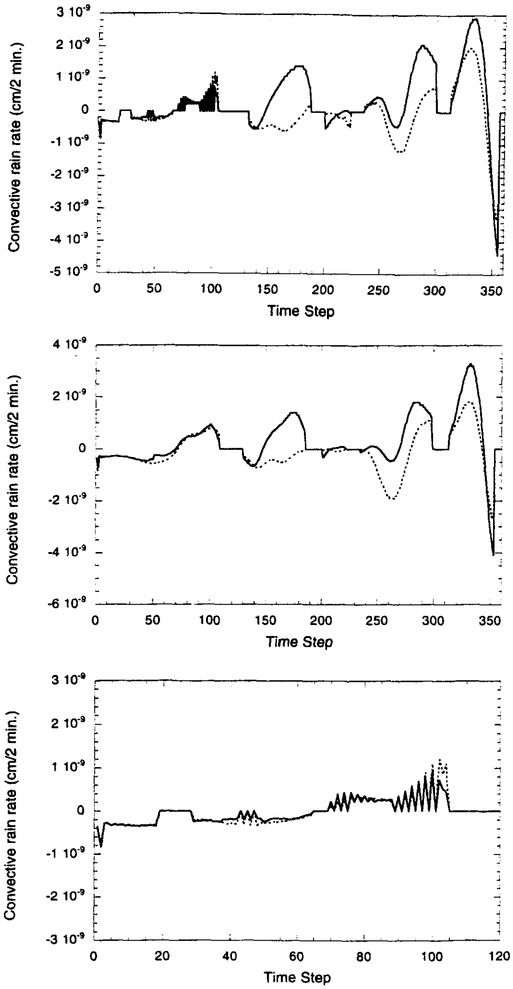


Fig. 7. The time variation of the true perturbation for convective rain rate from NLM integration ($P_n'^{\mathrm{ptb}}$, solid line) and the TLM solution ($P_n'^{\mathrm{tlm}}$, dotted line) using (a) the original Kuo scheme and (b) the smoothed Kuo scheme during the 12-h forecast. (c) is the same as (a) except the x-axis scale is increased to plot only the first 4-h result in order to see more clearly the details when $2\Delta t$ oscillations occurred.

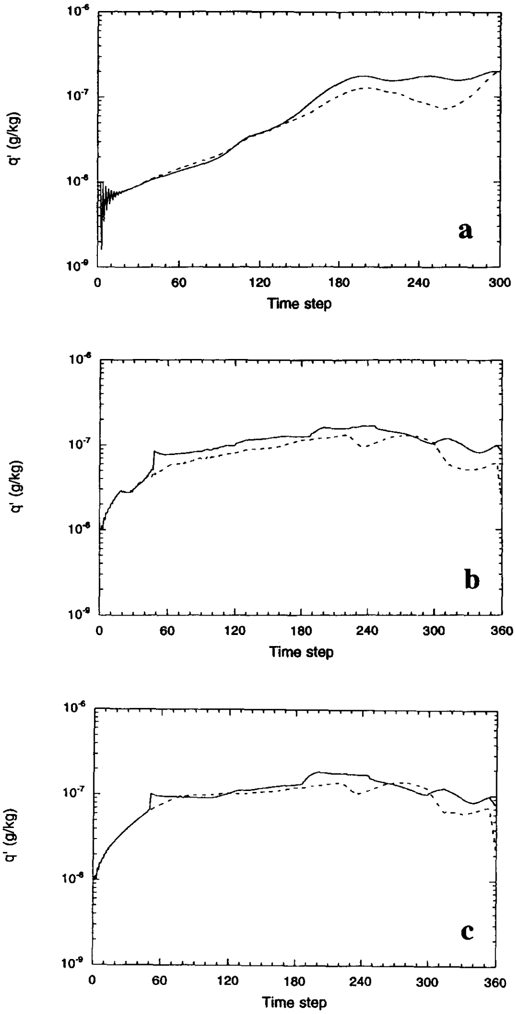


Fig. 8. The time variation of the true perturbation from NLM integration ($q_n'^{\mathrm{ptb}}$, solid line) and the TLM solution ($q_n'^{\mathrm{tlm}}$, dashed line) at $\sigma = 0.95$ using (a) an adiabatic version of MM5, (b) a diabatic version of MM5 with moist physics, and (c) a diabatic version of MM5 with the smoothed Kuo scheme.

$\boldsymbol{x}_0'$, which could exist in the adiabatic model, are not considered as "on-off" switches in this paper. The "on-off" switches in the perturbed solution $\boldsymbol{x}_n(\boldsymbol{x}_n + \Delta \boldsymbol{x})$ occur at the same time and same location as in the basic state solution $\boldsymbol{x}_n(\boldsymbol{x}_0)$ (figure omitted). A significant increase for the value of $\boldsymbol{x}_n'^{\mathrm{ptb}}$ (about an order of magnitude) occurs in the initial 2 h in both diabatic runs with or without the $2\Delta t$ oscillation. In the adiabatic integration, such an increase is delayed by 3 h. A general feature is that the TLM solution (dashed line) differs from the nonlinear solution (solid line) at the time when the perturbation become larger. The magnitudes of such discrepancy between TLM solution and the true perturbation (distances between the solid and dotted line) are similar after the magnitude of the initial perturbation having had a one order of magnitude increase with three versions of the forward model. Since "on-off" switches occur during the initial two hours (see Fig. 2) and the run with the smoothed Kuo scheme in which $2\Delta t$ oscillation eliminated (Fig. 8c) gives similar results to the run with a complete Kuo scheme (Fig. 8b), it seems that the "on-off" switches are not a major contributor to the difference between the TLM approximation and the total perturbation, comparing to the increase of model's nonlinearity due to the introduction of moist physics.

From (5.1) and (5.2), we observe that $\boldsymbol{x}_n'^{\text{tlm}}$ is the first-order approximation to $\boldsymbol{x}_n^{\text{ptb}}$ obtained through (5.1) and satisfies:

$$\boldsymbol{x}_{n+1}^{\text{ptb}} = \boldsymbol{P}_{n+1}^{n-1} \begin{pmatrix} \boldsymbol{x}_{n-1}' \\ \boldsymbol{x}_n' \end{pmatrix} + \varepsilon(\boldsymbol{x}_{n-1}', \boldsymbol{x}_n') \tag{5.8}$$

where

$$\lim_{\boldsymbol{x}_{n-1}', \boldsymbol{x}_n' \to 0} \frac{\varepsilon(\alpha \boldsymbol{x}_{n-1}', \alpha \boldsymbol{x}_n')}{\alpha} = 0. \tag{5.9}$$

Therefore, $\boldsymbol{x}_n'^{\text{tlm}}$ can approximate $\boldsymbol{x}_n^{\text{ptb}}$ to a high degree only if both $\boldsymbol{x}_{n-2}'$ and $\boldsymbol{x}_{n-1}'$ are sufficiently small, which is not ensured for a model solution as time continues. For a highly nonlinear model, $\boldsymbol{x}_n'^{\text{ptb}}$ can become rather large starting from a small initial perturbation $\boldsymbol{x}_0'$. Therefore, we conclude that the fact that the moist processes contain "on-off" switches does not necessarily mean a loss of accuracy of the TLM and adjoint model. However, since moist processes are highly nonlinear, it may substantially increase the nonlinearity to the model solution, as seen from Fig. 8a and 8b. This is also illustrated in Fig. 9, which shows the time evolution of the following quantity

$$|\boldsymbol{q}_n'^{\text{ptb}} - \boldsymbol{q}_n'^{\text{tlm}}| \tag{5.10}$$

at $\sigma = 0.95$ for the three models used in Fig. 8. We notice that the complete diabatic model has a maximum increase of the nonlinear part in the perturbation solution $\boldsymbol{x}_n'$ within 60 time steps (two hours) of model integration. It shows an increase of about 4 orders of magnitude and then remains as the highest during the subsequent 4 h, while the adiabatic model has the lowest level of nonlinearity during the first 6 h. All three models arrive at the same level of nonlinearity after 6 h of integration.

So far, we only examine the moist linearization problem at one particular point in the domain (point A in Fig. 1c). Fig. 10 shows the horizontal distribution of the specific humidity at $\sigma = 0.75$ for q^{ptb} and q^{tlm} for the 2-h forecasts. We find that the largest perturbation area is concentrated over southern Texas and the southeast coast where moisture advection is important. The second maximum is located at the border of Texas and

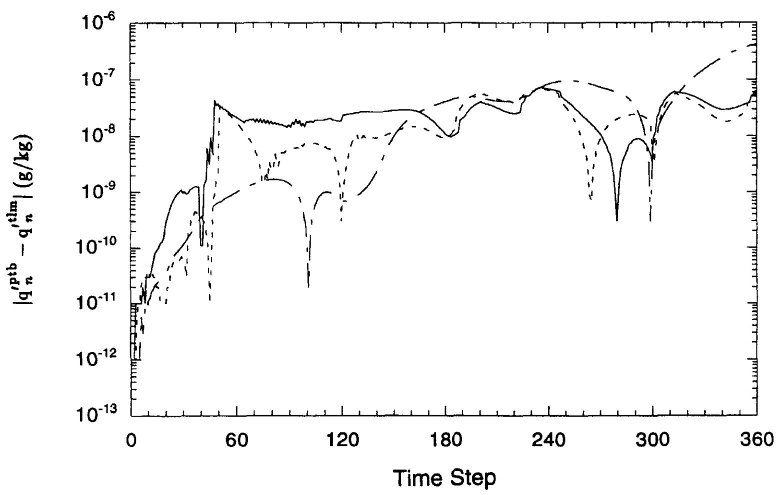


Fig. 9. The time evolution of $|q_n'^{\text{ptb}} - q_n'^{\text{ptb}}|$ (g/kg) at $\sigma = 0.95$ for the diabatic model including the large-scale precipitation scheme and the Kuo scheme (solid line), a diabatic model with the smoothed Kuo scheme (dashed line), and an adiabatic model (dash-dotted line).

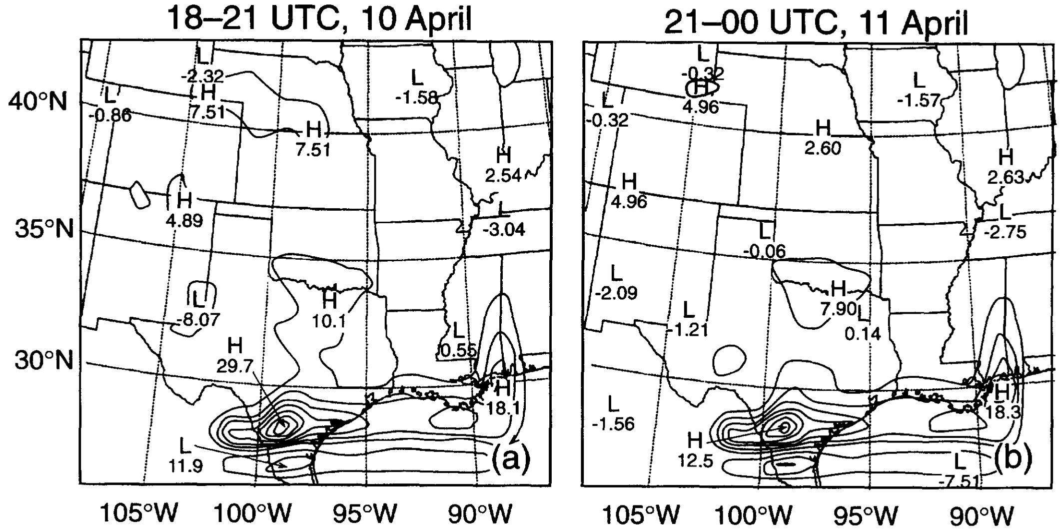


Fig. 10. The horizontal distribution of the specific humidity at $\sigma = 0.95$ for (a) q^{ptb} and (b) q^{tlm} (b and d) at 2-h forecasting times. Contour interval is 4×10^{-3} g/kg. Values in the figure are multiplied by 10^8.

Oklahoma, where precipitation occurs. The TLM approximates the NLM solution quite well, except at the Texas-and-Oklahoma border, where the perturbation from the TLM is smaller than the true perturbation. From the previous discussions, we believe that this discrepancy between q^{tlm} and q^{ptb} is due to the higher-order nonlinear terms, which is not accounted for by the TLM approximation q^{tlm}.

Therefore, we concluded that the "on-off" switches presented in the moist physical parameterization schemes will not prevent developing a meaningful TLM and adjoint model for use in different adjoint applications as far as only the first-order information is required. By definition, the TLM and adjoint model will not account for errors in the higher-than-first-order terms. However, the presence of "on-off" processes along with the moist processes may increase the forward model's nonlinearity to a large extent. Such increase of nonlinearity in the model solution due to the moist physical processes will make the forecast from the TLM deviate from the NLM solution after a certain time. It may also render the minimization problem more difficult to solve, especially over a long assimilation window, since the behaviour of the minimization procedure depends implicitly on the Hessian matrix, which is the second-order information of J. For adjoint sensitivity study, the poor approximation of the TLM solution to the NLM perturbation solution becomes more serious than that in 4DVAR since the adjoint calculated gradient contains less information and is less accurate for significant perturbations.

Another issue to be addressed is related to how to check the correctness of the TLM when moist physics is present. As we pointed out, due to the rapid increase of perturbation with time, (5.6) may not be valid for long time integrations. Since the TLM code at each time step is the same except for the first time step, a short-time integration check is sufficient. If convection does occur during the first few hours, we can break the TLM check into several short time periods starting from different time steps.

5.3. 4DVAR experiments

Three 4DVAR experiments using the diabatic MM5 and its adjoint including "on-off" processes are conducted to examine the performance of the minimization procedure. Analyses of wind, temperature, surface moisture, precipitable water (PW), and 3-h rainfall using data collected during SESAME 1979 were assimilated into the model. These observations were available at an interval of 3 h, and we thus performed 4DVAR data assimilation with a 3-h window.

In the first experiment (Exp. 1), only 3-h rainfall between 1800 UTC and 2100 UTC 10 April is assimilated into the model, which includes the large-scale precipitation and the Kuo cumulus convective parameterization scheme and their adjoints. The cost function is defined as

$$J(\boldsymbol{x}_0) = (P_{3\mathrm{h}} - P_{3\mathrm{h}}^{\mathrm{obs}})^2, \tag{5.11}$$

where $P_{3\mathrm{h}}$ and $P_{3\mathrm{h}}^{\mathrm{obs}}$ represent model-predicted and observed 3-h rainfall from 1800 UTC to 2100 UTC 10 April, respectively. The initial guess for the minimization ($\boldsymbol{x}_0^{(0)}$) is taken as the analysis. This problem is obviously an underestimated problem, since the number of observations (m) is much less than the degrees of freedom of the control variable ($n = 60\ m$, see Subsection 2.3). In such case, minimization is easier to converge and adjustments in the ICs could be rather arbitrary. In fact, with infinite number of iterations, the model will exactly fit to the observations. Nevertheless, we would like to see what changes such rainfall assimilation will bring to the guess IC and how much improvement the rainfall assimilation may have to the short-range rainfall prediction comparing to the forecast starting from analysis by assimilating 3-h rainfall data.

The minimization converges very fast. The cost function decreases one-and-a-half orders of magnitude in 10 iterations when a minimum is approached. The gradient of the cost function with respect to IC decreases two-and-a-half orders of magnitude during 30 iterations. As pointed out in the previous sections, the adjoint-derived "gradient" is the G-differential when "on-off" switches are included in the model since the gradient is discontinuous in such a case. This experiment shows that the adjoint-calculated G-differential is sufficient to provide the minimization procedure a descent direction. Fig. 11a,b shows the model-predicted 3-h accumulated rainfall during the 6-h forecast, starting from the optimal IC after 30 iterations. We find that the position of the model-predicted 3-h rainfall compares very well with the observed 3-h rainfall (Fig. 1a). The intensity of the simulated rain (9.35 mm) is weaker than the observed rain (14.5 mm). This may be due to the model's deficiency since we used a rather simplified version of MM5 and a relatively coarse resolution (40 km). The subsequent 3-h rainfall predicted from the optimal IC (Fig. 11b) also occurs at the right place (see Fig. 1b), although the amount of the rain is less than the observed value. Comparing Fig. 11b with Fig. 1d, we find that the fictitious rain over Colorado and south Wyoming in the no-data assimilation run (Fig. 1d) is removed after assimilation (Fig. 11b).

In the second experiment (Exp. 2), not only 3-h rainfall, but also wind, temperature, surface moisture and PW analyses at time 1800 UTC and 2100 UTC 10 April are assimilated into the model during the 3-h assimilation period. The cost function consists of

$$\begin{aligned} J(\boldsymbol{x}_0) = \sum_{r=1}^{2} (&W_{\mathrm{u}}(\boldsymbol{u}(t_r) - \boldsymbol{u}^{\mathrm{obs}}(t_r))^2 \\ &+ W_{\mathrm{v}}(\boldsymbol{v}(t_r) - \boldsymbol{v}^{\mathrm{obs}}(t_r))^2 \\ &+ W_{\mathrm{T}}(\boldsymbol{T}(t_r) - \boldsymbol{T}^{\mathrm{obs}}(t_r))^2 \\ &+ \boldsymbol{W}_{q_{\mathrm{s}}}(\boldsymbol{q}_{\mathrm{s}}(t_r) - \boldsymbol{q}_{\mathrm{s}}^{\mathrm{obs}}(t_r))^2) \\ &+ \gamma_1 (P_{3\mathrm{h}} - P_{3\mathrm{h}}^{\mathrm{obs}})^2 \\ &+ \sum_{r=1}^{2} \gamma_2 (\mathrm{PW}(t_r) - \mathrm{PW}^{\mathrm{obs}}(t_r))^2, \end{aligned} \tag{5.12}$$

where W_{u}, W_{v}, W_{T}, and $W_{q_{\mathrm{s}}}$ are weighting coefficients for the wind, temperature and surface moisture, respectively, which are diagonal matrices obtained by taking the inverse of the square of the maximum difference between the observations at 1800 UTC (t_1) and 2100 UTC (t_2) 10 April. γ_1 and γ_2 are penalty coefficients taking the values of $\gamma_1 = 500$ and $\gamma_2 = 30$. Notice that we are dealing with the following situation: no observational error statistics, an unperfect assimilation model with simple physics, and rather coarse horizontal (40 km) and vertical (10 levels) resolutions.

Compared to Exp. 1, Exp. 2 produces a closer fit to the observed 3-h rainfall during the assimilation period (Fig. 11c), with a maximum value of 11.4 mm. The subsequent 3-h rainfall prediction (Fig. 11d) resembles the observations (Fig. 1d) very well. It is very encouraging, since the period between 2100 UTC 10 April and 2400 UTC 10 April is not in the assimilation window. Comparing Fig. 11b and 13d, we find that combining rainfall observation with conventional wind, temperature, and surface moisture and PW measurements produces a better short-range rainfall prediction. Fig. 12 shows the threat scores of the model predicted 3-h rainfall from analyzed IC,

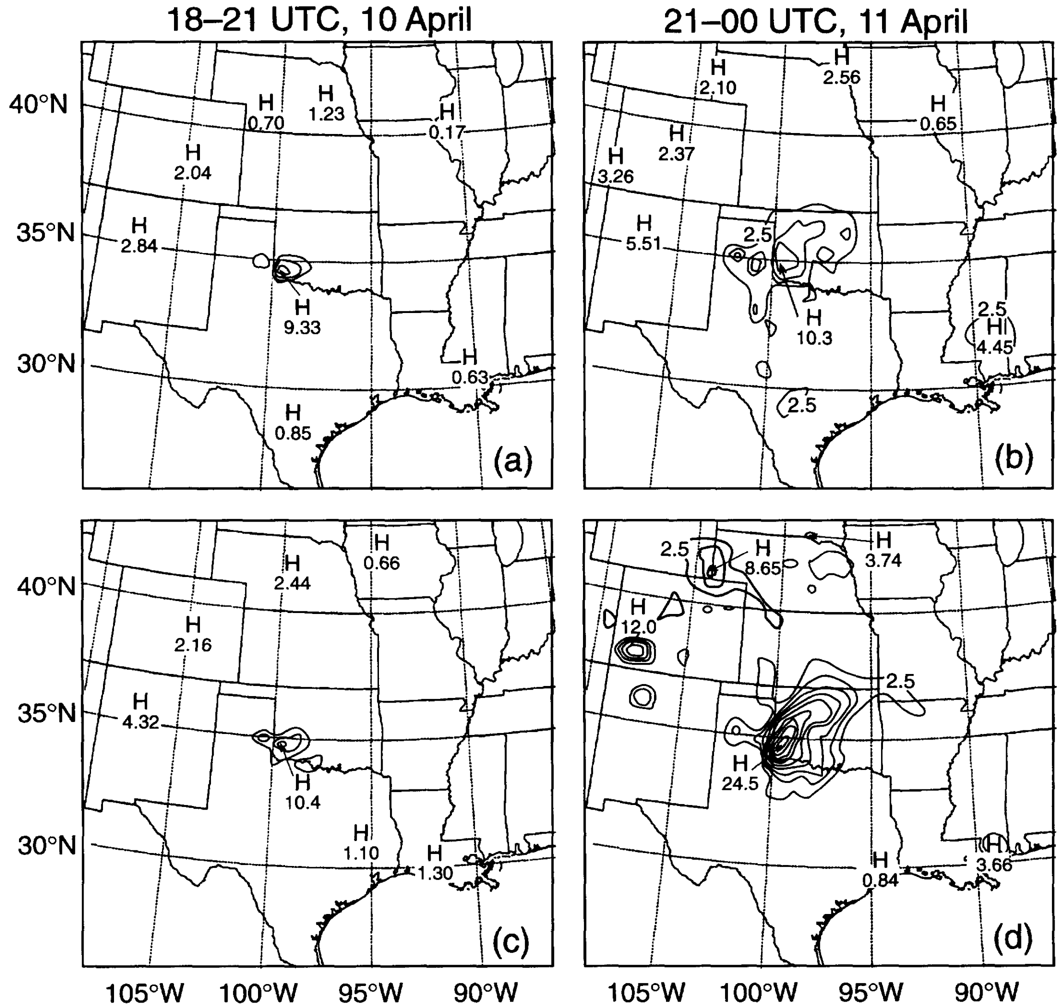


Fig. 11. The model-predicted 3-h rainfall (mm) during the 6-h forecast, starting from the optimal IC in Exp. 1 (a and b) and 2 (c and d) after 30 iterations. Contour interval is 2.5 mm.

optimal IC in Exp. 1, and the optimal IC in Exp. 2 at a threshold value of 2.5 mm, 5.0 mm and 10.0 mm. We observe that the forecast in Exp. 2 gives the highest threat score. However, the forecast skill drops quickly after 6-h prediction.

In order to see what causes the improvement of the 3-h rainfall prediction after 3-h 4DVAR, we plot the difference fields between the optimal IC and the analysis in Fig. 13. The left panel is the difference between Exp. 1 and analysis and the right panel is the difference between Exp. 2 and analysis. We find a maximum of 1°C warm area over southern Oklahoma at $\sigma = 0.85$ near the precipitation area in Exp. 2 (Fig. 13b). Not much difference is found in the low-level moisture field in Exp. 2 (Fig. 13d). The u-component of the upper-level jet ($\sigma = 0.35$) over northwest Texas is intensified by an amount of 2.94 ms^{-1}, and the zonal wind over eastern Texas is weakened in Exp. 2 (Fig. 13f). The difference in the low-level ($\sigma = 0.85$) meridional wind has a maximum on the Texas-and-Oklahoma border, which allows more moisture to be brought to the initial precipitation region (Fig. 13h). The difference in IC of Exp. 1 is

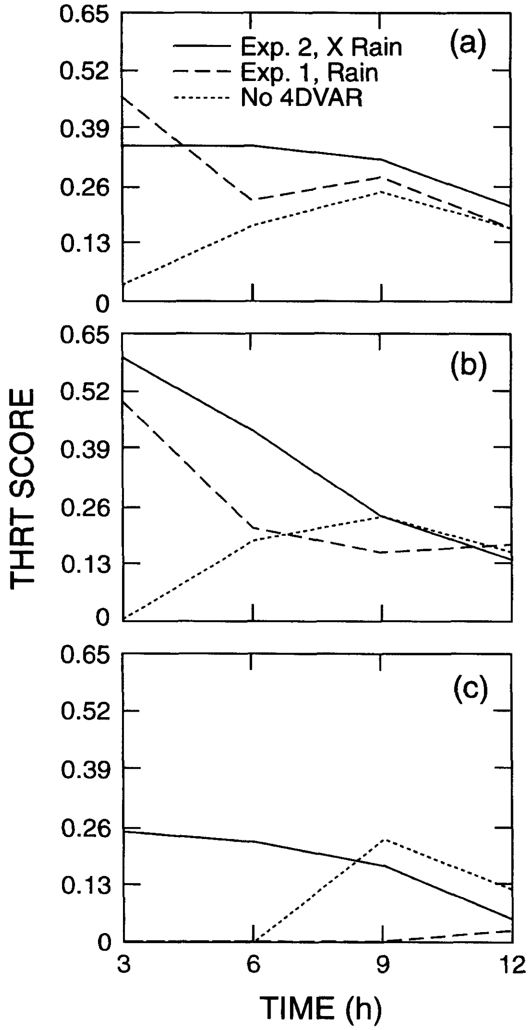


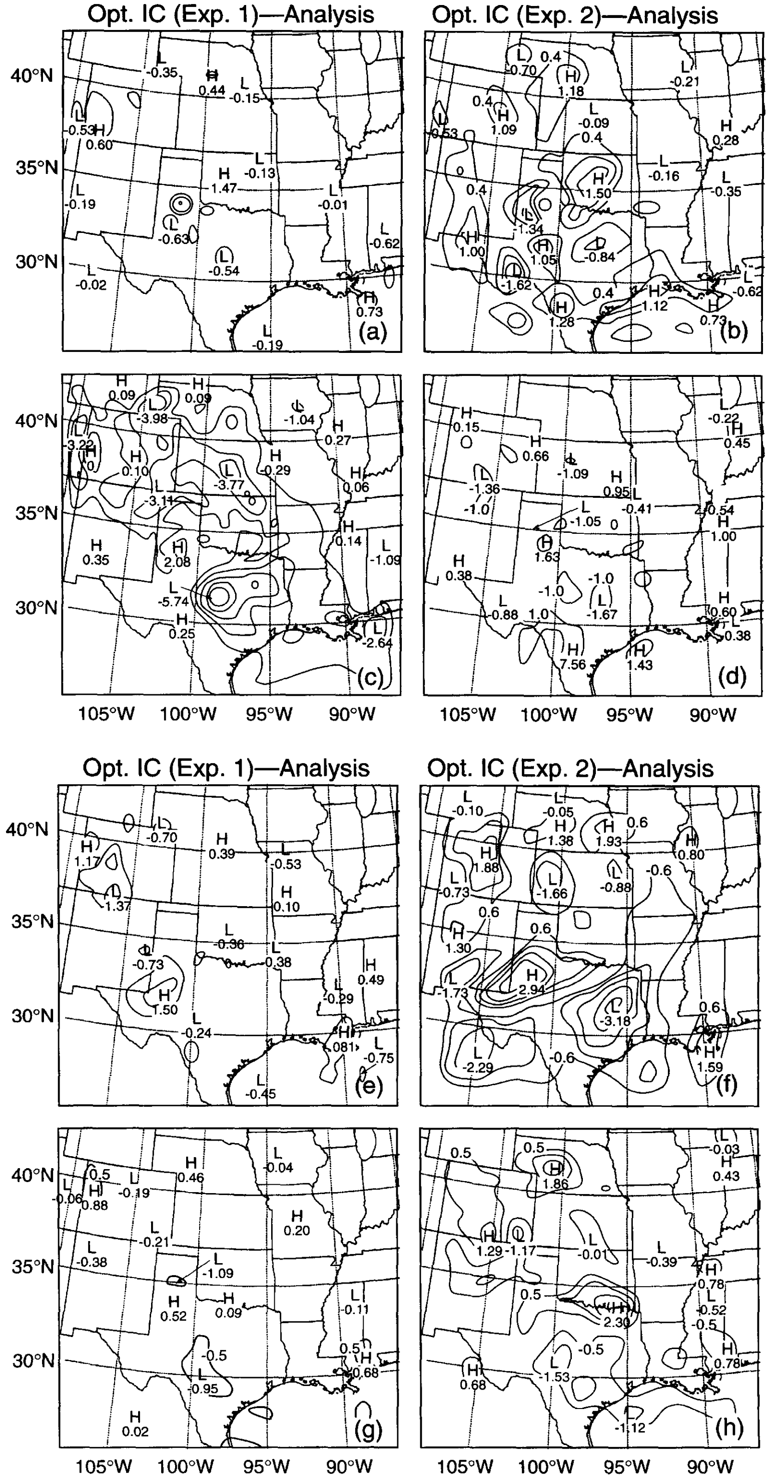
Fig. 12. The threat score of the model predicted 3-h rainfall from analyzed IC (dotted line), optimal IC in Exp. 1 (dashed line), and optimal IC in Exp. 2 (solid line) at the threshold value of (a) 2.5 mm, (b) 5.0 mm and (c) 10.0 mm.

relatively small and localized, except that a significantly large negative area of low-level specific humidity is found over central Texas.

One of the reasons that Exp. 2 outperforms Exp. 1 is that the wind, temperature, surface moisture, and PW observations at 2100 UTC 10 April are assimilated into the model, which will limit the freedom of the minimization procedure to arbitrarily adjust the IC. At 2100 UTC 10 April the RMS error of q in Exp. 2, calculated over all the 35 sounding stations, is smallest among 4DVAR experiment, Exp. 1, and Exp. 2 (figure omitted). The RMS error of q in Exp. 1 is higher than the no-data-assimilation result. It seems that in order to produce a realistic precipitation location in Fig. 1a as opposed to the one obtained in the straight model simulation (Fig. 1c), the 4DVAR procedure with the Kuo scheme generating convective rain decides to decrease the amount of moisture content over central Texas at the initial time. This will reduce the northward moisture advection and avoid precipitation occurring over the north of Texas as in Fig. 1c. This happens because there is no additional moist data constraint in the cost function. Therefore, it is better to add other moist data to rainfall observations during the assimilation procedure to reduce the arbitrariness of 4ADVR in adjusting the initial moisture field.

In Exp. 1 and 2, the smoothed Kuo scheme is used, which removes the $2\Delta t$ oscillations in the Kuo scheme as discussed in Section 5.1. In order to examine the impact of "badly behaved" "on-off" switches, an experiment (Exp. 3) similar to Exp. 2 is carried out, except that the original Kuo scheme is used during the assimilation procedure. Fig. 14 compares the variation of the cost function and the norm of the gradient with respect to IC between Exp. 2 and 3. Both Fig. 14a,b reveal that Exp. 2 and 3 have the same convergence rate, and Exp. 3 has a smaller value of the cost function through the assimilation procedure resulting from a closer fit to the observation with a complete Kuo scheme than with a modified version of Kuo scheme in which the stability check is turned off. Fig. 15 shows the distribution of the model-predicted 3-h rainfall in Exp. 3. Comparing to Fig. 11c,d and Fig. 1a,b, we find that the same improvement for the short-range rainfall predic-

Fig. 13. The difference fields between the optimal IC and the analysis. The left panel is the difference between Exp. 1 and analysis (no data assimilation) and the right panel the difference between Exp. 2 and analysis. (a) $T^{\mathrm{ptb}}_{\sigma=0.85}$, (b) $T^{\mathrm{tlm}}_{\sigma=0.85}$, (c) $q^{\mathrm{ptb}}_{\sigma=0.85}$, (d) $q^{\mathrm{tlm}}_{\sigma=0.85}$, (e) $u^{\mathrm{ptb}}_{\sigma=0.35}$, (f) $u^{\mathrm{tlm}}_{\sigma=0.35}$, (g) $v^{\mathrm{ptb}}_{\sigma=0.85}$, and (h) $v^{\mathrm{tlm}}_{\sigma=0.85}$. The units for T are °C, for q is g/kg, and for u and v is m/s. Contour intervals for T, q, u and v are 0.4°C, 1 g/kg, 0.6 m/s and 0.5 m/s, respectively.


Opt. IC (Exp. 1)—Analysis
Opt. IC (Exp. 2)—Analysis
40°N
35°N
30°N
(a)
(b)
(c)
(d)
105°W
100°W
95°W
90°W
Opt. IC (Exp. 1)—Analysis
Opt. IC (Exp. 2)—Analysis
(e)
(f)
(g)
(h)

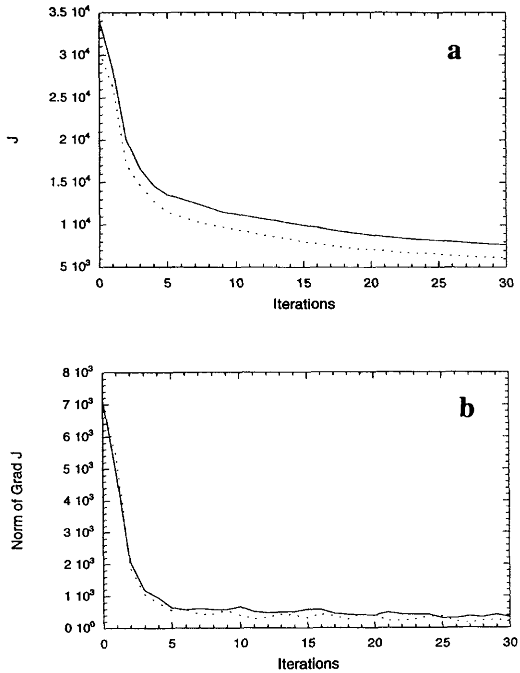


Fig. 14. Variation of the values of (a) the cost function and (b) the norm of the gradient with the number of iterations in Exp. 2 (solid line) and 3 (dotted line), in which the wind, temperature, surface moisture, PW and 3-h rainfall measurements are assimilated, during the minimization procedure with a 3-h assimilation window.

tion is observed in both Exp. 2 and 3. These results show that the "badly behaved" version of Kuo scheme does not decelerate the convergence rate, nor degrade the 4DVAR results.

6. Summary and conclusions

The tangent linear and adjoint versions of moist parameterization with "on-off" switches fixed and their feasibility for use in data assimilation are discussed by using both simple examples and the Penn State/NCAR mesoscale model MM5. Through a few analytic examples, we proved that for a continuous model, the TLM and adjoint model, developed by assuming that the "on-off" switches are kept the same as in the basic state and then linearizing the NLM around the basic state (TLM) and transposing it (adjoint model), may be feasible for 4DVAR under certain conditions. A sufficient condition for the TLM to provide a correct first-order approximation to the NLM solution, and the adjoint model to provide the *G*-differential of a cost function is that the nonlinear tendency equation is continuous at switching point. If not, errors may exist in the TLM solution and the adjacent calculated gradient. The magnitude of such errors depends on the magnitude of the increment added to the tendency equations at the switching point and the duration

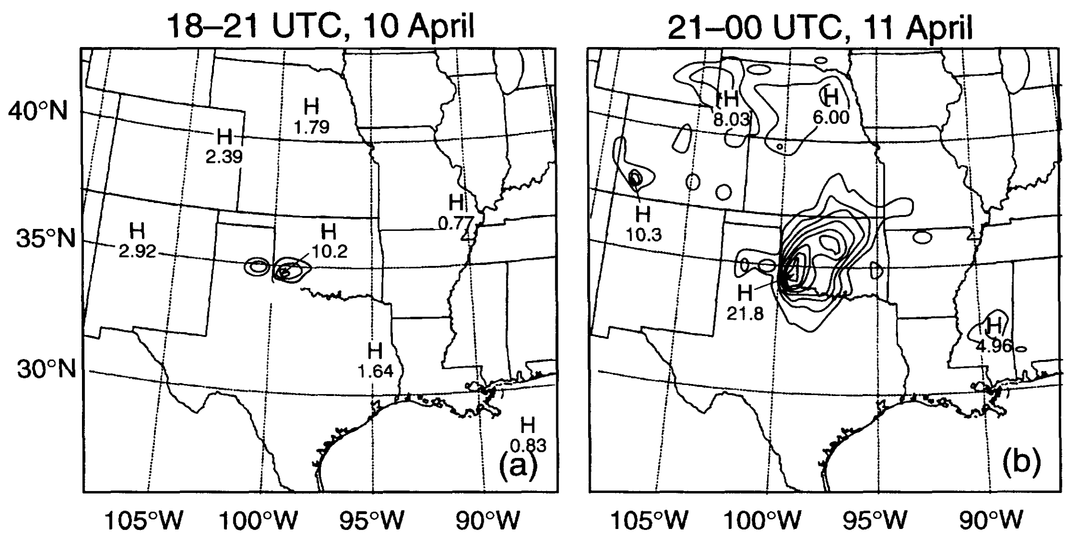


Fig. 15. The model-predicted 3-h rainfall (mm) in Exp. 3 during (a) 1800–2100 UTC 10 April, and (b) 2100–2400 UTC 10 April. Contour interval is 2.5 mm.

of moist processes being turned on. The larger the increment is comparing to other terms in the tendency equations, the bigger the errors will be.

However, examination of the discretized model — MM5, including the grid-resolvable non-convective precipitation and the Kuo cumulus parameterization, indicates that the moist physics containing "on-off" switches seem to have no impact on the validity of the TLM and the adjoint model for use in 4DVAR, in which the gradient information is required. Presence of "on-off" switches in moist parameterizations increases the degree of nonlinearity of the forward NLM solution, due to the strong nonlinearity introduced by moist physics. The forecast from a small perturbation may thus experience quick growth during the first few hours. Such an increase in the magnitude of the perturbation solution $x_n' = \boldsymbol{x}_n(\boldsymbol{x}_0 + \Delta \boldsymbol{x}_0) - \boldsymbol{x}_n(\boldsymbol{x}_0)$ will cause the TLM solution $x_n'^{\text{tlm}}$ to become a poor approximation to x_n' after a certain time of integration, since the TLM is linearized around the time-dependent basic state, and is accurate for small enough perturbations at every time step. Such an increase in the nonlinearity might cause the minimization problem to have more difficulty converging since the performance of the minimization depends upon the structure of the Hessian matrix of the cost function (Zou et al., 1993), which is not experienced in this research.

The 4DVAR experiments assimilating the wind, temperature, surface moisture, PW and 3-h rainfall observations using a diabatic version of MM5 are carried out over a 3-h window from 1800 UTC to 2100 UTC 10 April during the SESAME 1979 storm case. We find that (i) the 4DVAR using a diabatic version of MM5 and its adjoint, including "on-off" switches, works successfully; (ii) combining the rainfall observations with the wind, temperature, surface moisture, and PW results in a better rainfall prediction than the one assimilating only the rainfall observations; and (iii) switching on and off at every other time step does not make the minimization fail, and similar results are obtained as far as the convergence rate and improvement to the short-range rainfall prediction are concerned.

In this research, we use the NLM solution as the basic state in the TLM and adjoint model at every time step. The "on-off" switches are recalculated at every time step. In practice, this may present a computational problem concerning both the disk space and CPU time. If MM5 and its adjoint model are to be used for very high-resolution problems, which is one of our future goals, we need to consider a less frequent basic-state update cycle while maintaining good accuracy of both the TLM and the adjoint model. Preliminary results show that using less frequent updates of the basic state, say, every half-hour, will still give a very high accuracy for the adiabatic version of MM5 and its adjoint. However, in the presence of "on-off" switches, this presents a problem for the accuracy of both the tangent linear model and adjoint model. Further research is underway to use a less frequent basic-state update cycle, yet to retain the accuracy of the TLM and adjoint model, probably by only updating "on-off" switches at every time step.

7. Appendix

The cloud top (K_{Ctop}) and the cloud base (K_{Cbase}) are determined by the following procedures:

(1) Given an amount of $T'(=1\text{ K})$ and $q'(=10^{-3}\text{ kg kg}^{-1})$ perturbation to the model temperature and specific humidity, calculate the equivalent potential temperature $(\theta_e)_k$ for the perturbed temperature T_k^m and the specific moisture q_k^m

$$(\theta_e)_k = \frac{T_k^m}{(p/1000)^{R/C_p}}\, e^{Lq_k^m/c_p T_k^m}, \tag{A1}$$

where $T_k^m + T_k + T'$, $q_k^m = q_k = q'$, T_k and q_k are the temperature and specific humidity of the air at level k before the cumulus parameterization, $L = 2.5 \times 10^6\text{ J kg}^{-1}$ is the latent heat of liquid water, $c_p = 1004\text{ J deg}^{-1}\text{ kg}^{-1}$ is the specific heat at constant pressure, and $R = 287\text{ J kg}^{-1}\text{ k}^{-1}$ is the Universal gas constant.

(2) Find the lifting condensation level (LCL) below 700 hpa at which $(\theta_e)_k$ reaches maximum

$$(\theta_e)_{\text{LCL}} \geqslant (\theta_e)_k \qquad \text{for} \qquad p_k \geqslant 700\text{ hpa}. \tag{A2}$$

The cloud base is defined as the layer where the LCL is located: $k_{\text{Cbase}} = \text{LCL}$.

(3) Compute the saturate potential temperature $(\theta_s)_k$ with T_k:

$$(\theta_s)_k = \frac{T_k}{(p/1000)^{R/c_p}}\, e^{L(g_s)_k/c_p T_k} \tag{A3}$$

The cloud top is defined as the layer above k_{Cbase} on which

$$(\theta_s)_k - (\theta_e)_k > 3K, \tag{A4}$$

if the above condition is not satisfied for all $k < k_{\mathrm{Cbase}}$, $k_{\mathrm{Ctop}} = 1$. Note that the model levels are counted from the top to the bottom.

8. Acknowledgments

The authors thank O. Talagrand, R. Errico, J. Dudhia, and two anonymous reviewers for their useful comments on the manuscript.

REFERENCES

Andersson, E., Pailleux, J., Thépaut, J-N., Eyre, J. R., McNally, A. P., Kelly, G. A. and Courtier, P. 1994. Use of cloud-cleared radiances in three/four dimensional variational data assimilation. *Q. J. R. Meteor. Soc.* **120**, 627–653.

Bao, J. W. and Kuo, Y.-H. 1995. On-off switches in the adjoint method. Step functions. *Mon. Wea. Rev.* **123**, 1589–1594.

Bao, J. W. and Warner, T. T. 1993. Treatment of on/off switches in the adjoint method: FDDA experiments with a simple model. *Tellus* **45A**, 525–538.

Cacuci, D. G. 1981. Sensitivity theory for nonlinear systems: 1. Nonlinear functional analysis approach. *J. Math. Phys.* **22**, 2794–2802.

Douady, D. and Talagrand, O. 1990. The impact threshold process on variational assimilation. Preprints. *WMO International Symposium on Assimilation of observations in meteorology and oceanography.* Clermont-Ferrand, France, 486–487.

Dudhia, J. 1993. A nonhydrostatic version of the Penn State-NCR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.* **121**, 1493–1513.

Gill, P. E., Murray, W. and Wright, M. H. 1981. *Practical optimization*, United States Edition published by Academic Press, Inc., Orlando, Florida, 401 pp.

Grell, G. A., Dudhia, J. and Stauffer, D. R. 1994. A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). *NCAR Technical Note*, NCAR/TN-398 + STR, National Center for Atmospheric Research, Boulder, CO, 138 pp.

Gustafsson, N. and Lönnberg, P. 1995. Data assimilation for high resolution limited area models. *The Geophysical Magazine*, series 2, Special Issue, collection of lecture notes presented at *The WMO 2nd International Symposium on Assimilation of observations in meteorology and oceanography*, 13–17 March 1995, Tokyo, Japan, V-i-1–V-i-20.

Kuo, Y.-H., Guo, Y.-R. and Westwater, Ed R. 1993. Assimilation of precipitable water measurements into a mesoscale numerical model. *Mon. Wea. Rev.* **121**, 1215–1238.

Kuo, Y.-H., Zou, X. and Guo, Y.-R. 1995. Variational assimilation of precipitable water using a nonhydrostatic mesoscale adjoint model. Part I: Moisture retrieval and sensitivity experiments. *Mon. Wea. Rev.* **124**, 122–147.

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

Lorenc, A. 1995. Development of an operational variational assimilation scheme. World Meteorological Organisation. *The 2nd International Symposium on Assimilation of observations in meteorology and oceanography*, 13–17 March 1995, Tokyo, Japan, 415–420.

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

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

Parrish, D. F. and Derber, J. C. 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.

Qiu, C. J. and Xu, Q. 1992. A simple adjoint method of wind analysis for single-Doppler data. *J. Atmos. & Oceanic Technology* **9**, 588–598.

Sagan, H. 1969. *Introduction to the calculus of variations.* New York, 449 pp.

Sun, J. Z., Flicker, D. W. and Lilly, D. K. 1991. Recovery of three-dimensional wind and temperature fields from single-Doppler radar data. *J. Atmos. Sci.* **48**, 876–890.

Thépaut, J. N. and Courtier, P. 1991. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive equation model. *Q.J.R.M.S.* **117**, 1225–1254.

Thépaut, J. N., Hoffman, R. N. and Courtier, P. 1993. Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.* **121**, 3393–3414.

Verlinde, J. and Cotton, W. R. 1993. Fitting microphysical observations of non-steady convective clouds to a numerical model: An application of the adjoint technique of data assimilation to a kinematic model. *Mon. Wea. Rev.* **121**, 2776–2793.

Vukicevic, T. and Errico, R. M. 1993. Linearization and adjoint of parameterized moist diabatic processes. *Tellus* **45A**, 493–510.

Xu, Q. 1996. Generalized adjoint for physical processes with parameterized discontinuities — Part I: Basic issues and heuristic examples. *J. Atmos. Sci.* **53**, 1123–1142.

Zou, X., Kuo, Y.-H. and Guo, Y.-R. 1995. Assimilation of atmospheric radio refractivity using a nonhydrostatic adjoint model. *Mon. Wea. Rev.* **123**, 2229–2249.

Zou, X., Navon, I. M. and Sela, J. G. 1993a. Control of gravitational oscillations in variational data assimilation. *Mon. Wea. Rev.* **121**, 272–289.

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

Zou, X., Navon, I. M., Berger, M., Phua, Paul K. H., Schlick, T. and Le Dimet, F. X. 1993c. Numerical experience with limited-memory quasi-Newton and truncated Newton methods. *SIAM J. on Optimization* **3**, 582–608.

Zupanski, M. and Zupanski, D. 1995. Recent developments of the NMC's regional four-dimensional variational data assimilation system. World Meteorological Organisation. *The 2nd International Symposium on Assimilation of observations in meteorology and oceanography*, 13–17 March 1995, Tokyo, Japan, 367–372.

Zupanski, D. 1993. The effect of discontinuities in the Betts–Miller cumulus convection scheme on four-dimensional data assimilation. *Tellus* **45A**, 511–524.

Zupanski, D. and Mesinger, F. 1995. Four-dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.* **123**, 1112–1127.