Tellus (1993), 45A, 511–524
Printed in Belgium – all rights reserved

Copyright © Munksgaard, 1993
TELLUS
ISSN 0280–6495

The effects of discontinuities in the Betts–Miller cumulus convection scheme on four-dimensional variational data assimilation

By DUŠANKA ŽUPANSKI*, *NMC/UCAR Visiting Research Program, NMC/WWB2, Room 204, 5200 Auth Road, Camp Springs, MD 20746, USA*

(Manuscript received 6 November 1992; in final form 24 May 1993)

ABSTRACT

A tangent linear and an adjoint of the large-scale precipitation and the cumulus convection processes in the National Meteorological Center's NMC/ETA regional forecast model are developed. The effects of discontinuities in the Betts–Miller cumulus convection scheme are examined and applicability of derivative minimization methods in four-dimensional variational (4D VAR) data assimilation is considered. It is demonstrated that discontinuities present in the control Betts–Miller cumulus convection scheme increase linearization errors to a large extent and have adverse effects on 4D VAR data assimilation. In the experiments performed, discontinuities in the cumulus convection scheme have the most serious effect in low layers. These problems can be reduced by modifying the scheme to make it more continuous in low layers. Positive effects of inclusion of cumulus convection in 4D VAR data assimilation are found in upper layers, especially in humidity fields. The "observations" used are optimal interpolation analyses of temperature, surface pressure, wind and specific humidity. By inclusion of other data, more closely related to the convective processes, such as precipitation and clouds, more benefits should be expected. Even with the difficulties caused by discontinuities, derivative minimization techniques appear to work for the data assimilation problems. In order to get more general conclusions, more experiments are needed with different synoptic situations. The inclusion of other important physical processes such as radiation, surface friction and turbulence in the forecast and the corresponding adjoint models could alter the results since they may reinforce the effects of discontinuities.

1. Introduction

In 4D VAR data assimilation, a misfit between the forecast and the observations is minimized. The mesure of this misfit is often called the functional, or objective function, and denoted J. Commonly, the forecast model is taken as a strong constraint in minimization of J. Then, by backward integration of the corresponding adjoint model, the first derivative of J with respect to some control variable $\mathbf{x}$ (the variable to be altered in the minimization procedure) can be calculated and used to minimize J. The vector of initial conditions for the forecast model is usually chosen to be the control variable, but other variables (e.g., boundary conditions, model parameters etc.) could be chosen as well.

The first successful 4D VAR data assimilation experiments (Lewis and Derber, 1985; LeDimet and Talagrand, 1986; Hoffman, 1986; Courtier and Talagrand, 1987; etc.) were performed using simple (quasi-geostrophic, shallow water) forecast and adjoint models. Later on, simple models were replaced by adiabatic primitive equations models (e.g., Thepaut and Courtier, 1991; Navon et al., 1992). Further progress was made by developing techniques to use the observations directly (Parrish and Derber, 1992; Thepaut et al., 1993). Model errors are taken into account (Derber, 1989; Wergen, 1992; M. Županski, 1993a) as well. A natural step forward is to begin using more com-

* On leave from the Institute of Physics, Belgrade, Yugoslavia.

plex models with physical processes in 4D VAR data assimilation. One possible way to do this is proposed in M. Županski (1993a). He uses the primitive equation forecast model with full physics to define basic states and functional J. The gradient of J is however, calculated using the adjoint model which includes only dry adiabatic processes and horizontal diffusion. The experimental results showed that this method was able to improve the forecast, but the final solution probably was not optimal since the diabatic processes were not properly accounted for in the adjoint model. Another serious limitation of this method is that only observations related to dry model dynamics can be assimilated. In order to use other observations related to the physics (humidity, precipitation, cloud water etc.) it is necessary to include physical processes into the adjoint model.

The major problem with using forecast models with physical processes is that they are highly nonlinear and often contain discontinuities in the parameterizations. The functional J might not be a differentiable function with respect to control variables in that case. As a consequence, "smooth" (or "derivative") minimization methods (e.g., quasi-Newton, conjugate gradient etc.) might not be applicable. These problems were studied in the context of 4D VAR data assimilation first by Douady and Talagrand (1990). Because of the difficulties caused by discontinuities in a one-dimensional convective adjustment model the use of other minimization methods was suggested as a possibility. These alternative minimization techniques, usually called "nonsmooth" or "nonderivative" methods (e.g., function comparison; simplex (Gill et al., 1981); subgradient method (Rockafellar, 1981) etc.) do not suffer of problems related to discontinuities. Their computational cost and memory requirements however, are extremely high and not practical in application to realistic 4D VAR data assimilation.

Numerical experiments with a simple one-dimensional convection model involving on/off switches associated with latent heat release are preformed by Bao and Warner (1993, this issue). No serious problems due to discontinuities are found, though the conditions of the experiments are academic and the conclusions might not be generally applicable. Discontinuities in a two-dimensional kinematic microphysical model were studied and the approach of making processes more continuous was examined by Verlinde (1992). A spline was fitted over the discontinuity to make the model continuous. The experimental results showed a substantial improvement of the performance of minimization in that case.

The same discontinuities might have different effects however, as studied in a more complex primitive equation models. The effects of discontinuities, present in the parameterization of convective processes, on 4D VAR data assimilation in that case might be small as compared to other effects because of local character of convection. The 4D VAR data assimilation experiments, done by Zou et al., 1993 (this issue), using a global spectral model which includes cumulus convection parameterization show negligible effects of discontinuities. The experiments performed by Vukićević and Errico, 1993 (this issue) using the full physics PSU-NCAR mesoscale grid point model, show that discontinuous transition to the convective regime could have negative effects on sensitivity experiments, but the effects on 4D VAR data assimilation are not studied. Obviously, these results make the question about the effects of such discontinuous processes even more intriguing.

In Section 2, the Betts–Miller cumulus parameterization scheme is described and data assimilation problem using this scheme in the forecast and corresponding adjoint model is explained. The approach of making convective processes more continuous is examined as a possible solution. Two modified, more continuous versions of the cumulus convection scheme are described and the effects of the modifications in reducing discontinuities are discussed in the same section. In Section 3 the numerical models: nonlinear, linear and adjoint, and data used in the experiments are described. Linearization and data assimilation experiments are presented and the results are discussed in Section 4. Finally, in Section 5 the conclusions are drawn and future experiments are discussed.

2. The Betts–Miller scheme and data assimilation problem

2.1. *The Betts–Miller cumulus parameterization scheme*

In the Betts–Miller cumulus parameterization scheme (Betts, 1986; Betts and Miller, 1986) is a nudging type adjustment of a grid scale temperature (T) and humidity (q) described by the following equations:

$$T^{n+1} = T^n + \frac{T_r^n - T^n}{\tau} \Delta t, \tag{1a}$$

$$q^{n+1} = q^n + \frac{q_r^n - q^n}{\tau} \Delta t, \tag{1b}$$

where τ is a prescribed relaxation time, superscript n denotes the time step in which the convection takes place and subscript r stands for the quasi-equilibrium ("referent") profile. The equations (1a, b) are applied to convectively unstable grid points in the layers between the bottom (L_B) and the top (L_T) of convection. The bottom and top layers are determined differently depending on the type (deep or shallow) of convection. In shallow convection L_B and L_T coincide with cloud base and cloud top, respectively. In deep convection L_T is the cloud top layer, but L_B is prescribed to be the second layer above the ground. Below L_B or above L_T and in convectively stable points no adjustment is done. Convectively unstable grid points are selected using a buoyancy criteria based on the moist adiabat θ_{ES} (Betts, 1982; 1983). A prescribed vertical profile of saturation pressure difference ($p_s - p$) and a constant value of θ_{ES} during the parcel ascent is assumed to define first guess values for T_r and q_r above the freezing level. To calculate first guess of T_r and q_r below the freezing level a different profile of $p_s - p$ is applied and θ_{ESV} (moist virtual adiabat, Betts, 1983) is used instead of θ_{ES}. Starting from the first guess values, the referent profiles are adjusted iteratively to satisfy the enthalpy conservation constraint:

$$\Delta H = \sum_{L=L_B}^{L_T} \left(c_p \frac{(T_r - T)_L}{\tau} + E_V \frac{(q_r - q)_L}{\tau} \right) \Delta p_L = 0, \tag{2}$$

where c_p is dry air specific heat at constant pressure, E_V is latent heat of vaporization and Δp_L is pressure thickness of the layer L. The adjusted referent profiles are used to calculate T^{n+1} and q^{n+1} by (1a, b). The precipitation rate is calculated as:

$$\mathrm{PR} = -\sum_{L=L_B}^{L_T} \frac{(q_r - q)_L}{\tau} \frac{\Delta p_L}{g} = \frac{c_p}{E_V} \sum_{L=L_B}^{L_T} \frac{(T_r - T)_L}{\tau} \frac{\Delta p_L}{g}. \tag{3}$$

The eqs. (2) and (3) are used in the deep convection. In the shallow convection, the precipitation rate (3) and each term in (2) are set to be equal zero.

The scheme used in this study (referred hereafter as the control scheme) is somewhat different then the original Betts–Miller scheme. As applied to the specific model (regional NMC/ETA grid point model with step-like mountains and "eta" vertical coordinate, described in Section 3), the Betts–Miller scheme was modified to improve the performance of the overall physical package. The modifications are partially described in a paper by Janjić, 1990. The major differences between the original Betts–Miller scheme and the control scheme can be summarized as follows. A distinction is made between equilibrium humidity ($p_s - p$) profiles over land and over water. A family of land and water profiles ranging from extremely "fast" and dry to extremely "slow" and moist has been introduced, rather than a single water or land profile. The "fast" or "slow" profiles were motivated by the fast or slow moving line wake, described by Betts (1986). The choice of an appropriate equilibrium humidity profile from a family of profiles is controlled by the nondimensional "cloud efficiency" (C_{ef}) function calculated in time step n at the sea points as:

$$C_{ef}^n = 0.5 \left(C_{ef}^{n-1} + \text{const} \frac{E_{adj}^n}{E_{prec}^n} \right), \tag{4}$$

where E_{adj} is the entropy change due to convective adjustment of T and q, and E_{prec} is the entropy change due to precipitation, both calculated as vertically integrated values. The cloud efficiency is allowed to vary between 0.2 and 1 in the deep convection. In the sallow convection, where the precipitation is zero, C_{ef} is set to be equal to 0.6 over water. Over land is $C_{ef} = 1$ for both types of confection.

The cloud efficiency was included into the convection scheme to reduce spurious heavy precipitation over warm water generated in some synoptic situations (Z. I. Janjić, personal communication).

The second term of the equations (1a, b) is multiplied by C_{ef}^{n} so that the overall effect of this function is in enlarging the relaxation time τ for the points with much precipitation and little entropy change due to convective adjustment ($C_{\mathrm{ef}} < 1$).

There are some additional small differences between the control and the original scheme. Our experience is that these differences do not affect substantially the properties we are interested in: (non)linearities and (dis)continuities. The control scheme, rather than the original Betts–Miller scheme will be examined. The conclusions made by this study are expected to be applicable to the original Betts–Miller scheme (Betts, 1986; Betts and Miller, 1986) as well.

2.2. *The assimilation problem*

A simple 4D VAR data assimilation can be defined by minimizing the functional J given by:

$$J(\boldsymbol{x}) = \frac{1}{2} \sum_{n=0}^{N_{\max}} \langle (\boldsymbol{x}_n - \boldsymbol{x}_n^{\mathrm{obs}}), \boldsymbol{W}(\boldsymbol{x}_n - \boldsymbol{x}_n^{\mathrm{obs}}) \rangle. \tag{5}$$

The analysis vector $\boldsymbol{x}$ and the observation vector $\boldsymbol{x}^{\mathrm{obs}}$ belong to a N-dimensional Hilbert space, with the inner product $\langle \, , \rangle$ defined on it. The index n stands for different time levels, defined by t_n, such that $t_0 \leqslant t_n \leqslant t_{N_{\max}}$. The $N \times N$ matrix $\boldsymbol{W}$ defines weighting factors.

The numerical prediction model can be written by:

$$\boldsymbol{x}_n = G\boldsymbol{x}_{n-1}, \tag{6}$$

where G is a nonlinear forecast model operator. This model equation is used as a strong constraint in the minimization problem.

If derivative minimization algorithms are applied the functional J should be a differentiable function with respect to the control variable. In the case with the strong constraint (6) the control variable is usually chosen to be the vector of the initial conditions, $\boldsymbol{x}_0$. The differentiability of J with respect to $\boldsymbol{x}_0$ requires therefore, differentiability of the model operator G with respect to the model state variables.

Additional condition, desirable to improve the convergence of the minimization and to eliminate possibility of appearance of multiple minima (e.g., Tarantola, 1987) is the approximate linearity of the strong constraint, expressed by:

$$G(\boldsymbol{x} + \delta\boldsymbol{x}) \approx G(\boldsymbol{x}) + G'_x(\boldsymbol{x})\,\delta\boldsymbol{x}, \tag{7}$$

where $\delta\boldsymbol{x}$ is a small magnitude perturbation vector and G'_x is the Jacobian operator (i.e., the first derivative of G with respect to $\boldsymbol{x}$).

In order to have defined first derivative of G with respect to model state variables, the first derivatives of (1a, b) with respect to T^n, T_{r}^n, q^n and q_{r}^n should be defined. The equations (1a, b) are applied only to convectively unstable points in the layers between L_{B} and L_{T} (convective regime). A small change in T^n or q^n can however, result in the point being changed from convective to nonconvective regime or vice versa and in large changes in T^{n+1} and q^{n+1}. This transition between the convective and nonconvective regime is the most discontinuous aspect of the Betts–Miller scheme and its maximum effect is in the layers of transition L_{B} and L_{T}. Discontinuities in the convective regime itself, such as discontinuous transition between deep and shallow convection, discontinuities in T_{r}, q_{r}, C_{ef} etc. appear to have less effects in the experiments performed.

To eliminate the most serious discontinuity one could redefine the relaxation time τ (eq. 1a, b) as a continuous function of model state variables which would tend to infinity (or to a very large number, several orders of magnitude larger then the original τ) for the nonconvective regime. The equations (1a, b) with this new τ would be used everywhere, though the effect of convection would be negligible in some points. It is not clear however, that it is really necessary to make the transition between the two regimes completely continuous. The desired results may be achieved through reducing the discontinuity. Two different simple ways to make the convection schemes more continuous ("smooth" and "downdraft" schemes) are described below and discussed as possible solutions.

2.3. *Modifications of the convective scheme*

2.3.1. Smooth scheme. To reduce the abrupt jumps from nonconvective to convective regimes, the differences $\Delta T^n = T_{\mathrm{r}}^n - T^n$ and $\Delta q^n = q_{\mathrm{r}}^n - q^n$ in (1a, b) can be smoothed in the layers of transition (L_B and L_T) using the following equations (superscript n is omitted):

$$\Delta T(\eta) = (1 - YB(\eta)) \times (1 - YT(\eta)) \times \Delta T(\eta), \quad (8a)$$

$$\Delta q(\eta) = (1 - YB(\eta)) \times (1 - YT(\eta)) \times \Delta q(\eta), \quad (8b)$$

where the functions $YB(\eta)$ and $YT(\eta)$ are defined by:

$$YB(\eta) = 0.5 \tanh\left(\frac{a}{\Delta\eta^*}(\eta - \eta_B)\right) + 0.5, \quad (9a)$$

$$YT(\eta) = 0.5 \tanh\left(\frac{a}{\Delta\eta^*}(\eta_T - \eta)\right) + 0.5. \quad (9b)$$

Here, η is the vertical coordinate, $\eta_B = \eta(L_B)$ defines the bottom and $\eta_T = \eta(L_T)$ the top of convection. The parameter a is the slope of the hyperbolic tangent (tanh) function. The normalizing factor $\Delta\eta^*$ is chosen as:

$$\Delta\eta^* = \eta_B - \eta_T. \quad (10)$$

The function $YT(\eta)$ is equal 0.5 for $\eta = \eta_T$ and it asymptotically approaches 0 or 1 as the distance from η_T increases. The same is true for the function $YB(\eta)$, but with respect to η_B. According to (8a, b) the smoothing of $\Delta T(\eta)$ and $\Delta q(\eta)$ is strongest in the layers of transition (η_B and η_T) and becomes negligible far from these layers. The effect of the functions $YB(\eta)$ and $YT(\eta)$ may be understood as a selective smoothing through discontinuities in layers L_B and L_T, equivalent to enlarging τ in eqs. (1a, b). All other aspects of the smooth scheme are identical to the control scheme. The amount of smoothing is controlled by the steepness (a) of the hyperbolic tangent function.

2.3.2. Downdraft scheme. This scheme is the recent update of the Betts–Miller scheme (Betts and Miller, 1993). The parameterization is made more complex in order to get a smoother transition between two regimes rather then doing a formal smoothing. A parameterization of the unsaturated boundary layer downdraft mass flux (Betts and Miller, 1993) is added to the deep convection of the control scheme. The shallow convection is kept unchanged. In convectively unstable points, selected using the same procedure as in the control scheme, the eqs. (1a, b) are modified to include downdraft and read:

- in layers between L_{BL} and L_T:

$$T^{n+1} = T^n + \frac{T_r^n - T^n}{\tau}\Delta t, \quad (11a)$$

$$q^{n+1} = q^n + \frac{q_r^n - q^n}{\tau}\Delta t; \quad (11b)$$

- in layers between L_B and L_{BL}:

$$T^{n+1} = T^n + \frac{T_{BL}^n - T^n}{\tau_{BL}}\Delta t, \quad (12a)$$

$$q^{n+1} = q^n + \frac{q_{BL}^n - q^n}{\tau_{BL}}\Delta t. \quad (12b)$$

Here, T_{BL} and q_{BL} are referent profiles for the downdraft, slightly different (usually cooler and dryer) from T_r and q_r. Layer L_{BL} is related to downdraft inflow level and prescribed to be between L_B and L_T at about 850 mb. In the layers below L_B or above L_T and in convectively stable points no adjustment is done. As it can be seen, the upper layers convective adjustment (eq. 11a, b) is identical to the control scheme, except for different T_r and q_r, resulting from the enthalpy conservation adjustment (2). In this case T_{BL}, q_{BL} and τ_{BL} are used below, and T_r, q_r and τ above L_{BL}, but only T_r and q_r are allowed to change during the adjustment in (2). Eqs. (12a, b) describe the convective adjustment in subcloud layers where the downdraft is active. The relaxation time τ_{BL} is calculated as a continuous function of precipitation and grid scale T and q (except for the case when the convection is shut down by the same on/off switches as in the control scheme).

Generally, τ_{BL} is an order of magnitude larger then τ. Theoretically, by using different equations for the adjustment below and above downdraft inflow level, new discontinuities are introduced. Practically, equations (12a, b) help to make the transition between convective and nonconvective regimes smoother in low layers because of larger relaxation time τ_{BL}. For this reason, the downdraft scheme is expected to have similar effect as the smooth scheme in reducing discontinuities. This scheme is however, physically more appealing because of the parameterization of the processes of cooling and drying by the downdrafts, supported by the observations.

3. Numerical models and data

The limited area primitive equation grid point model (Mesinger et al., 1988; Janjić, 1990) with step-like mountains and η vertical coordinate is used in this study. The domain of integration is the North American continent with adjacent waters. The horizontal resolution is approximately 80 km. There are 16 eta layers in the vertical, with the top at 100 hPa. This is the future operational (or quasi-operational) NMC version of the model. The model used in the experiments includes dynamics, horizontal diffusion large scale precipitation and cumulus convection. The time step is 200 s for gravity-inertia waves, 400 s for advection and large scale precipitation and 800 s for other physical processes (cumulus parameterization in our experiments). Depending on the particular experiment, a more or less complex version of the model is used. In all experiments, forecast, tangent linear and adjoint models were consistent with each other (i.e., the same processes included). The use of inconsistent forecast and adjoint models in 4D VAR data assimilation would result in the use of inappropriate first derivative of the forecast model operator G (eq. (6)) and incorrect gradient of the functional, so that it was necessary to eliminate these effects by excluding from the forecast model all processes not present in the adjoint model.

The tangent linear NMC/ETA model and its conjugate transpose, the adjoint model, include in the most complex versions, the following processes: dynamics, horizontal diffusion, large scale precipitation and cumulus convection (control, smooth, or downdraft) scheme. The domain of integration, space and time resolution are the same as in the nonlinear forecast model. The linearization is made with respect to the basic state defined by the nonlinear model, updated every time step. It was necessary to update some basic state variables in each iteration of the adjustment in cumulus convection scheme. Considerable reduction of the tangent linear model errors is found by including this update. The on/off switches are determined by the basic state and the convection of the linearized models is effective in the same points as in the nonlinear forecast model.

Gridded data, provided by NMC RDAS (Regional Data Assimilation System), (Di Mego, 1988) are used in our experiments. OI analyses of temperature, wind (u, v), surface pressure (p_s) and specific humidity are chosen for two different synoptic situations (winter and summer). The first data set consists of analyses, available every 3 h, starting from 1200 UTC 27 December and ending 00 UTC 28 December 1991. This is the same data set (except for humidity) used in assimilation experiments by M. Županski (1993a, b). The second data set (used only in linearization experiments) consists of OI analyses of T, u, v, p_s and q valid 1200 UTC 9 July 1991.

4. Experiments

Two groups of experiments are performed: (1) linearization and (2) assimilation experiments.

4.1. *Linearization experiments*

Linearization experiments are performed to test the applicability of the assumptions of linearity and differentiability. The tangent linear model is run and compared to a pair of perturbed and unperturbed nonlinear forecasts. OI analyses (described in Section 3) are used to define the initial conditions for the nonlinear model. The unperturbed forecast is used to define the basic state for the tangent linear model. To define perturbations the nonlinear forecast model is run twice starting from the same analysis, but with different integration time: 12 or $12 + \Delta H$ hours (ΔH given in Table 1). The difference between these two nonlinear forecasts serves as the initial perturbation for the tangent linear model and for the perturbed nonlinear forecast model. The comparison between the nonlinear forecast difference and the linear forecast is made after additional 12 hours of integration of the two nonlinear and the tangent linear models. The forecast difference, rather than a random perturbation vector, is used to define initial perturbation to reduce the effect of spurious gravity waves. The boundary conditions for both nonlinear forecast models are identical and perturbations at the boundaries are set to be equal to zero throughout the integration of the tangent linear model. For convenience, the boundary conditions for the nonlinear models are kept constant. Our hypothesis is that constant boundary conditions do not change substantially a degree of linearity and continuity of the forecast model. The

Table 1. *Conditions of the 4 experiments to test linearization (Section 4.1): synoptic situation, perturbed and unperturbed forecast integration time difference ΔH, and the initial perturbation norm $\|\Delta \boldsymbol{x}\|$, scaled by 10^{-2}, for five different models*

Exp. no.	Synoptic situation and ΔH	model	$\|\Delta x\| = \|\Delta x\| \times 10^{-2}$
1	winter $\Delta H = 1$ h	NO PHYSICS	$\|\Delta x\| = 6.79$
		NO CONVECTION	$\|\Delta x\| = 6.72$
		CONTROL	$\|\Delta x\| = 6.93$
		SMOOTH	$\|\Delta x\| = 6.78$
		DOWNDRAFT	$\|\Delta x\| = 6.76$
2	winter $\Delta H = 6$ h	NO PHYSICS	$\|\Delta x\| = 27.23$
		NO CONVECTION	$\|\Delta x\| = 26.61$
		CONTROL	$\|\Delta x\| = 27.12$
		SMOOTH	$\|\Delta x\| = 26.80$
		DOWNDRAFT	$\|\Delta x\| = 26.72$
3	summer $\Delta H = 1$ h	NO PHYSICS	$\|\Delta x\| = 4.06$
		NO CONVECTION	$\|\Delta x\| = 4.02$
		CONTROL	$\|\Delta x\| = 4.20$
		SMOOTH	$\|\Delta x\| = 4.11$
		DOWNDRAFT	$\|\Delta x\| = 4.09$
4	summer $\Delta H = 6$ h	NO PHYSICS	$\|\Delta x\| = 17.88$
		NO CONVECTION	$\|\Delta x\| = 17.58$
		CONTROL	$\|\Delta x\| = 17.76$
		SMOOTH	$\|\Delta x\| = 17.62$
		DOWNDRAFT	$\|\Delta x\| = 17.59$

tests performed with variable boundary conditions in the nonlinear model showed very similar results.

As in the experiments conducted to test linearization of the PSU-NCAR meso-scale dry adiabatic model (Errico and Vukićević, 1992), correlation coefficients between the nonlinear forecast difference and tangent linear forecast are calculated. Unlike in Errico and Vukićević (1992), a shorter (12 h long) integration period is chosen because of nonlinearities and discontinuities present in our more complex model.

Five different models are used in the experiments. The differences are basically in the degree of complexity. The simplest model is the one with dynamics and horizontal diffusion only, called here NO PHYSICS. A more nonlinear and more discontinuous model is one including dynamics and horizontal diffusion + large scale precipitation, but no convection (model called NO CONVECTION). The most discontinuous model is the most complex one including dynamics and horizontal diffusion + large scale precipitation + cumulus convection. Depending on the particular choice of cumulus convection scheme the three versions of the most complex model are used, called CONTROL, SMOOTH or DOWNDRAFT. The same level of complexity is kept in the nonlinear and tangent linear models.

Experiments with the five different models conducted to examine the effects of different initial perturbations are listed in Table 1. Experiments no. 1 and no. 2 are performed for the winter and the experiments no. 3 and no. 4 for the summer case. The only difference between the exp. no. 1 and no. 2, as well as between the no. 3 and no. 4, is in the initial perturbation vector. The perturbation norms ($\|\Delta \boldsymbol{x}\|$) are larger for $\Delta H = 6$ h (exp. nos. 2 and 4) than for $\Delta H = 1$ h (exp. nos. 1 and 3) for all models. In calculating the norms the same weighting factors (diagonal matrix $\boldsymbol{W}$) as in the functional J_{obs} (eq. (17) in Subsect. 4.2) are used. The average magnitudes of the initial perturbations in the exp. no. 1 are 1 K for temperature, 0.5 mb for pressure and 0.5 m/s for wind. In the exp. no. 2, the initial perturbations are about 4× larger. These perturbations seem to be a reasonable choice because of necessity to deal with perturbations of about the same order of magnitude in minimization of J. In the experiments with summer case the perturbations are even smaller than for the winter case (see Table 1). In all experiments, the perturbations are at least one or two orders of magnitude smaller then the magnitudes of the basic state variables, and still belong to a small perturbation regime.

The correlation coefficients calculated for T, q and meridional wind (v) are shown in Figs. 1, 2 and 3, respectively, for the 4 experiments described above. The correlation coefficients for the zonal wind component (u) are quite similar to these obtained for meridional wind and are not shown. The correlation coefficients calculated for temperature (Fig. 1) and humidity (Fig. 2), show a discontinuity problem in the CONTROL model in the winter case (exp. no. 1 and no. 2). Higher correlations are obtained for larger initial perturbations, which is in disagreement with the expected results of the Taylor series expansion for small perturbations and differentiable functions. In the summer case, the same problem can be seen in Fig. 3 for meridional wind component, obtained again in the CONTROL model experiments (nos. 3 and 4). This means that some serious discon-

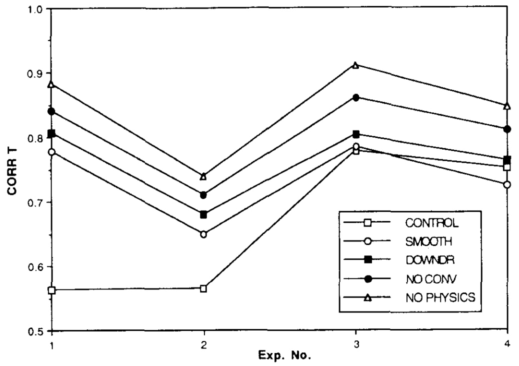

Fig. 1. The correlation coefficients between the nonlinear forecast difference and the tangent linear forecast, calculated for temperature T after 12-h forecast of tangent linear and a pair of perturbed and unperturbed nonlinear forecast models in 4 different experiments, for 5 different models (Table 1). The nonlinear and the corresponding tangent linear models include the same processes.

tinuities present in the CONTROL model are not present (or are reduced to a similar degree) in all other models. The modification done in the DOWNDRAFT and the SMOOTH schemes help to reduce the problem.

As it can be seen in the figures correlation coefficients for temperature and wind are 75–90% for the model NO PHYSICS. These values are comparable to the results obtained by Errico and Vukićević (1992) with a dry adiabatic mesoscale grid-point model. The inclusion of moist processes (large scale precipitation and cumulus convection) obviously adds nonlinearities and discontinuities

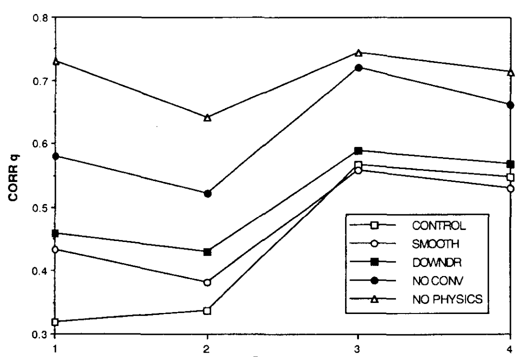

Fig. 2. As in Fig. 1, but for specific humidity q.

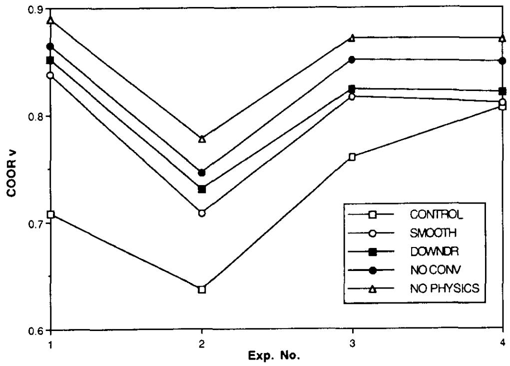

Fig. 3. As in Fig. 1, but for meridional wind v.

and reduces correlations. For the CONTROL model humidity correlations are as low as about 30% for the winter case. The basic contribution to the reduction of correlations comes from locally very large errors (e.g., for the CONTROL model maximum linearization errors for humidity are 3–4 orders of magnitude larger then the perturbations) in the bottom and top layers, L_B and L_T, where a discontinuous transition from convective to non-convective regime exists. The correlations are generally smaller in the winter than in the summer case. As can be seen comparing exp. nos. 1 and 4, this is not entirely due to the larger perturbations in the winter case. The differences are much larger for the CONTROL model, where considerable effects of discontinuities are found in the winter case. One would expect these large local errors and small correlations to have a negative impact in 4D VAR data assimilation. The effects of these problems are examined in the next subsection.

4.2. Assimilation experiments

The inclusion of more complex physics into the forecast models often increases the nonlinearity and the number of discontinuities. This can result in slower convergence rate or even failure of the minimization algorithm. These problems are examined in the 4D VAR data assimilation experiments performed using the same models as in Section 4.1. The following functional is minimized:

$$J(x) = J_{\mathrm{obs}} + J_{\mathrm{p}}, \tag{13}$$

where J_{obs} and J_{p} are the observational and penalty part, defined by:

$$J_{\mathrm{obs}} = \frac{1}{2} \sum_{n=0}^{N_{\max}} \langle (\boldsymbol{x}_n - \boldsymbol{x}_n^{\mathrm{obs}}), \boldsymbol{W}(\boldsymbol{x}_n - \boldsymbol{x}_n^{\mathrm{obs}}) \rangle, \qquad (14)$$

$$J_{\mathrm{p}} = \frac{1}{2} \sum_{n=0}^{N_{\max}} \left\langle \frac{\Delta\,\mathrm{div}_n}{\Delta t_n}, \varepsilon \frac{\Delta\,\mathrm{div}_n}{\Delta t_n} \right\rangle. \qquad (15)$$

In J_{p}, the time derivative of the divergence $(\Delta\,\mathrm{div}/\Delta t)$ is minimized to reduce the amount of spurious gravity waves. The weighting factors, defined by the diagonal matrices $\boldsymbol{W}$ in J_{obs} and ε in J_{p} are chosen empirically. The particular values chosen are: $\boldsymbol{W}_{\mathrm{p}} = 5 \times 10^{-4}\,\mathrm{Pa}^{-2}\,\boldsymbol{I}$ for pressure, $\boldsymbol{W}_{\mathrm{T}} = 4.0\,\mathrm{K}^{-2}\,\boldsymbol{I}$ for temperature, and $\boldsymbol{W}_{\mathrm{V}} = 0.8\,\mathrm{m}^{-2}\,\mathrm{s}^2\,\boldsymbol{I}$ for wind, where $\boldsymbol{I}$ is the identity matrix. The weighting matrix for humidity, is chosen to vary from $\boldsymbol{W}_{\mathrm{q}}(\eta) \approx 10^7\,\boldsymbol{I}$ in the lowermost layer, to $\boldsymbol{W}_{\mathrm{q}}(\eta) \approx 10^{12}\,\boldsymbol{I}$ in the uppermost layer. The weight for the penalty term is $\varepsilon = 10\,\mathrm{s}^4\,\boldsymbol{I}$. With this choice of the weights, J_{p} is about 3 orders of magnitude smaller than J_{obs}. The matrix $\boldsymbol{W}$ used in this study provide reasonably well behavior of the minimization and approximately reflect analysis errors. More appropriate choice of the weights needs further attention and will be done in our future studies. The same values of weighting factors (except for humidity) were used in M. Županski (1993a) and farther details can be found in this paper.

The minimization is performed for the winter case, starting at 1200 UTC 27 December 1991. The analyzed data (p_s, T, u, v, q), available every 3 h, over a 12-h assimilation period are assimilated. A quasi-Newton memoryless minimization algorithm (Shanno, 1978) is applied. The calculation of the grad J is consistent with the particular forecast model chosen for the experiment. The control variable is the vector of initial conditions for the numerical forecast model, $\boldsymbol{x}_0$, defined at the beginning of the assimilation period. The values of the vector $\boldsymbol{x}_0$ are not allowed to change at the boundaries during the minimization. The boundary conditions for the forecast model are variable, taken from the analyses available in the data assimilation period. The maximum number of iterations of the minimization is 10 for all experiments, except DOWNDRAFT, where 20 iterations are performed.

The functional J, normalized by the initial value J_0, is plotted in Fig. 4 as a function of the iteration number, for the five different models. The values of J_0 are 0.87522×10^7, 0.86434×10^7 and 0.85507×10^7 for CONTROL, SMOOTH and DOWNDRAFT model, respectively. For the NO CONVECTION model is $J_0 = 0.83916 \times 10^7$ and for NO PHYSICS $J_0 = 0.87480 \times 10^7$. The initial values of the functional for the models with convection are larger then for the NO CONVECTION model. This is because the quasi-equilibrium profiles of the Betts–Miller scheme are defined assuming that radiation, turbulence and surface processes interact with the convection. Exclusion of these processes, for the reasons explained in Section 3, had an undesirable effect on the forecast and in order to get more realistic 4D VAR data assimilation experiments a full physics forecast model should be used. One would expect that the inclusion of these processes might alter the results of the experiments not only through improvement of the forecast, but through reinforcement of the effects of discontinuities in the convective scheme as well. The experiments performed show that, even in this simpler case, there are problems due to discontinuities. As can be seen in Fig. 4, the best convergence is obtained for the simplest models: NO PHYSICS and NO CONVECTION, the worst for the CONTROL model. The differences are becoming smaller with increasing iteration number. A possible explanation for larger differences in the first few iterations is as

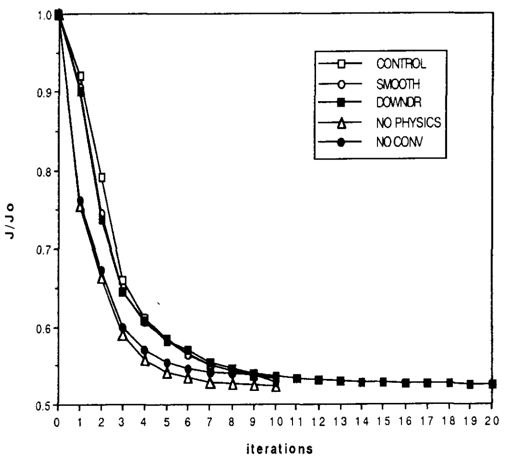

Fig. 4. Functional J, normalized by the initial value J_0, plotted as function of iteration number, for the 5 models (Table 1).

follows. The more nonlinear the forecast model is, the more different from quadratic is the functional J (e.g., Tarantola, 1987) which slows down the convergence of the minimization. Because of discontinuities present in the cumulus parameterization scheme (especially in the CONTROL scheme) the calculated grad J is erroneous, which also slows down the convergence. In the case of very strong discontinuities the functional might even diverge (e.g., Verlinde 1992). Because of local character of convection in our experiments, the minimization process has the ability to recover correct gradients from one iteration to another and to proceed without major difficulties. Thus, even the gradient is erroneous to some degree, as long as it is in descent direction, the minimization procedure converges.

The maximum reduction of J in Fig. 4 is about 50%. The amount of reduction crucially depends on the initial value J_0, i.e., on the distance from the minimum at the beginning. We used the analysis at the beginning of the assimilation period to define the first guess value of the control variable $\boldsymbol{x}_0$, which appeared to be relatively close to the optimal value. Using model independent data (OI analyses) rather then model generated data contributes to lessen the reduction of the functional. Boundary condition, model and analysis errors also have negative effects on reduction of J. One might expect greater reduction of J by allowing the adjustment of boundary conditions during the minimization, by assuming that the model is not perfect (e.g., Derber, 1989; M. Županski, 1993a), or directly using observations instead of analyses. As in the mentioned papers, the forecast experiments should be preformed in order to see if there are benefits of such 4D VAR data assimilation.

Small differences between different experiments after 10 iterations of minimization (Fig. 4) should not be understood that the effects of discontinuities are only in slowing down the convergence. Larger differences between different experiments can be seen by examining particular parts of J. The humidity part, J_q (the misfit between q and q^{obs}), calculated after 10 iterations of minimization, for all models with physical processes, is shown in Fig. 5 as a function of vertical layers. The values of J are normalized by the values obtained in the NO CONVECTION experiment. As can be seen in Fig. 5, all three convective schemes have similar positive effects in reducing J ($J < 1$ means an improvement as compared to the NO CONVECTION case) in upper layers. In low layers however,

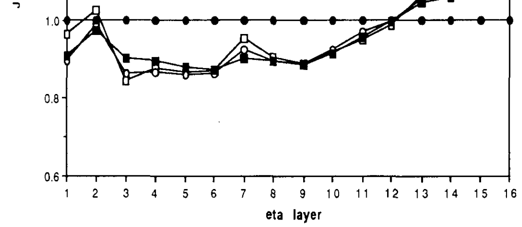

Fig. 5. The humidity part of the functional (J_q), normalized by the value (J_{nc}) calculated for the NO CONVECTION model, after 10 iterations of minimization, plotted for four models with physics as a function of vertical η layers. The layer numbers are increasing from the model top to the bottom.

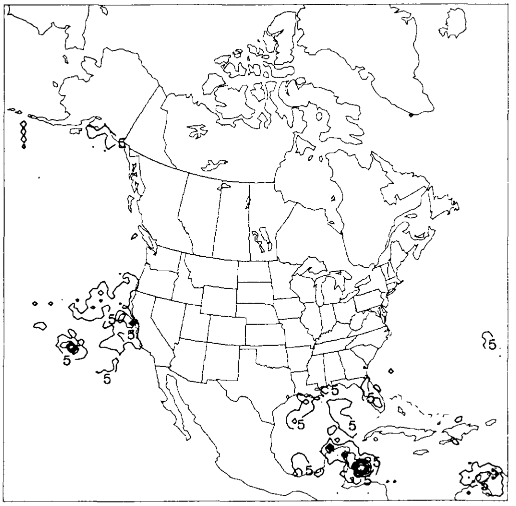

Fig. 6. Absolute value of G_q in lowermost η layer, calculated in 8th iteration of the minimization using the CONTROL model. The gradient is normalized by the weight W_q. The minimum contour is 5 g/kg, the contour interval is 20 g/kg and the maximum value of $|G_q|$ is 157.7 g/kg.

the inclusion of convection has a negative effect ($J > 1$). In Fig. 5, the functional for the CONTROL model is up to 50% larger than for the NO CONVECTION model in the lower layers. These negative effects in low layers are reduced for the smooth and downdraft schemes, but the results are still worse than in the NO CONVECTION case.

The differences in Fig. 5 are not only due to different models but due to their different abilities to handle discontinuities as well. In order to demonstrate that discontinuities are a problem in the assimilation with the CONTROL model, we examined the gradients of J with respect to humidity (G_q) in the lowermost η layer, where the major differences between different experiments are found. It is especially important to see the differences between the gradients calculated in last iterations, when the convergence of J is almost reached. The perturbations Δq to the control variable q are small in these iterations (Δq is two or three orders of magnitude smaller then q in most of the domain) and therefore, well within linear regime. The absolute values of G_q for the CONTROL experiment in 8th, 9th and 10th iterations of the minimization are plotted in Figs. 6, 7 and 8, respectively. As can be seen comparing the figures, very rapid changes in the gradients could be seen south of Alaska, west of California and in the Gulf of Mexico from one iteration to another, despite only small changes in the initial conditions. For example, $|G_q|$ changes from 5 g/kg to 577 g/kg in the region south of Alaska in a single iteration (Figs. 6 and 7). These large changes in the gradient indicate nonsmoothness of J, a consequence of the discontinuities in the control cumulus convection scheme. This problem of the CONTROL model

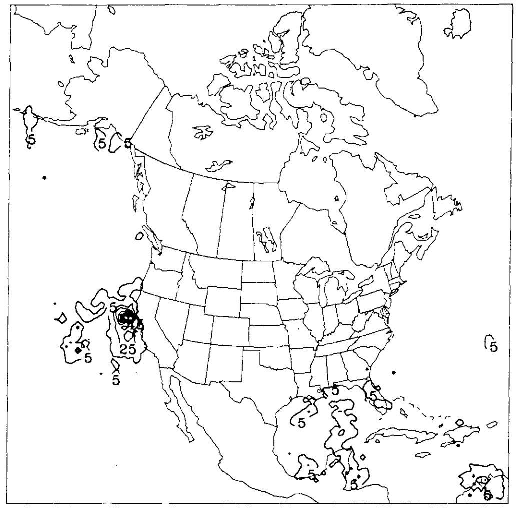

Fig. 8. As in Fig. 6, but for 10th iteration. The maximum value of $|G_q|$ is 202.6 g/kg.

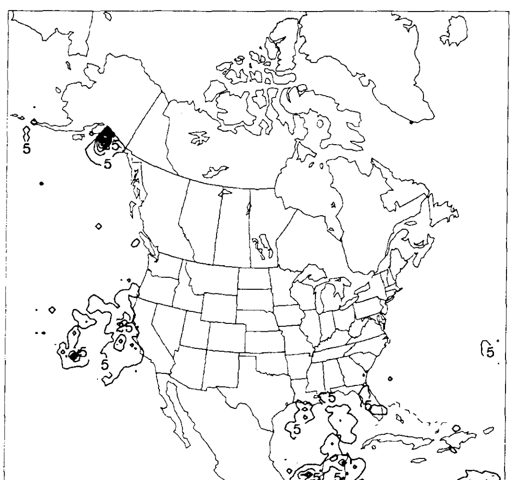

Fig. 7. As in Fig. 6, but for 9th iteration. The maximum value of $|G_q|$ is 577.7 g/kg.

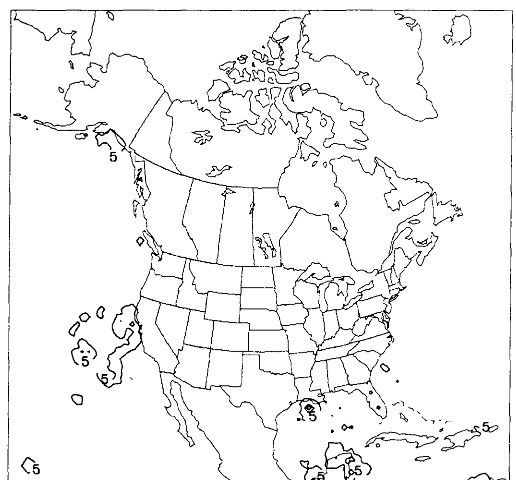

Fig. 9. As in Fig. 6, but for the DOWNDRAFT model. The maximum value of $|G_q|$ is 62.5 g/kg.

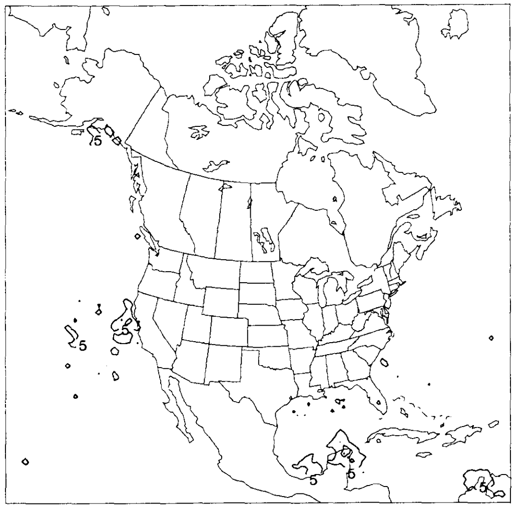

Fig. 10. As in Fig. 7, but for the DOWNDRAFT model. The maximum value of $|G_q|$ is 46.6 g/kg.

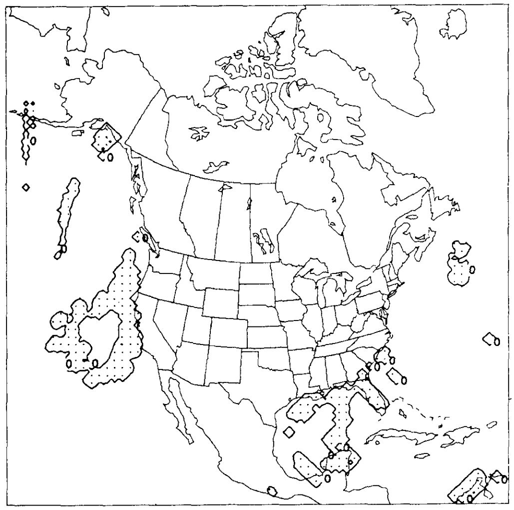

Fig. 12. The 12-h accumulated convective precipitation (in mm) calculated in 10th iteration of the minimization, obtained by DOWNDRAFT model. The minimum contour is 0.1 mm, contour interval is 10 mm and the maximum precipitation amount is 14.9 mm. Dotted areas indicate precipitation amount greater then 0.1 mm.

is serious because it cannot be eliminated by reducing the magnitude of the perturbation. In some experiments that we performed, random perturbations of the order of computer round off errors (10^{-13}) caused the changes of gradients in the CONTROL model case as large as in the figures presented. The most serious effects of the discontinuities of the CONTROL model are in convectively active regions. The convectively active regions can be seen in Fig. 12, which shows 12-h accumulated convective precipitation field, calculated in 10th iteration of the minimization using DOWNDRAFT model. The convective precipitation fields for other two models (CONTROL and SMOOTH) are very similar and are not shown.

Similar large changes caused by small perturbations imposed to the initial conditions are not noted in the DOWNDRAFT case (Figs. 9, 10 and 11 show $|G_q|$ calculated in 8th, 9th and 10th iteration of the minimization, respectively) and in all other cases (not shown). Even though the discontinuities are not completely eliminated in the DOWNDRAFT and the SMOOTH models, their effects are much smaller than in the CONTROL model in the experiments preformed.

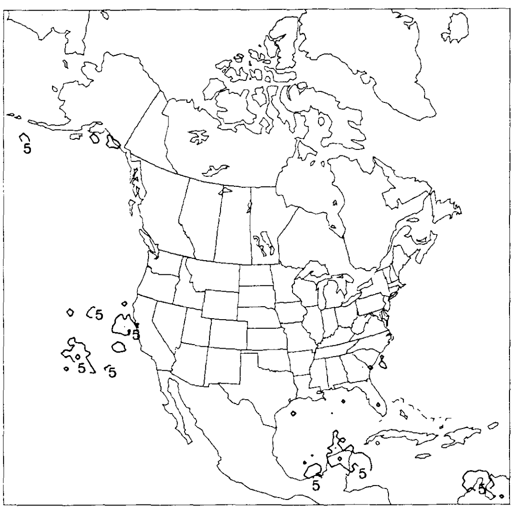

Fig. 11. As in Fig. 8, but for the DOWNDRAFT model. The maximum value of $|G_q|$ is 37.7 g/kg.

5. Conclusions

The following questions are considered in this study. How do discontinuities present in state-of-the-art forecast models with physics affect the 4D VAR data assimilation? It is possible to use derivative minimization methods? Is it worth making discontinuous processes more continuous, and if it is so, how should it be done?

The effects of discontinuities in the Betts–Miller cumulus convection scheme are examined. The operational OI analyses and the quasi-operational version of the NMC/ETA regional forecast model (excluding some physical processes as explained in the text) are used in our experiments. Results of the experiments showed that discontinuities in cumulus convection parameterization increase linearization errors to a large extent and have adverse effects on 4D VAR data assimilation. The gradients calculated by the adjoint model which includes convective processes are locally incorrect due to discontinuities in the forecast model. In our experiments these problems are most serious in control cumulus convection scheme in low layers. Very rapid changes from one iteration of the minimization to another are found in the gradients in this case, even with the perturbation to the control variable of the order of computer round off errors. These problems can be reduced by modifying the cumulus parameterization scheme to make it more continuous in low layers as in the downdraft or smooth scheme.

Positive effects resulting from the inclusion of cumulus convection in 4D VAR data assimilation are found in upper layers, especially in the humidity fields. Taking the convective processes into account in the adjoint model does have an impact on the solution and making the most discontinuous transitions more continuous is beneficial. We expect more benefits by inclusion of other data, more closely related to convection, such as precipitation, clouds, surface parameters etc. Even though this may be more strongly affected by the lack of continuity, the inclusion of radiation, surface processes, turbulence and other important physical processes into the forecast and the corresponding adjoint models should result in additional benefits.

The experiments performed showed that the minimization process converges reasonably well, even with discontinuity problems. Thus, derivative minimization methods appear to be applicable, at least for similar cases. In order to get more general conclusions the assimilation/forecast experiments with full physics forecast and corresponding adjoint models in different synoptic situations using real observations are needed. The forecast results however, may be affected by additional problems that need attention. It is especially important to define the functional J to measure the distance from the truth, as close as possible, and to account for model errors (or at least model bias as in Derber, 1989 and M. Županski, 1993a). These improvements will be examined in future work.

6. Acknowledgments

The author is grateful to Milija Županski and John C. Derber for countless hours of stimulating discussions and valuable suggestions during this research as well as for their careful reading of the article and helpful extensive comments. The idea of making the cumulus convection scheme more continuous was born after discussing with John C. Derber, Fedor Mesinger and Eugenia Kalnay. Special thanks are extended to the many people from NMC that have contributed to this work. Valuable comments on the convection scheme by Alan K. Betts and Zaviša I. Janjić are very appreciated. The program code for the control cumulus convection scheme is provided by Zaviša I. Janjić and the code for the downdraft scheme by Martin J. Miller. The tangent linear and adjoint models, adiabatic versions, as well as the minimization programs are provided by Milija Županski. This work was supported by the UCAR/NMC Visiting Research Program.

REFERENCES

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.

Betts, A. K. 1982. Saturation point analysis of moist convective overturning. *J. Atmos. Sci. 39*, 1484–1505.

Betts, A. K. 1983. Thermodynamics of mixed stratocumulus layers: saturation points budgets. *J. Atmos. Sci. 40*, 2655–2670.

Betts, A. K. 1986. A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. R. Met. Soc. 112*, 677–691.

Betts, A. K. and Miller, M. J. 1986. A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Quart. J. R. Met. Soc. 112*, 693–709.

Betts, A. K. and Miller, M. J. 1993. The Betts-Miller Scheme. In: *The Representation of cumulus convection in numerical models of the atmosphere* (eds. Emanuel, K. A. and Raymond, D. J.), American Meteorological Society (in print).

Courtier, P. and Talagrand, O. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equations. II: Numerical results. *Quart. J. R. Met. Soc. 113*, 1331–1350.

Derber, J. C. 1989. A variational continuous assimilation technique. *Mon. Wea. Rev. 117*, 2437–2446.

Di Mego, G. J. 1988. The Nationl Meteorological Center Regional Analysis System. *Mon. Wea. Rev. 116*, 977–1000.

Douady, D. and Talagrand, O. 1990. The impact of threshold processes on variational assimilation. Preprints, WMO Int. Symposium on Assimilation of Observations in Meteorology and Oceanography. Clermont-Ferrand, France, 486–487.

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

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

Hoffman, R. N. 1986. A four-dimensional analysis exactly satisfying equations of motion. *Mon. Wea. Rev. 114*, 388–397.

Janjić, Z. I. 1990. The step-mountain coordinate: Physical package. *Mon. Wea. Rev. 118*, 1429–1443.

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

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

Mesinger, F., Janjić, Z. I., Ničković, S., Gavrilov, D. and Deaven, D. 1988. The step mountain coordinate: model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Mon. Wea. Rev. 116*, 1493–1518.

Navon, I. M., Zou, X., Derber, J. C. 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.

Rockafellar, R. T. 1981. *The theory of subgradients and its applications to problems of optimization: convex and non-convex functions.* Berlin, Heldermann Verlag, 107 pp.

Shanno, D. F. 1978. Conjugate gradient methods with inexact line searches. *Math. Operations Res. 3*, 244–256.

Tarantola, A. 1987. *Inverse problem theory. Methods for data fitting and model parameter estimation.* Amsterdam, Oxford, New York, Tokyo, Elsevier, 613 pp.

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

Thepaut, J.-N., Vasiljević, D., Courtier, P. and Pailleux, J. 1993. Variational assimilation of conventional meteorological observations with a multilevel primitive equation model. *Quart. J. R. Met. Soc. 119*, 153–186.

Verlinde, J. 1992. Fitting microphysical observations to a numerical model through an optimal control theory technique. PhD Thesis, Colorado State University, Fort Collins, Colorado 80523, Paper No. 493, 83 pp.

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

Wergen, W. 1992. The effect of model errors in variational assimilation. *Tellus 44A*, 297–313.

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

Županski, M. 1993a. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev. 121*, 2396–2408.

Županski, M. 1993b. A preconditioning algorithm for large scale minimization problems. *Tellus 45A*, 578–592.