Performance of 4D-Var strategies using the FSU Global Spectral Model with its full physics adjoint

Zhijin Li

Supercomputer Computations Research Institute

Florida State University, Tallahassee, Florida

I. M. Navon

Department of Mathematics and

Supercomputer Computations Research Institute

Florida State University, Tallahassee, Florida

To be submitted to Mon. Wea. Rew.

November, 1998

Corresponding author:

I. Michael Navon

Supercomputer Computations Research Institute

Florida State University

Tallahassee, FL 32306-4052

Email: navon@scri.fsu.edu

Phone: (904) 644-6560

Fax: (904) 644-0098

Abstract

A set of four-dimensional variational assimilation (4D-Var) experiments were conducted using both a standard method and an incremental method. The full physics adjoint model of the FSU Global Spectral model (FSUGSM) was used in the standard 4D-Var, while the adjoint of only selected physical parameterizations was used in the incremental method. We examined in detail the impact of physical processes on 4D-Var by comparing the results of above experiments. As a whole, the inclusion of full physics in the adjoint model was detrimental to the rate of convergence of the minimization process prior to iteration 50, while full convergence was attained only after a large number of iterations. The detrimental impact was found to be primarily related to precipitation physics in the adjoint model, and was limited over precipitation regions. The inclusion of full physics turned out to be significantly beneficial in terms of assimilation error to the lower troposphere during the entire minimization process. The beneficial impact was found to be primarily related to boundary layer physics. The precipitation physics in the adjoint model also tended to have a beneficial impact starting from iteration 50. Experiment results confirmed that assimilation analyses with the full physics adjoint model display a shorter precipitation spin-up period. The beneficial impact on precipitation spin-up did not result solely from the inclusion of the precipitation physics in the adjoint model, but rather from the combined impact of several physical processes.

A truncated Newton-like incremental approach was introduced for examining the possibility of circumventing the detrimental aspects using the full physics in the adjoint model in 4D-Var but taking into account its positive aspects. This algorithm is based on the idea of the truncated Newton minimization method and the sequential cost function incremental method introduced by Courtier et al. (1994), consisting of an inner loop and an outer loop. The inner loop comprised the incremental method, while the outer loop consisted of the standard 4D-Var method using the full physics adjoint. The limited-memory quasi-Newton minimization method (L-BFGS) was used for both inner and outer loops, while information on the Hessian of the cost function was jointly updated at every iteration in both loops. In an experiment with a two-cycle truncated Newton-like incremental approach, the quality of the assimilation analyses turned out to be better than that obtained from either the standard 4D-Var or the incremental 4D-Var in all aspects examined. The CPU time required by this two cycle approach was larger by 35% compared with that required by the incremental 4D-Var without physics in the adjoint model, while the CPU time required by the standard 4D-Var with the full physics adjoint model was more than twice that required by the incremental 4D-Var. Several hypotheses concerning the impact of using standard 4D-Var full physics on minimization convergence were discussed.

1 Introduction

Four dimensional variational data assimilation (4D-Var) is used for generating estimates of dynamical variables, taking into account both information about the dynamics from a dynamical model, and available information about the true state of the atmosphere contained in the observations and the background. Usually, a cost-function is defined as measuring the distance between a model trajectory and the observations over an assimilation interval or window. 4D-Var is implemented mathematically by minimizing the defined cost function with respect to the control variables. The minimization algorithm used in 4D-Var employs the adjoint equations for the computation of the gradient of the cost function with respect to control variables.

Thus, a successful 4D-Var system hinges upon three main aspects: the use of observations and background information, the accuracy of the model used, and the efficiency of the minimization algorithm. Theoretically, improvement in any aspect may improve the 4D-Var system. A sizable progress has been accomplished in every one of the above aspects during the last decade. However, due to the enormous size of meteorological problems, such improvement is greatly restricted by presently available computer platforms.

In the present study, we focus on how to use a full physics adjoint model in 4D-Var. The inclusion of the totality of the physical processes in the adjoint model is a required step towards obtaining a better estimation of the gradient of the cost-function. Thus this research aims to address improvements in the second aforementioned aspect, though we will not address here the issue of model errors.

4D-Var was first applied to simple models (Le Dimet and Talagrand 1986; Lewis and Derber 1985; Courtier and Talagrand 1987; Talagrand and Courtier 1987), before being tested in the context of adiabatic primitive models (Thepaut and Courtier 1991; Navon et al. 1992; Chao and Chang 1992; Zupanski 1993). Recently, increasingly sophisticated physical parameterizations have been introduced into adjoint systems (e.g., Zou et al. 1993; Zou and Kuo 1996; Zou 1997; Zhu and Navon 1998; Tsuyuki 1997; Zupanski and Messinger 1995; Li and Navon 1997; Mahfouf et al. 1998; Fillion and Errico 1998). These studies have provided encouraging evidence that the inclusion of physics in the adjoint may improve the performance of 4D-Var.

In the context of an operational environment, Rabier et al. (1997, 1998) used the ECMWF 4D-Var incremental assimilation system to examine the influence of physical processes. The inclusion of physics was found to have the largest impact on the analysis of humidity. A positive impact was also found on the performance of analyses in the tropics, with a reduction of the spin-down of precipitation in the subsequent forecast, and improved wind scores. Also, a slight improvement in extratropical scores was noted. ECMWF is planning to implement 4D-Var with physics in the near future (Rabier et al. 1997). It should be pointed out that the physics in the ECMWF's adjoint model consists of a simplified version of the physics package used in the nonlinear forecasting model (Mahfouf et al. 1998, Mahfouf 1998).

While recent research has shown a beneficial impact on the quality of assimilated data, the effect of nonlinearities in physical processes on the convergence rate of 4D-Var minimization is still not adequately understood. Theoretically, an adjoint model with full physics should be completely consistent with the nonlinear forecasting model, thus providing exact gradients. However, physical parameterizations display much stronger nonlinearities than the model dynamical part, even when they do not contain discontinuities or on-off switches. While the presence of nonlinearities in the dynamic part does not cause severe problems, those in physical parameterizations may cause a lack of validity of linearization approximation even for very limited sizes of perturbations. Thus, the accurate gradient may still not provide an effective descent direction for a minimization process. Further, discontinuities may directly lead to errors in the gradients, unless generalized gradients are used (Xu, 1996a,b; 1997).

The 4D-Var experiments in the present research will show that, for a large number of minimization iterations during the starting stages of the minimization process, relatively large errors in the assimilation analysis are found to be closely related to precipitation when full physics is included in the adjoint model, and these errors are larger than those present in assimilation analyses where no physics is included in the adjoint model. On the other hand, the absence of physics causes large errors in assimilation analyses in the lower troposphere which cannot be reduced by performing additional minimization iterations. The inclusion of full physics into an adjoint model on 4D-Var has thus a mixed effect.

The inclusion of full physics in the adjoint model requires a 4D-Var algorithm to overcome the negative effect of strong nonlinearities present in physics while being able to take advantage of the positive aspects resulting from consistency between the forecasting nonlinear model and adjoint model.

Several approaches have been proposed for mitigating the negative effect of strong nonlinearities in physical processes included in the adjoint model by either direct modifications or simplifications to physical parameterizations themselves. Zupanski and Mesinger (1993) and Tsuyuki (1997) showed a beneficial effect when smoothing formulas are used to replace those with discontinuities. ECMWF uses simplified physics in the adjoint model (Mahfouf et al. 1998; Rabier et al. 1997). A concern is that these modifications or simplifications may lead to a degree of inconsistency between the nonlinear forecasting model and the adjoint model.

An alternative approach is to deal with the nonlinear problem of physical processes in the adjoint model in the framework of the minimization procedure. In 4D-Var, physical processes are indeed of secondary importance, compared with the dynamical processes for short-span assimilation windows (less than 24 hours). This fact was validated by the success obtained by 4D-Var at ECMWF where only simple horizontal and vertical diffusions are the sole physical processes included in the adjoint model used (Rabier et al. 1997, 1998). Further, physical processes are controlled by dynamical processes to a large degree for large scale dynamics. These features allow us to deal with physical processes in a fashion differing from that used with dynamical processes. This basic view constitutes the basis for this research. As an immediate consequence of this tenet, we naturally resort to the idea of a progressive inclusion of physical processes in 4D-Var as the minimization process proceeds, an idea first proposed by Courtier et al. (1994).

The progressive inclusion of physical processes can be implemented in 4D-Var using the incremental approach devised by Courtier et al. (1994). In this approach, the minimization is performed with respect to increments. An increment is defined as a deviation from a background (guess) trajectory. The background trajectory is computed with the full nonlinear forecasting model which has a high resolution and comprehensive physics. A simplified linear model with low resolution and/or simplified physics is used for solving the minimization problem in the vicinity of the trajectory. This method is further extended by defining a sequence of cost functions for different stages of the minimization process. Then the approach allows a progressive inclusion of physical processes, using simplified physics adjoint models during the initial stages and full physics adjoint models towards the final stages of the minimization process. Due to the fact that effects of nonlinearities are weaker for perturbations of smaller sizes, this sequential cost function incremental method offers an alternative approach for alleviating effects of strong nonlinearities present in physics. This assertion will be validated using 4D-Var experiments.

Courtier et al. (1994) mentioned that there is no guarantee for convergence of the incremental approach using a sequence of cost functions. Also, the convergence rate may be negatively affected by using different cost functions in the course of a minimization iteration procedure. Minimization algorithms used in 4D-Var usually belong to limited-memory quasi-Newton methods. The basic motivation behind limited memory quasi-Newton methods is to try obtaining the rapid convergence associated with Newton's method without explicitly evaluating the Hessian of a cost function at every step. This is accomplished by constructing approximations to the inverse Hessian based on information gathered during the descent process. In 4D-Var, the minimization is terminated after a number of iterations much smaller than the dimension of the problem. Using a sequence of cost functions may cause additional difficulties for obtaining a good approximation to the inverse Hessian.

Truncated Newton methods have been extensively investigated and applied (e.g., Dembo and Steihaug 1983; Nash and Nocedal 1991; Schlick and Fogelson 1992; Navon et al. 1992a; Zou et al. 1993; Wang et al. 1992, 1995). The standard truncated Newton method consists of nested iterations: an outer iteration and inner iteration. The theoretical framework of truncated Newton methods presents a suitable tool for examining performance of the incremental method using a sequence of cost functions. In fact, the incremental method using a sequence of cost functions can be viewed as an algorithmic variant of the truncated Newton method. As such, we expect to refine the sequential cost function algorithm and better analyze its convergence properties by using the theoretical framework of truncated Newton methods.

The aforementioned effort requires a specific and detailed understanding of the impact of physical processes in the adjoint model on 4D-Var. Here we use an adjoint model with full physics based on the Florida State University Global Spectral model (FSUGSM). This model has been successfully applied to carry out both 4D-Var and optimal parameter estimation by (Zhu and Navon 1998a, b) and Tsuyuki (1997,1996). Having a full physics adjoint model at our disposal, we can use it as a benchmark with which to compare other adjoint models with partial physical processes. Carrying out such comparisons enables us to examine how physical processes in the adjoint model impact on 4D-Var as the minimization process proceeds.

The outline of this paper is as follows. In section 2, we briefly summarize the essential features of the FSUGSM model and its full physics adjoint model. The results of the standard 4D-Var approach using the full physics adjoint model are given in Section 3. Section 4 deals with the incremental method that involves no physics in the adjoint. Section 5 details comparisons between results obtained via the incremental 4D-Var and those obtained using the standard 4D-Var, followed by discussions on the effect of physical processes on 4D-Var. Section 6 examines the incremental method with a sequence of cost functions, and a new truncated Newton-like incremental method is presented and tested. Section 7 discusses the results obtained in this study.

2 A brief description of FSUGSM and its adjoint

FSUGSM has been used in numerical weather forecasts for operational purposes for more than a decade. Forecasts using this model especially emphasize tropical aspects such as monsoon and tropical storms (e.g., Krishnamuti et al. 1991).

The model has a comprehensive advanced physical parameterization package. The main physical parameterizations include a fourth order horizontal diffusion (Kanamitsu et al. 1983), a modified Kuo-type convective scheme (Krishnamuti et al. 1983), dry convective adjustment, large-scale condensation (Kanamitsu et al. 1975), surface flux via similarity theory (Businger et al. 1971), vertical distribution of fluxes utilizing diffusive formulation where the exchange coefficients are functions of the Richardson number (Louis 1979), long-wave and shortwave radiative fluxes based on a band mode (Harshvardan and Corsetti 1984; Lacis and Hansen 1974), computation of low, middle and high clouds based on threshold relative humidity for radiative calculation, and surface energy balance coupled to the similarity theory (Krishnamuti 1991).

The adjoint system of FSUGSM has been developed as a result of several years efforts. The adjoint system includes the dynamic core (Wang 1993) and all abovementioned physical parameterizations (Tsuyuki 1996; Zhu and Navon 1997). The adjoint codes have been derived in consistence with the original nonlinear codes. In order to improve the performance of the corresponding linearized model, based on which the adjoint model was derived, a number of smoothing techniques were introduced to remove discontinuities in some physical parameterization formulations prior to derivation of the adjoint model (Tsuyuki 1996, 1997; Zhu and Navon 1997). This adjoint system has been successfully applied to 4D-Var experiments (Tsuyuki 1995, 1996), sensitivity analysis (Zhu and Navon 1998b) and parameter optimal estimation (Zhu and Navon 1998a).

The dependent variables of the forecasting model as well as the adjoint model include vorticity, divergence, the logarithm of surface pressure, temperature and dewpoint depression which is defined as the difference between temperature and dewpoint temperature. In the following variational data assimilation experiments, the spectral expansion of the model variables is triangularly truncated at the wave number 42 (T42). A sigma (σ) coordinate is used in the vertical and the vertical resolution consists of 12 layers roughly between 100 and 1000 hPa.

3 Standard 4D-Var using full physics adjoint model

3.1 A formulation description

The standard 4D-Var minimizes the cost (objective) function J that measures the weighted sum of squares of distances to the background state x^b and to the observations y^o distributed over a time interval $[t_0, t_n]$ (Ide et al. 1997; Lorenc 1986),

$$J[\mathbf{x}(\mathbf{t_0})] = \frac{\mathbf{1}}{\mathbf{2}}[\mathbf{x}(\mathbf{t_0}) - \mathbf{x^b}(\mathbf{t_0})]^{\mathbf{T}}\mathbf{B_0^{-1}}[\mathbf{x}(\mathbf{t_0}) - \mathbf{x^b}(\mathbf{t_0})] + \frac{\mathbf{1}}{\mathbf{2}}\sum_{\mathbf{i=0}}^{\mathbf{n}}(\mathbf{y_i} - \mathbf{y_i^o})^{\mathbf{T}}\mathbf{R_i^{-1}}(\mathbf{y_i} - \mathbf{y_i^o}), \quad (1)$$

where $\mathbf{x}$ is the control variable vector, and $\mathbf{y_i} = \mathbf{H_i}[\mathbf{x}(\mathbf{t_i})]$. $\mathbf{H_i}$ is an observation operator which maps model state variables to the observations. If the state variables are observed directly then $\mathbf{H_i}$ is just an interpolation operator. Here $\mathbf{B}$ is the error-covariance matrix of $\mathbf{x^b}$, and $\mathbf{R}$ the error-covariance matrix of $\mathbf{y^o}$. The minimization of the cost function (1) is carried out with respect to the initial state $\mathbf{x}(\mathbf{t_0})$.

The evaluation of J involves time integrations of a nonlinear forecasting model from t_0 to t_i,

$$\mathbf{x}(\mathbf{t_i}) = \mathbf{M}(\mathbf{t_i}, \mathbf{t_0})\mathbf{x}(\mathbf{t_0}), \quad (2)$$

where M is a nonlinear dynamics operator. Efficient algorithms available for performing minimization of J require availability of its gradient with respect to $\mathbf{x}(\mathbf{t_0})$. Efficient computation of the gradient requires integrating the adjoint model backward in time from t_i to t_0 (e.g., see Le Dimet and Talagrand 1985; Talagrand and Courtier 1987).

Various simplifications to the standard 4D-Var are often required for operational implementation. For currently available computer platforms, the standard 4D-Var imposes heavy requirements for its operational implementation due to the enormous size of meteorological problems. The purpose of the simplifications to the standard 4D-Var is to reduce the heavy requirements in computer resources. The incremental method proposed by Courtier et al. (1994) is one of the most successful in reducing computational cost, and has been operationally used at ECMWF (Rabier et al. 1997). It may also serve another purpose, namely to alleviate the negative effect of nonlinearities, especially those originating in physical parameterizations, as discussed in the following sections. Since time integrations of the adjoint model encompass the major part of the computational cost of 4D-Var, simplifications are generally implemented by using a simplified adjoint model. Theoretically, the standard 4D-Var requires the use of an adjoint model consistent with the forecasting model. In this section, we carry out experiments with the standard 4D-Var using the full physics adjoint model, i.e., the adjoint model is consistent with the forecasting model. This result serves as a benchmark for comparison with results obtained from simplified forms of 4D-Var. One main purpose of this research is to examine the impact of physical processes on minimization processes and on errors in resulting assimilation analyses.

To examine errors in the assimilation analyses, the experiments are designed in an identical twin framework. Thus assimilation errors are just the differences between assimilation analyses and 'observations'. A reference forecast is carried out from $t_0 = 0$ to $t_1 = 6h$ using the full physics forecasting model, and the initial field and the results of the forecast at $t_1 = 6h$ are used as 'observations'. The 4D-Var is carried out to recover the state at time t_0. The initial condition of the reference forecast is the initialized analysis valid at 00UTC 03 September 1996. The guess initial condition (or the background field) is constructed by adding random perturbations with a mean-root-error of $2.5 \times 10^{-5} s^{-1}$ for vorticity, $1.6 \times 10^{-5} s^{-1}$ for divergence, $4.0K$ for dewpoint depression, and $2.1K$ for temperature.

In the identical twin framework, the cost function (1) is simplified to read as

$$J_s[\mathbf{x(t_0)}] = \frac{\mathbf{1}}{\mathbf{2}} \sum_{\mathbf{i=0}}^{\mathbf{1}} (\mathbf{H_i x(t_i)} - \mathbf{y_i^o})^{\mathbf{T}} \mathbf{R_0^{-1}} (\mathbf{H_i x(t_i)} - \mathbf{y_i^o}). \tag{3}$$

Here y_i^o are defined on all model Gaussian grid points, and thus $\mathbf{H_i}$ are the identity matrix. The covariance matrix of observation error is defined as a diagonal matrix, the diagonal elements of which correspond to the variances of random errors used to construct the guess initial condition. In such a case, the exact minimization solution is y_i^o, while $J_s[\mathbf{x(t_0)}] = \mathbf{0}$. The experiments have been performed using the limited memory quasi-Newton minimization algorithm (L-BFGS) described by Liu and Nocedal (1989), which is equivalent to the M1QN3

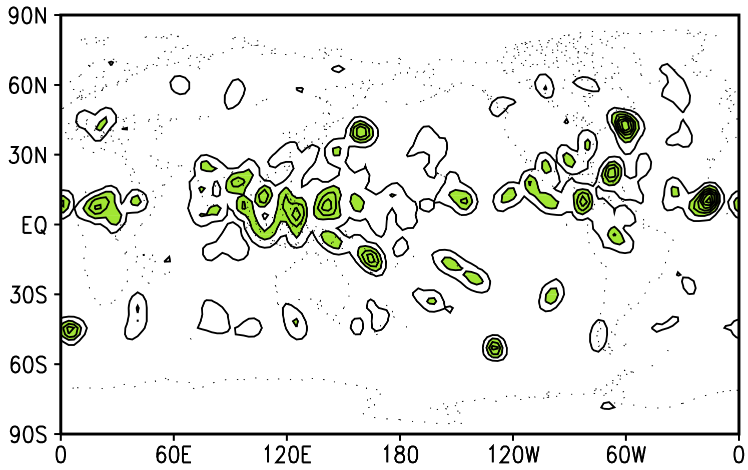

Figure 1: Accumulated precipitation of the reference forecast ('observation') over the last 3-h within the assimilation window. The contour interval is $1.0mm$. Values larger than 1.5 mm are shaded.

used at ECMWF (Gilbert and Lemarechal 1989).

3.2 Results of the standard 4D-Var with full physics adjoint model

Previous results of some experiments have shown that 4D-Var displays different behavior in the tropics and in the extratropics (e.g., Rabier et al. 1997). It is often assumed that this originates in differences in convective activities. We analyze the differences between the tropics and extratropics in several aspects of 4D-Var. Figure 1 presents the 3h accumulated precipitation of the reference forecast ('observations') over the last 3h within the assimilation window. The major precipitation events occur between $30°S$ to $30°N$. Thus in what follows we represent the tropics by the region between $30°S$ to $30°N$, and the extratropics by all the other regions.

Figure 2 displays the evolution of the cost function along with its tropical and extratropical parts versus the number of minimization iterations. The cost functions have been normalized using their corresponding initial values. The minimization process displays a good rate of convergence, even for the tropics. The value of the cost function decreases by one order of magnitude after the first 30 iterations, and decreases to 0.5% of its initial value

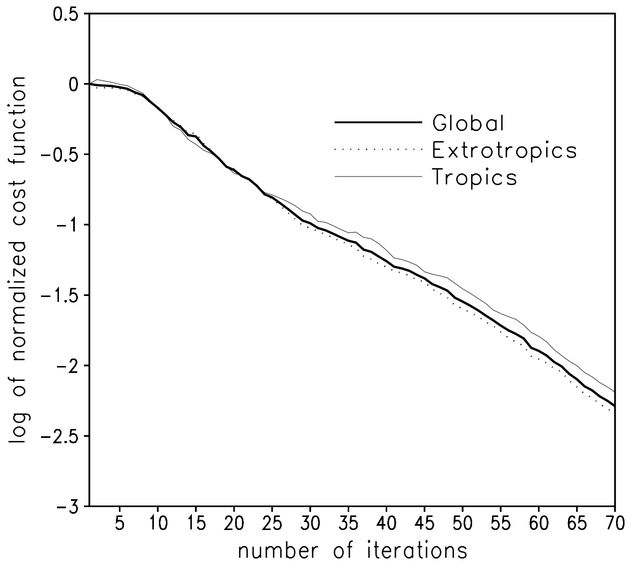

Figure 2: The logarithm (base 10) of the normalized cost function with number of iterations for the standard 4D-Var, in which the full physics adjoint model is used. The cost function is partitioned into the tropical and extratropical parts, which are also displayed. A total of 70 minimization iterations are carried out, while 75 evaluations of the cost function and its gradients are required.

after 70 minimization iterations. Thus, the inclusion of physics into the adjoint model is not detrimental to the final convergence of the minimization process. It is interesting to note that the tropical part of the cost function decreases at a slightly slower rate after 20 iterations.

We also compute contributions of single model variables to the cost function. Figure 3 provides the evolution of the normalized values of various model variables with the minimization iterations. For both the tropics and extratropics, the vorticity and divergence parts exhibit a faster rate of decrease than either the dewpoint depression or temperature parts. In other words, the dynamic field is easier to recover than either the thermal or moisture fields. It is of interest to note that the evolution of single variable contributions exhibits also a difference between the tropics and extratropics. In the tropics, the divergence part and the dewpoint depression part display a slower rate of descent, mainly between iteration 5 and iteration 25 of the minimization process. This difference may be attributed to a stronger negative effect of nonlinearities related to convective activities in the tropics than in the

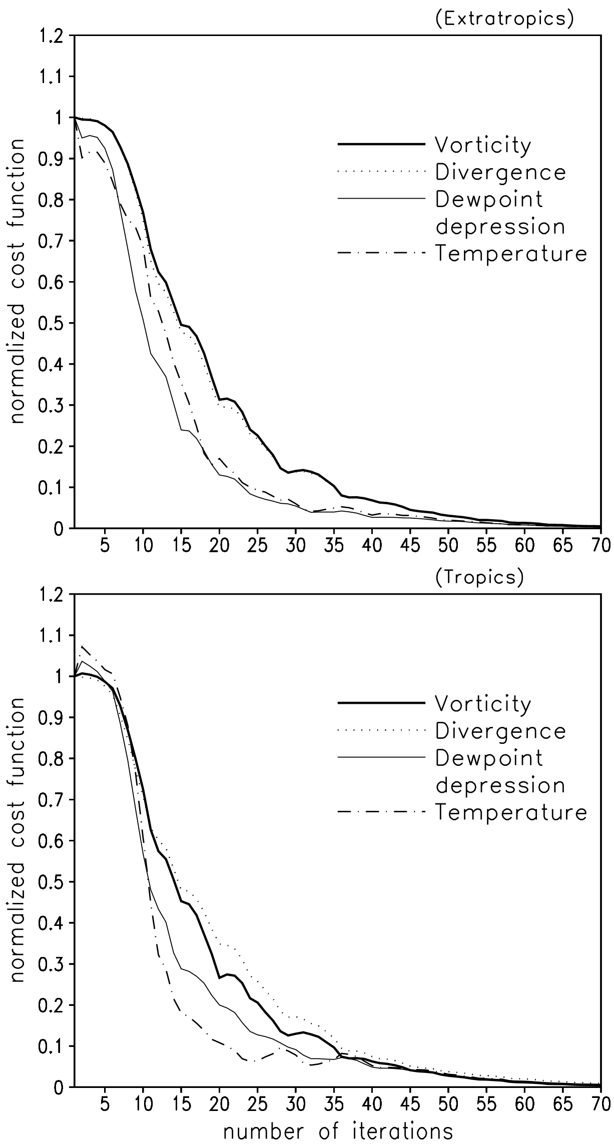

Figure 3: Values of single model variable contributions to both the tropical and extratropical parts of the cost function for the standard 4D-Var. The others are the same as in Fig. 2.

extratropics, since convective activities directly impact divergence and moisture fields.

Although the cost function reaches a rather small value at the end of the minimization process, there still exists a possibility for large localized errors in the assimilation analysis. Figure 4 illustrates the spatial distribution of error of the assimilation analysis of dewpoint depression on representative model σ-levels at iteration 70. The σ level 1 is taken to represent the level near the ground, σ level 3 the lower troposphere, and σ level 9 the upper troposphere. As expected, the error of the assimilation analysis does exhibit a localized character. The maximal error attains a value of up to 1.6 K. The errors in vorticity, divergence and temperature fields exhibit spatial distributions similar to that of dewpoint depression, with maxima of 0.7×10^{-5}, 0.7×10^{-5} and 1.2 K respectively. Thus the minimization does display a much slower rate of convergence in some limited areas.

Intense physical processes such as deep convections, usually occur in limited regions. The influence of physical processes on 4D-Var may also be regional. We carry out a comparison between the error of the assimilation analysis in Fig. 4 and the accumulated precipitation in Fig. 1. The strong precipitation region in Southern Asia and the western Pacific corresponds to large errors in the assimilation analysis. Also, the strong precipitation region in Southern America corresponds to large errors in the assimilation analysis. Furthermore, it is evident that these isolated strong precipitation centers correspond to large error centers. The errors which correlate with the strong precipitation dominate the error field. From the spatial correlations between precipitation and errors of assimilation analysis, we conclude that the moisture physics may lead to large errors in the assimilation analysis over some limited regions where strong precipitation occurs.

The vertical variation of errors of the assimilation analysis does not seem to be significant. The evidence for this is also provided by the similarity of the dewpoint depression error between model levels 1, 3 and 9 in Fig. 4. In order to provide a sounder evidence, we present the vertical distribution of the root-mean-square (RMS) error of the assimilation analysis and 6-h forecasts from the assimilation analysis (Fig. 5). The vertical variation is indeed very small.

4 Incremental 4D-Var

The incremental method proposed by Courtier et al. (1994) allows us to perform 4D-Var using a simplified adjoint model. The standard 4D-Var consists in finding the model trajectory which best fits the available observations over the assimilation window. In the incremental method, an increment is defined by

$$\delta \mathbf{x}(\mathbf{t_0}) = \mathbf{x}(\mathbf{t_0}) - \mathbf{x^b}(\mathbf{t_0}). \tag{4}$$

Here $\mathbf{x^b}(\mathbf{t_0})$ is the background (guess) field at the initial time t_0. Then the incremental 4D-Var is designed to find the analysis increment $\delta \mathbf{x^a}(\mathbf{t_0})$ which minimizes a cost function. This

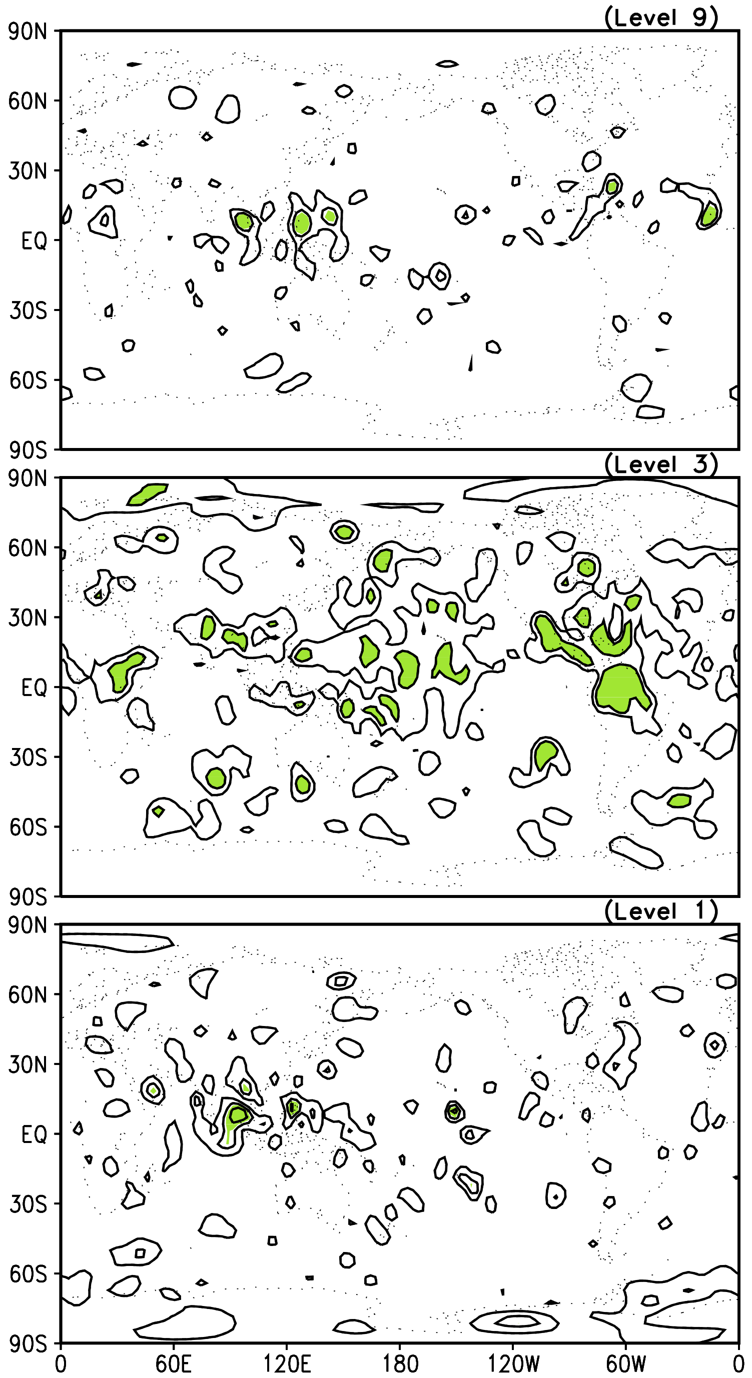

Figure 4: Errors in the assimilation analysis of dewpoint depression for the standard 4D-Var for representative model σ-levels after 70 minimization iterations. Level 1 represents model levels near the ground, level 3 in the lower troposphere and level 9 in the upper troposphere. The contour interval is $0.25K$. Values larger than $0.5K$ are shaded.

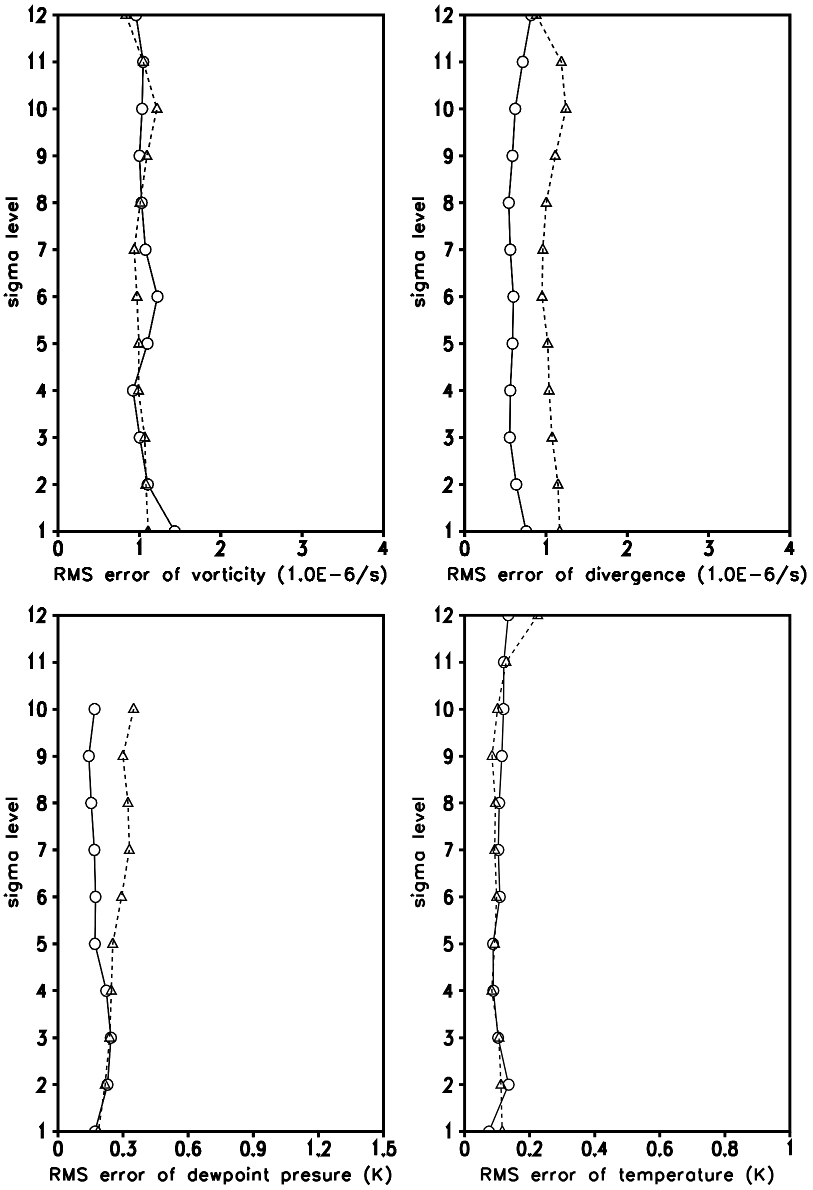

Figure 5: Vertical distributions of RMS errors for the standard 4D-Var analyses at minimization iteration 70 and 6-h pre-forecasts starting from the assimilation analyses. The full physics adjoint model is used. Solid lines are for the assimilation analyses, and dashed lines for the 6-h forecasts.

cost function is defined as the function of $\delta\mathbf{x^a(t_0)}$, instead of $\mathbf{x(t_0)}$ as in the cost function (1) for the standard 4D-Var (Courtier et al. 1994; see also Veerse and Thepaut 1998). The assimilation analysis is obtained by adding the analysis increment $\delta\mathbf{x^a(t_0)}$ to the background

$$\mathbf{x^a(t_0)} = \mathbf{x^b(t_0)} + \delta\mathbf{x^a(t_0)}. \tag{5}$$

The idea central to the incremental method is to use a simplified and linearized version of the forward forecasting model to compute the evolution of the increment $\delta\mathbf{x^a(t_0)}$, while the evolution of the background is predicted with the full forward forecasting model. The impact of the simplification and linearization is thus reduced to a minimal level.

In accordance with the cost function (3) in the standard 4D-Var, we define the incremental cost function as

$$J[\delta\mathbf{x(t_0)}] = \mathbf{\frac{1}{2}\sum_{i=0}^{1}(H_i\delta\tilde{x}(t_i) - d_i)^T R_i^{-1}(H_i\delta\tilde{x}(t_i) - d_i)} \tag{6}$$

where $\delta\tilde{\mathbf{x}}(t_i) = \tilde{\mathbf{x}}(t_i) - \tilde{\mathbf{x}}^b(t_i)$, and the innovation vectors $d_i = \mathbf{y_i^o} - \mathbf{H_i x^b(t_i)}$. $\tilde{\mathbf{x}}(t_i)$ and $\tilde{\mathbf{x}}^b(t_i)$ are governed by the simplified model

$$\tilde{\mathbf{x}}(t_i) = \tilde{M}(t_i, t_0)[\mathbf{x^b(t_0)} + \delta\mathbf{x(t_0)}], \tag{7}$$

$$\tilde{\mathbf{x}}^b(t_i) = \tilde{M}(t_i, t_0)[\mathbf{x^b(t_0)}]. \tag{8}$$

Here $\tilde{\mathbf{M}}(t_i, t_0)$ is the dynamics operator representing the time integration of the simplified forecasting model from time t_0 to time t_i. It is worth emphasizing that $\mathbf{x^b(t_i)}$ is predicted by the full forecasting model starting from $\mathbf{x^b(t_0)}$. In this way, the nonlinearity in the full forecasting model is partially taken into account.

The definition of (6) corresponds to the one in Courtier et al. (1994). As mentioned by Courtier et al. (1994), the practical advantage of this formulation is that there is little technical development required once the standard 4D-Var problem has already been implemented. Another advantage is that (6) becomes identical to the cost function used in the standard 4D-Var when we switch from the simplified model to the full model.

In (6), $\tilde{\mathbf{x}}^b(t_i)$ and $\mathbf{d_i}$ are kept constant during the minimization process. Only $\tilde{\mathbf{x}}(t_i)$ is changed as the minimization iteration proceeds, since it is determined by $\delta\mathbf{x(t_0)}$. Correspondingly, only the adjoint model associated with $\tilde{\mathbf{M}}(t_i, t_0)$ is required. Thus, the adjoint model does not involve physical processes which are not used in the simplified model.

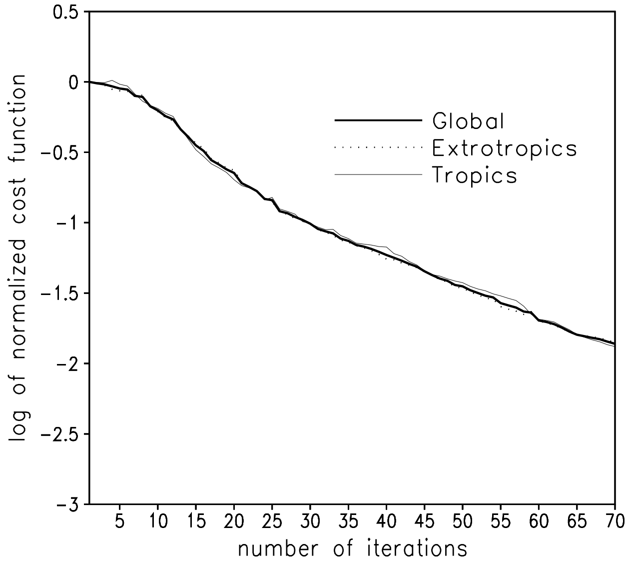

Figure 6: The logarithm (base 10) of of the normalized cost function with the number of iterations for the incremental 4D-Var. The physical processes are not included in the adjoint model. A total of 70 minimization iterations require 75 evaluations of the cost function and its gradients. The cost function is partitioned into the tropical and extratropical parts, which are also displayed.

Here we present a 4D-Var experiment implementing the incremental method. In this experiment, the simplified model does not use any physical processes except for horizontal diffusion, while all other aspects remain the same as in the full forecasting model. Accordingly, the adjoint model includes only the dynamic core and the horizontal diffusion. Thus the difference between the results of this experiment and those of the standard 4D-Var results is due only to the physics package being absent in the adjoint model.

Figure 6 shows the evolution of the normalized cost function (6) versus the number of minimization iterations. Here we use the L-BFGS minimization algorithm as in the standard 4D-Var. The cost function displays a satisfactory rate of decrease. After 70 minimization iterations, it decreases to 1.3% of its initial value.

5 Comparisons between the standard and incremental 4D-Var

The incremental method provides us with a powerful tool to examine the impact of inclusion of physical processes in the adjoint model on 4D-Var. In the incremental 4D-Var setting, we may define the simplified model by using a lower model resolution and/or by using only part of physical processes or their simplified forms (e.g., Courtier et al. 1994; Rabier et al. 1997; Veerse and Thepaut 1998). In this section, we are only concerned with the incremental 4D-Var without inclusion of physical processes in the adjoint model. With the standard 4D-Var analysis described in the previous section, we can obtain the difference between the standard 4D-Var and the incremental method. Only when the performance of the standard 4D-Var is superior to that of the incremental 4D-Var, does the inclusion of the full physics package in the adjoint model have a beneficial impact, and vice versa.

5.1 Evolution of cost functions

In this section, we compute the values of the standard cost function defined by (3) in terms of the update analysis at each iteration of the incremental 4D-Var. These values are comparable with those obtained via the standard 4D-Var, while the incremental cost function (6) differs from the standard cost function (3). Figure 7 shows the evolution of the normalized standard cost function values versus the number of minimization iterations. After 70 minimization iterations, the standard cost function value of the incremental 4D-Var decreases to 0.8% of its initial value, which is larger than the final value obtained using the standard 4D-Var. The incremental 4D-Var also exhibits a satisfactory convergence rate measured by the standard cost function.

From Fig. 8, we can infer that a descent direction obtained by the incremental method is also a descent direction of the standard 4D-Var, since the cost function is almost monotonically decreasing with the number of iterations. Figure 8 presents the cost function change for every iteration, defined as the difference between two subsequent iterations k and $k+1$. We note that all the changes are positive apart from the one at iteration 53. It is interesting

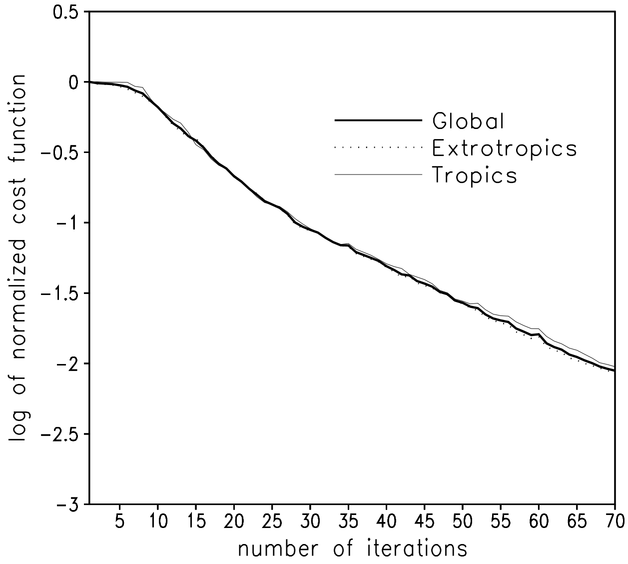

Figure 7: The logarithm (base 10) of the standard cost function evaluated in terms of the update analyses of each iteration of the incremental 4D-Var corresponding to Fig. 6. The values of the standard cost function are comparable with those obtained via the standard 4D-Var in Fig. 2.

to note that the largest changes occur between iteration 5 and iteration 25. This result is consistent with those obtained by Navon et al. (1992) amongst others, that the minimization process balances large structures during the first 15-20 minimization iterations. A similar evolution of the cost function changes with the number iterations is observed in the standard 4D-Var (Fig. 9). Qualitatively, the descent process of the incremental 4D-Var bears a strong similarity to that of the standard 4D-Var.

These results may tempt us to conclude that the physical process impact is negligible for the computation of the gradient of the cost function. In order to probe this issue in depth, we compute a distance between the "observation" and the trajectory of the model without physics. This trajectory starts from the same initial condition as the one used to produce the "observation". This distance is evaluated using the cost function of (3), where $\mathbf{x}(\mathbf{t_i})$ is the trajectory of the model without physics. The distance is solely a consequence of the absence of physics in the model. The distance has a value of 545.5, which is 11.0% of the initial cost function value. This value is one order of magnitude larger than the value of the standard cost function at iteration 70. These results can be explained by the fact that strong physical processes occur over some limited areas comprising only a small part of the

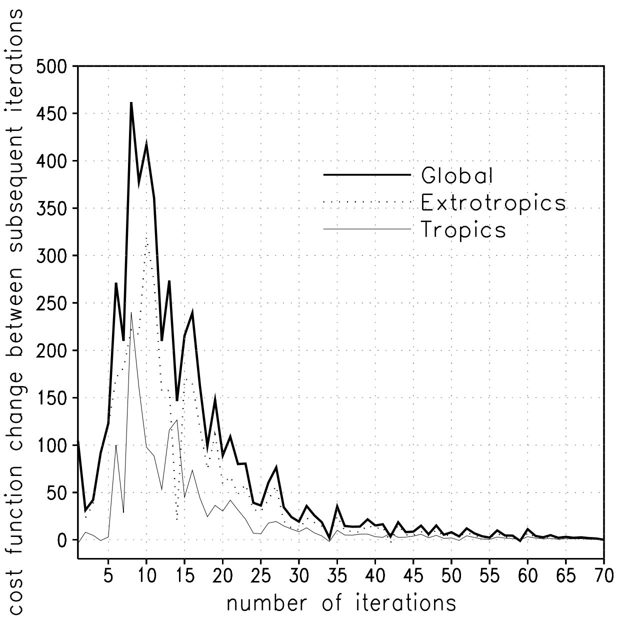

Figure 8: The cost function changes for each minimization iteration for the incremental 4D-Var. They are defined as the differences of values of the standard cost function between two subsequent iterations k and $k+1$.

model domain.

However, from a quantitative examination, we observed that the cost function displays a faster rate of descent at the first stage of the minimization process (ie., the first 25 iterations) in the incremental 4D-Var than in the standard 4D-Var. Figure 10 illustrates the difference between the incremental and standard 4D-Var versus the number of minimization iterations. Prior to iteration 50, the cost function of the standard 4D-Var is larger than the one of the incremental 4D-Var. However, after 50 iterations, the cost function of the standard 4D-Var becomes smaller than that of the incremental 4D-Var. Thus, when considered in terms of the decrease in the cost function, the standard 4D-Var with full physics adjoint is not superior in terms of rate of decrease to the incremental method until reaching iteration 50, when the cost function has decreased to about 2.5% of its initial value.

5.2 Assimilation analysis errors

We now proceed to examine the difference between assimilation errors of the standard and incremental 4D-Var after 21 minimization iterations. This is done in order to study in depth the stage of the minimization process where the cost function of the incremental 4D-Var reaches smaller values than those of the standard 4D-Var for the same number of

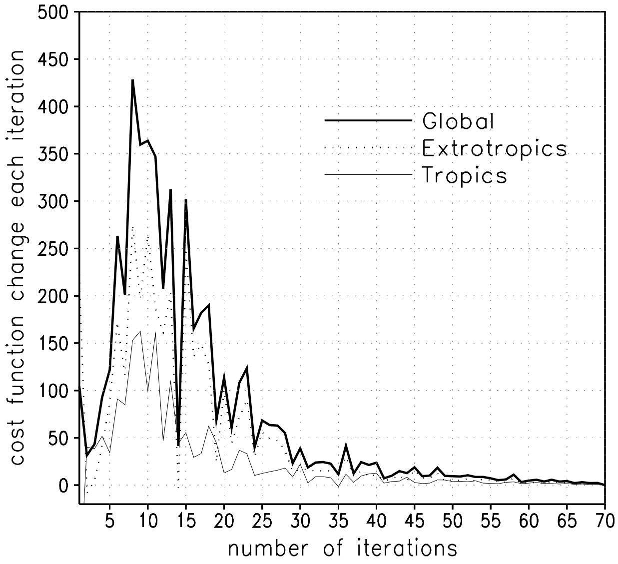

Figure 9: Same as in Fig. 8, but for the standard 4D-Var.

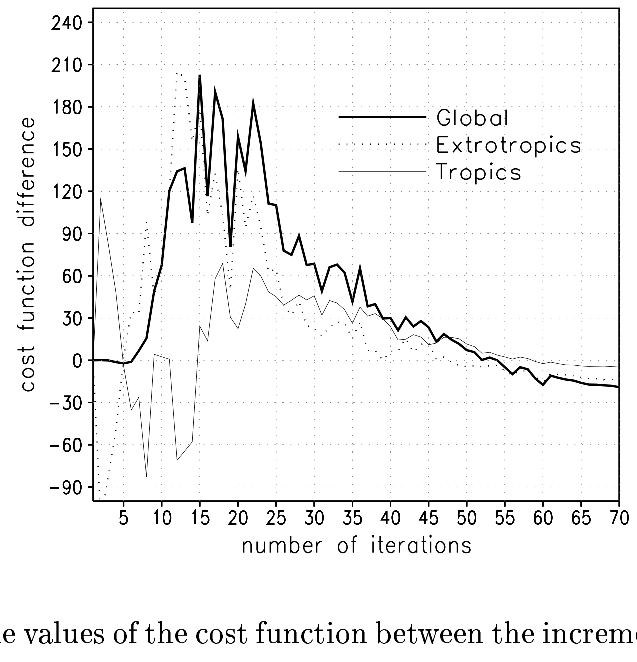

Figure 10: Differences in the values of the cost function between the incremental and standard 4D-Var versus the number of minimization iterations.

minimization iterations.

Figures 11 and 12 show the vertical distribution of the assimilation analysis RMS error along with that of the forecast RMS error at the end of the assimilation window. The RMS errors of the assimilation analyses of the standard 4D-Var are consistently larger than those of the incremental 4D-Var for all model variables and model levels. Especially, the RMS error of dewpoint depression of the standard 4D-Var is much larger than that of the incremental 4D-Var in the lower and middle troposphere. The forecast RMS errors of the incremental 4D-Var are overall smaller than that of the standard 4D-Var.

Figures 13 and 14 show the assimilation analysis error of dewpoint depression for the standard and incremental 4D-Var respectively. At σ-levels 1 and 3, the largest errors for the standard 4D-Var are located in Southern Asia and the western Pacific. Also, the largest errors are observed to be located in Southern America. The areas with the largest errors correspond to areas with a large rate of precipitation (Fig. 1). In fact, strong precipitation centers evidently correspond to centers of large assimilation analysis errors. Vukicevic and Bao (1998) found a similar local performance of 4D-Var. However, unlike the result obtained after 70 iterations where the analysis errors correlate well with strong precipitation, the precipitation is not significantly correlated with assimilation analysis errors at level 9.

A striking feature of the assimilation analysis errors in the incremental 4D-Var is that they are relatively small in regions where relatively large errors occur in the assimilation analyses of the standard 4D-Var. These regions include the precipitation areas. Further, unlike the error of the standard 4D-Var, the largest errors in the incremental 4D-Var at all model levels are not found in the strong precipitation areas in the tropics, but they rather appear to be related to special orographic features such as the Tibetan Plateau and the Rocky Mountains. Correspondingly, the major large error areas in the incremental 4D-Var are located in the extratropics, and the errors in these areas are substantially larger than those in the standard 4D-Var.

The relationship between the large assimilation errors and some special orographic features in the incremental 4D-Var strongly suggests that the large errors result from the absence of the boundary layer physical processes in the adjoint model. To validate this assertion, we

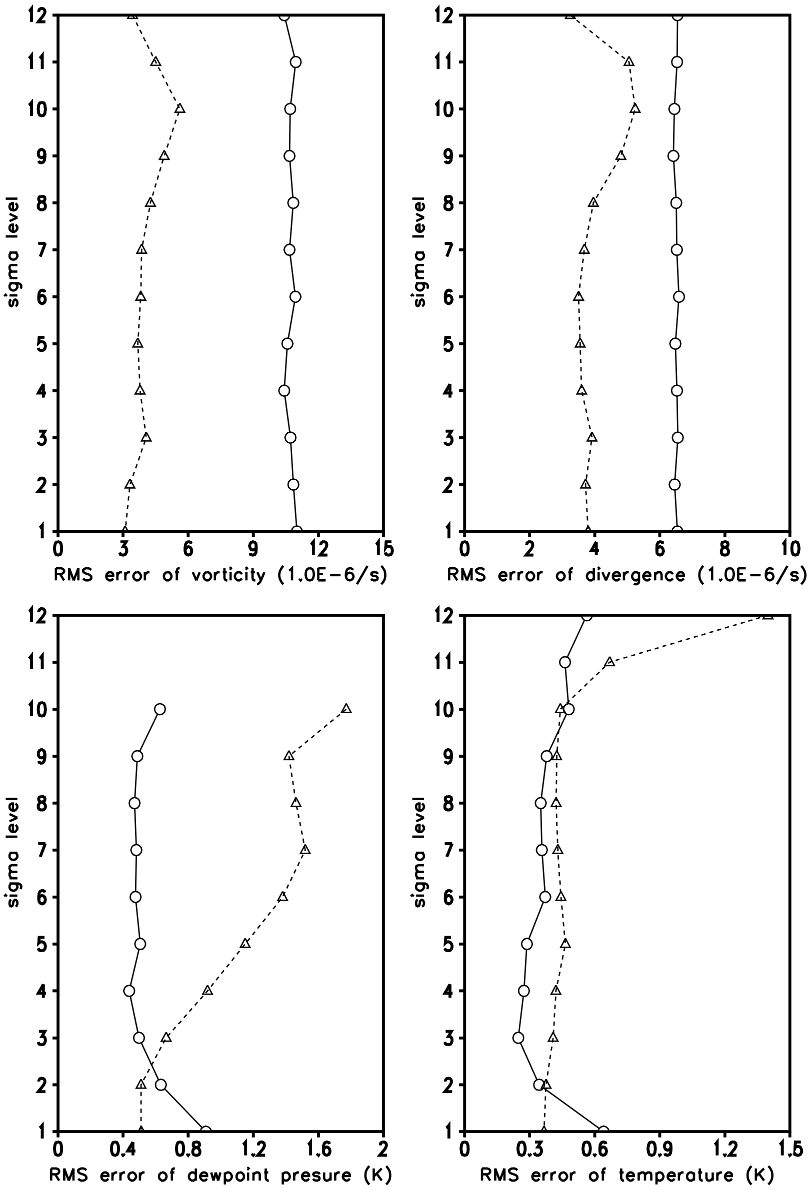

Figure 11: Vertical distributions of RMS errors for the incremental 4D-Var analyses after minimization iteration 21 and 6-h pre-forecasts starting from the assimilation analyses. The physical processes are not included in the adjoint model. Solid lines are for the assimilation analyses, and dashed lines for the 6-h forecasts.

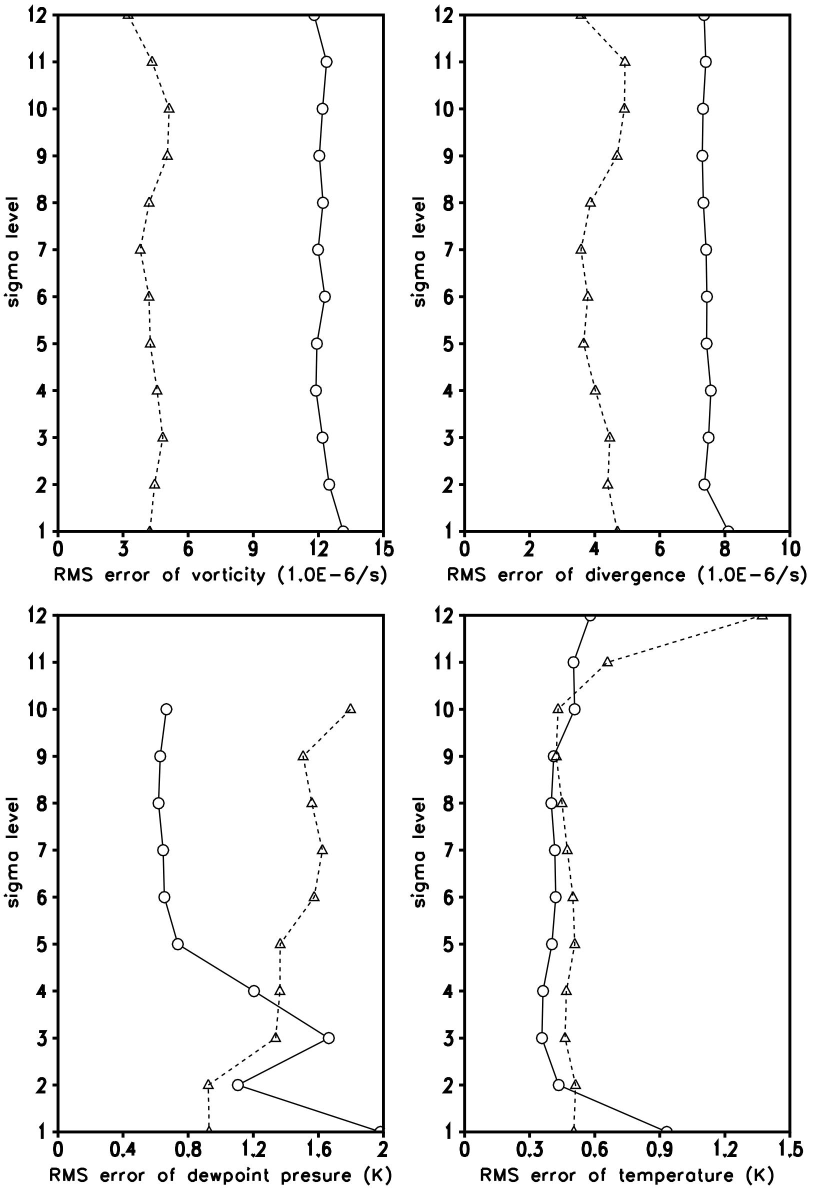

Figure 12: Same as in Fig. 11, but for the standard 4D-Var, where the full physics adjoint model is used.

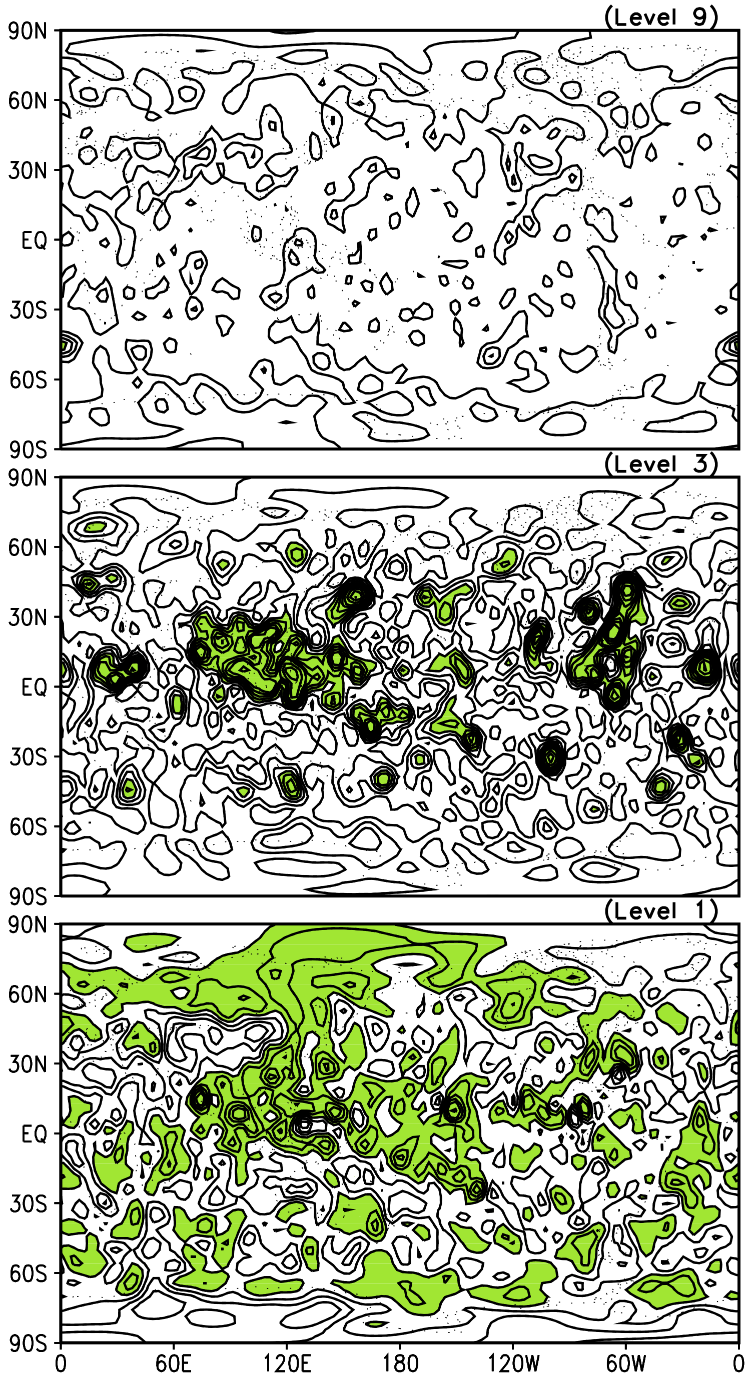

Figure 13: Errors in the assimilation analysis of dewpoint depression for the standard 4D-Var at minimization iteration 21. The contour interval is $0.5K$. Values larger than $2.0K$ are shaded. The others are the same as in Fig. 4.

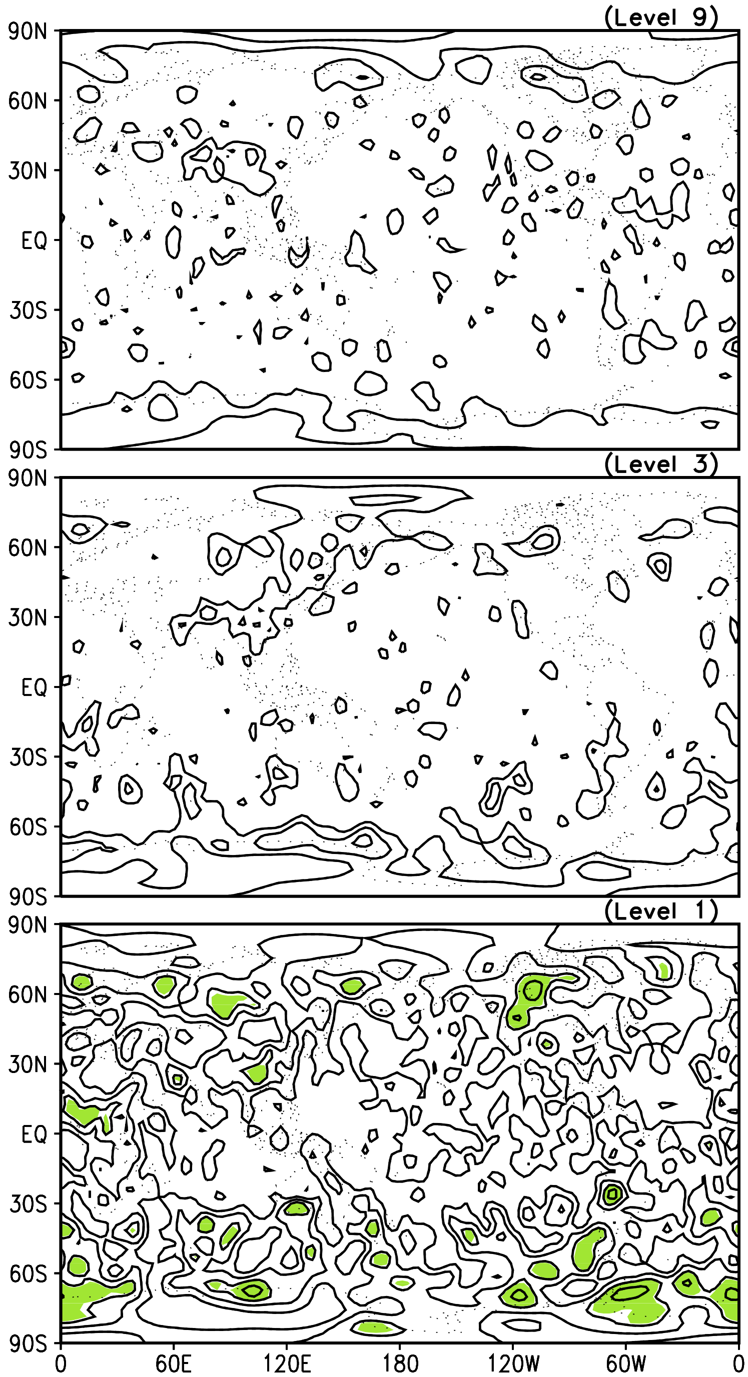

Figure 14: Same as in Fig. 13, but for the incremental 4D-Var.

perform an additional experiment. In this experiment, we repeat the incremental 4D-Var as above, but with the boundary layer physics being included into the adjoint model. The results obtained show that the assimilation analysis error at σ level 1 is indeed substantially reduced.

These results show that at the beginning stages of the minimization the inclusion of the precipitation physics displays a negative impact on the 4D-Var, and the absence of the precipitation physics in the adjoint model does not affect significantly the 4D-Var. However, the absence of the boundary layer physics in the adjoint leads to large errors in the assimilation analyses in the lower troposphere.

We now analyze the assimilation analysis errors after 70 minimization iterations, when the standard 4D-Var attains a smaller value of the cost function than that of the incremental 4D-Var. Figure 15 presents the vertical distribution of the assimilation analysis errors. Contrary to the uniform distribution in the vertical for the standard 4D-Var (Fig. 3), the RMS errors of the incremental method display significant vertical variations. The RMS errors are two to three times larger at model levels near the ground than in the middle troposphere for all model variables. Also, at model levels near the ground the RMS errors in the incremental method are larger than those in the standard 4D-Var. In the middle and upper troposphere, there is no substantial difference between the standard and incremental 4D-Var in as far as the RMS error is concerned. These results indicate that the absence of physics in the adjoint causes substantial errors at model levels in the lower troposphere.

Figure 16 illustrates the assimilation analysis error of dewpoint depression at σ levels 1, 3 and 9. For levels 1 and 3, the assimilation analyses of the incremental 4D-Var tend to have relatively small errors in the regions associated with the areas of the largest error in the assimilation analyses of the standard 4D-Var. The error distribution at level 9 is of importance, since the large errors there evidently correlate with large precipitation rates such as over Southern Asia. Remarkably, most of the large precipitation areas correspond to large error centers. Similar to the previous discussion for iteration 21, we conclude that the absence of the boundary layer physics in the adjoint model leads to the dominant assimilation error in the lower troposphere, while the absence of the precipitation physics causes a dominant

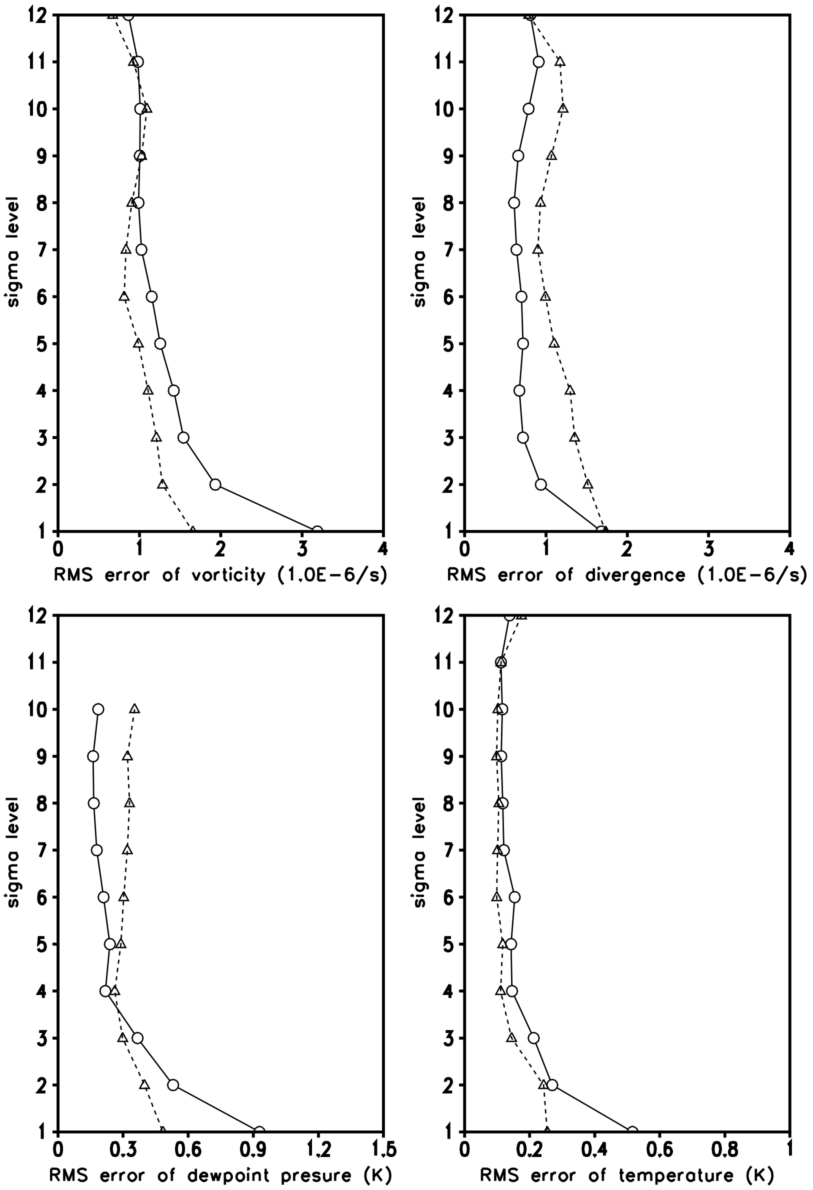

Figure 15: Same as in Fig. 5 but for the incremental 4D-Var analyses at minimization iteration 70 and the 6-h pre-forecasts starting from the assimilation analyses. The physical processes are not included in the adjoint model.

assimilation error in the upper troposphere. It is of interest to note that at levels 3 and 9, the error of the incremental 4D-Var over some regions of precipitation is larger than that of the standard 4D-Var, especially at middle and high latitudes.

The change in the assimilation analysis errors as the minimization iteration proceeds exhibits some important features. The assimilation analysis errors of the standard 4D-Var are largest over intense precipitation regions, but they are primarily confined between model levels 1 to 4, i.e., in the lower troposphere. After 70 minimization iterations, the largest errors are still concentrated over the precipitation regions, but their sizes are of a similar magnitude at all model levels. The sizes of assimilation analysis errors become consistently smaller in the lower and middle troposphere as the minimization iteration progresses from iteration 21 to iteration 70. In the upper troposphere, the error is not reduced over some small areas associated with the largest error centers.

For the incremental 4D-Var, the spatial distribution of assimilation analysis errors changes very little from minimization iteration 21 to iteration 70 at model levels 1 and 3. Corresponding to the RMS error, the error sizes do not display any reduction. At level 9, the size of the error is significantly reduced. However, relatively large error centers appear over some small areas related to precipitation. These relatively large error centers are not found at iteration 21. They either display a larger magnitude or have no counterpart in the standard 4D-Var. These observations suggest that the assimilation analysis errors due to the absence of the physics in the adjoint model can not be reduced by performing additional minimization iterations. The errors over some precipitation areas, especially at middle and high latitudes, may become dominant in the middle and upper troposphere, and have larger sizes than those present in the standard 4D-Var.

We carry out two 24 hour forecasts starting from the assimilation analyses after 70 iterations, obtained from the standard and incremental 4D-Var respectively. The difference between the two forecasts does not seem to be significant in terms of the RMS error, but the RMS errors of all the model variables are slightly smaller for the standard 4D-Var than for the incremental 4D-Var in the lower troposphere. The major difference is in the precipitation spin-up. The accumulated precipitation of the forecast in the first 6 hours is significantly

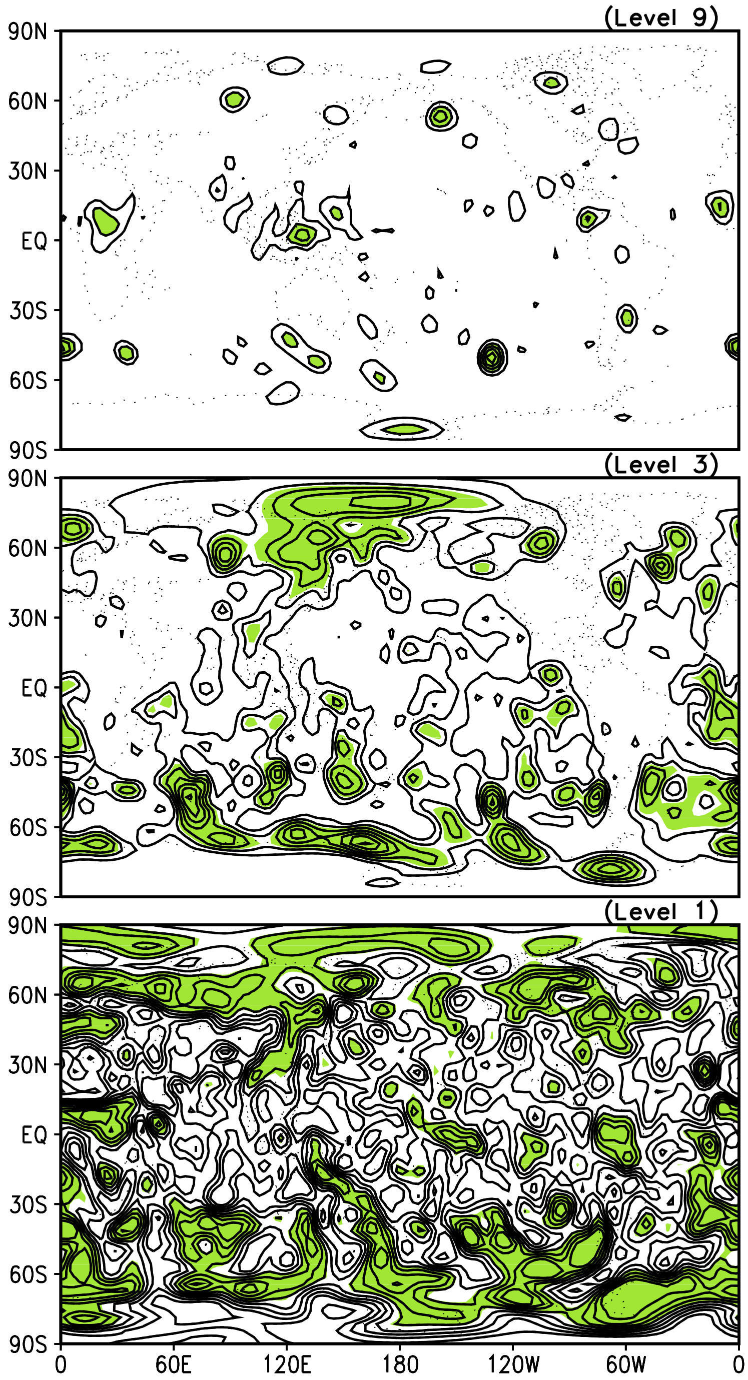

Figure 16: Same as in Fig. 4 but for the incremental 4D-Var, where no physical process is included into the adjoint model.

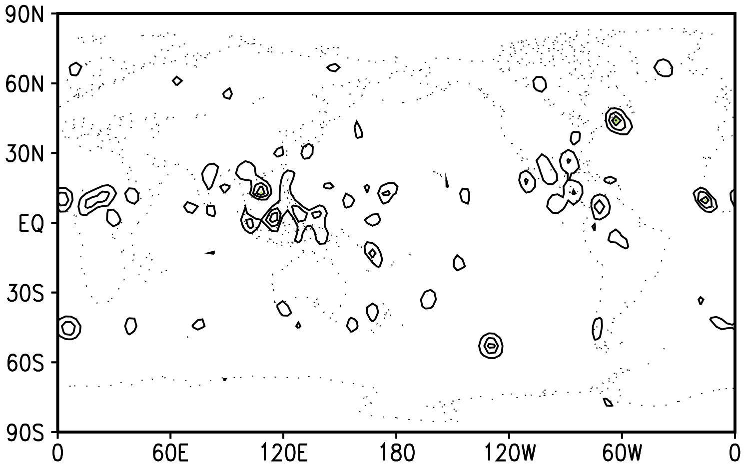

Figure 17: Errors in 3-h accumulated precipitation for the pre-forecast of the last 3-h within the 6-h assimilation window for the incremental 4D-Var. The contour interval is $0.5mm$. Values larger than $0.75mm$ are shaded.

insufficient. Figure 17 shows the errors in the 3-h accumulated precipitation during the last three hours within the assimilation window. Comparing with Fig. 1, we see that the forecast using the assimilation analyses of the incremental 4D-Var is about 20% less than the accurate amount, while the error is very small when using the assimilation analyses of the standard 4D-Var.

6 A 4D-Var strategy using full physics adjoint

As mentioned previously, inclusion of physical processes into the adjoint model sizably increases the computational cost of 4D-Var. In this model, the standard 4D-Var requires an amount of CPU time twice as large as that required by the incremental method which does not include the physics in the adjoint model. This ratio of CPU time between the standard 4D-Var and the incremental method is representative of operational models as estimated by Courtier et al. (1994) and Rabier et al. (1997). The increase in the CPU time when the full physics package is included into adjoint models may preclude the operational implementation of 4D-Var.

Another concern is the presence of strong nonlinearities in physical parameterizations. Even worse, they may contain on-off switches (Xu 1996a, b). We have shown that the inclusion of the full physics into the adjoint model is characterized by a negative impact on the assimilation analyses over precipitation regions prior to minimization iteration 50, while it has a significant positive impact on the lower troposphere during the entire minimization process. Thus, when full physics is included in the adjoint model, an adequate 4D-Var strategy is necessary in order to circumvent the abovementioned negative impact while taking advantage of the positive impact.

As mentioned in the introduction, the sequential cost function incremental approach proposed by Courtier et al. (1994) provides a possible method for alleviating the effect of nonlinearities and reducing the computational load due to the inclusion of physics into the adjoint model. We will discuss this algorithm by comparing it with the algorithmic features of the truncated Newton method, and we then introduce and test an new algorithm suitable for dealing with physical processes in the adjoint model.

6.1 The standard truncated Newton method and its variants

The truncated-Newton method is used to solve the problem

$$\text{minimize} \quad J(\mathbf{x}), \tag{9}$$

where $\mathbf{x}$ is a vector. This method computes a search direction by finding an approximate solution $\mathbf{p}$ to the Newton equation

$$\mathbf{H_k p} = -\nabla \mathbf{J}(\mathbf{x_k}) \tag{10}$$

using some iterative method, usually the conjugate-gradient method or the Lanczos method (Dembo and Steihaug 1983). Here $\mathbf{H_k}$ is the Hessian of $J(\mathbf{x})$ at $\mathbf{x_k}$. The iterative method is stopped (truncated) prior to attaining the exact solution to the Newton equation. When the exact solution is obtained, the search direction is just an Newton direction.

Thus, the standard truncated Newton method consists of nested iterations. There is an outer iteration (loop) that corresponds to a general optimization method. At each outer iteration (loop) we compute a search direction and perform a line search. The computation of the search direction uses an inner iteration corresponding to the iterative method used to solve the Newton equations.

Finding an approximate solution to the Newton equations corresponds to the computation of the approximate minimizer of the cost function

$$\tilde{J}(\mathbf{x}) = \mathbf{J}(\mathbf{x_k}) + \nabla\mathbf{J}(\mathbf{x_k})\mathbf{p} + \frac{\mathbf{1}}{\mathbf{2}}\mathbf{p}^{\mathbf{T}}\mathbf{H_k}\mathbf{p} \tag{11}$$

obtained by approximating $J(\mathbf{x})$ by a quadratic approximation consisting of the first three terms in Taylor expansion where $\mathbf{p} = \mathbf{x} - \mathbf{x_k}$. We refer to it as an inner cost function. The inner cost function has identical gradient and Hessian as those used in the outer iteration (we will refer to it as the outer cost function).

Here we do not directly use (11) but define instead another inner cost function $\hat{J}(\mathbf{x})$ approximating $\tilde{J}(\mathbf{x})$. This inner cost function $\hat{J}(\mathbf{x})$ is required to have the gradients and Hessian which are an approximation to the outer cost function, i.e.,

$$\frac{\nabla J(\mathbf{x_k})}{\mathbf{H_k}} \approx \frac{\nabla \hat{J}(\mathbf{x})}{\hat{\mathbf{H}}_k}, \tag{12}$$

where $\hat{\mathbf{H}}_k$ is the Hessian of $\hat{J}(\mathbf{x})$ at $\mathbf{x_k}$. Equation (12) guarantees that the Newton descent direction of $\hat{J}(\mathbf{x})$ is also a descent direction of $J(\mathbf{x})$. We refer to a truncated Newton method as the truncated Newton-like method where $\hat{J}(\mathbf{x})$ is used to find a search direction instead of (11).

6.2 A truncated Newton-like incremental method

A key aspect in the implementation of the truncated Newton-like method is the selection of the approximate inner cost function $\hat{J}(\mathbf{x})$. We have shown that an important feature of the incremental cost function (6) is that its descent direction is the descent direction of the standard cost function (3) when the L-BFGS algorithm is used both in the incremental and in the standard 4D-Var. The incremental cost function (6) is selected to serve as an inner cost function.

The algorithmic set-up of the truncated Newton-like method can be formulated and outlined in the following manner:

a) Use the adjoint model without physics to carry out a number of minimization iterations in terms of the incremental cost function (6), where no physics is included in the adjoint model. Here use a limited-memory quasi-Newton method (Liu and Nocedal 1988). Then obtain an analysis increment.

b) Perform a line search in terms of the standard cost function (3), using the analysis increment as the search direction, and then update the assimilation analysis.

c) Carry out a number of minimization iterations of the outer loop where the full physics adjoint model is used. Here also use the limited-memory quasi-Newton method (Liu and Nocedal 1988).

d) Repeat the cycle consisting of steps a)-c).

The standard truncated Newton method does not include step c). Step c) is introduced here in order to take into account the effect of inclusion of full physics in the adjoint model. Due to step c), the truncated Newton-like method can be viewed as a variant of the sequential cost function incremental approach proposed by Courtier et al. (1994). Thus we refer to this method as a truncated Newton-like incremental method. We also refer to both steps b) and c) as being the outer loop. The number of minimization iterations to be conducted in each loop should be determined according to the total number of minimization iterations permitted. Generally, during the first cycles, a relatively small number or no iterations at all are conducted in step c), while a relatively large number of iterations is conducted in step a).

The crucial aspect for speeding up the rates of descent of this algorithm is the Hessian update information. The limited-memory quasi-Newton method is based on the idea of computing the descent direction p_k as $-Mg_k$, where k is the iteration step, g_k is the gradient of the cost function at $\mathbf{x_k}$, and M is an approximation to the Hessian. The matrix M is a positive-definite matrix obtained by updating a prescribed diagonal matrix with a limited number (5-7) of quasi-Newton corrections (Nocedal 1980; Liu and Nocedal 1988). Under the approximation of (12), the updating quasi-Newton corrections gathered in both loops can be used as if they were all within the same loop.

6.3 4D-Var experiments using the truncated Newton-like incremental method

To avoid the strong effect of nonlinearities in the physical processes, a relatively large number of iterations in step a) should be carried out in the first cycle. Thus the distance between the updated analysis and the solution would have been substantially reduced when the standard 4D-Var starts. This is mandated by the fact that the standard 4D-Var is subject to nonlinearities, and smaller perturbations may weaken the effect of nonlinearities. We perform 40 minimization iterations in step a), and 30 minimization iterations in step c). Then only one cycle of the truncated Newton-like incremental method is carried out. This experiment is mainly intended to examine the expected benefit of its application towards alleviating the effect of nonlinearities in the physics.

Figure 18 shows the normalized cost function versus the number of iterations. The final normalized cost function reaches a value 0.4% of its initial value at the end of the entire cycle. The normalized cost function value reaches a value of 0.5% of its initial value in the standard 4D-Var using full physics in the adjoint model after 70 iterations, and 0.8% of its initial value in the incremental method without physics in the adjoint model. In terms of the rate of decrease of the cost function, the one cycle truncated Newton-like method is superior to either the standard 4D-Var or the incremental method.

Figure 19 presents the vertical distribution of the RMS errors of assimilation analyses along with those for the forecast RMS error at the end of the assimilation window. The RMS errors of the one cycle truncated Newton-like method are smaller than those of the standard 4D-Var. It is more interesting to compare the results of the one cycle truncated Newton-like method with the incremental method without physics in the adjoint model (Fig. 3). The assimilation analysis error for model levels near the ground is much smaller than that in the incremental method. The assimilation analysis error in the new method is also somewhat smaller for the divergence field in the upper troposphere compared with the incremental 4D-Var without physics in the adjoint model. Corresponding to these improvements, the error in the 3-h accumulated precipitation during the last 3-h within the assimilation window also become very small, being close to that obtained from the standard 4D-Var. It should be pointed out that the error at model levels near the ground is still considerably larger than the one present in the standard 4D-Var using full physics in the adjoint model.

To a large degree, the one cycle truncated Newton-like incremental method accomplishes the aim of alleviating the negative effects of the nonlinearities while taking into account the influence of physical processes in a satisfactory manner. The CPU time required increases by only 40% compared with that required by the incremental method which involves no physics in the adjoint model.

We have assumed that continuously updating Hessian information at every iteration in both the inner loop and the outer loop may accelerate the convergence of the minimization procedure. To verify this point, we conduct an experiment in which the minimization is restarted without using Hessian information in the inner loop when the outer loop starts. Figure 18 shows that over the entire outer loop the cost function decreases in average by 10% faster when the Hessian information is continuously updated.

The major deficiency in the one cycle truncated Newton-like incremental method is that the error at model levels near the ground is still larger than that present in the standard 4D-Var. Indeed, the main cause of this deficiency is that the effect of the boundary physics is not sufficiently taken into account. This is due to the fact that the boundary physics starts impacting the 4D-Var from the starting stages of the minimization process. To remedy this deficiency, we devise a two-cycle experiment, in which the inner loop consists of 25 minimization iterations, and the outer loop consists of 10 minimization iterations in the first cycle. In the second cycle, the inner loop uses 20 minimization iterations and the outer loop 15 minimization iterations. The computational cost increases by 35% compared with the incremental method without physics in the adjoint model.

The resulting assimilation analyses using the two cycle truncated Newton-like incremental method experience an improvement in all the aspects examined. The value of the cost function at the end of the minimization process decreases to 0.3% of its initial value, which is smaller than the values obtained by both the standard and incremental 4D-Var. Figure 20 presents the vertical distribution of the RMS errors of assimilation analyses along with those for the forecast RMS error at the end of the assimilation window. As expected, the large

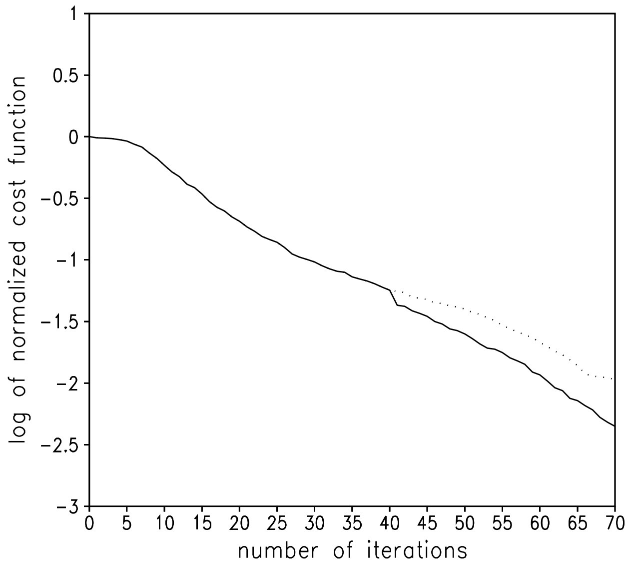

Figure 18: The logarithm (base 10) of normalized cost function for the one-cycle truncated Newton-like incremental method versus the number of minimization iterations. The incremental approach is used for the first 40 minimization iteration, where no physics is included in the adjoint model. The standard 4D-Var minimization approach is used for the last 30 minimization iterations, where the full physics adjoint model is used. For the solid line, the information on the Hessian is jointly updated at every iteration in both the inner and outer loops. For the dotted line, the minimization process is restarted when the minimization process switches from one to another loop.

error at model levels near the ground disappears. It is very encouraging to find that the RMS error is consistently smaller than that present in both the standard and incremental 4D-Var for every model level and for all model variables. This advantage is also preserved in ensuing 24-h forecasts as well as in precipitation spin-up, when using the two cycle truncated Newton-like incremental method.

7 Summary and Discussions

In this paper, we presented results of several 4D-Var experiments. These experiments included both the standard 4D-Var where full physics was used in the adjoint model, and versions of incremental 4D-Var where only selected physical processes were used. The comparison between these experiments provided us a detailed understanding of how physical processes act on a 4D-Var procedure as the minimization process proceeds.

Results obtained showed that during a rather large number (about 50) of iterations from the starting stages of the minimization process, the analysis errors in the standard 4D-Var were overall larger than those in the incremental 4D-Var analyses where no physics was included in the adjoint model. The major assimilation analysis errors in the standard 4D-Var were found to be located over intense precipitation regions. On the contrary, for the incremental 4D-Var analyses without any physics in the adjoint model the major errors did not correlate with precipitation, and the analysis errors over the large precipitation regions were smaller than those in the standard 4D-Var analyses. Only after 50 minimization iterations, did the errors over some precipitation regions gradually become larger in the incremental 4D-Var analyses than those in the standard 4D-Var. Thus, the inclusion of precipitation physics in the adjoint model appeared to be detrimental to convergence rates of the minimization process prior to minimization iteration 50, and then gradually turned out to become beneficial. Interestingly, the 4D-Var with full physics in the adjoint model did indeed speed-up the precipitation spin-up compared with the incremental 4D-Var without physics in the adjoint model after iteration 70. However, the reduction of the precipitation spin-up period cannot be solely attributed to the inclusion of the precipitation physics in the adjoint model, but may be the result of the combined impact of several physical processes.

In the incremental 4D-Var where no physics was involved in the adjoint model, major assimilation analysis errors were found to be primarily located in the lower troposphere, and exhibited a correlation with some special orographic features. Especially, the errors in the dewpoint depression and temperature analyses could not be further reduced after 20 minimization iterations. These large errors can be primarily attributed to the absence of the boundary physics in the adjoint model. Further, the major assimilation analysis errors

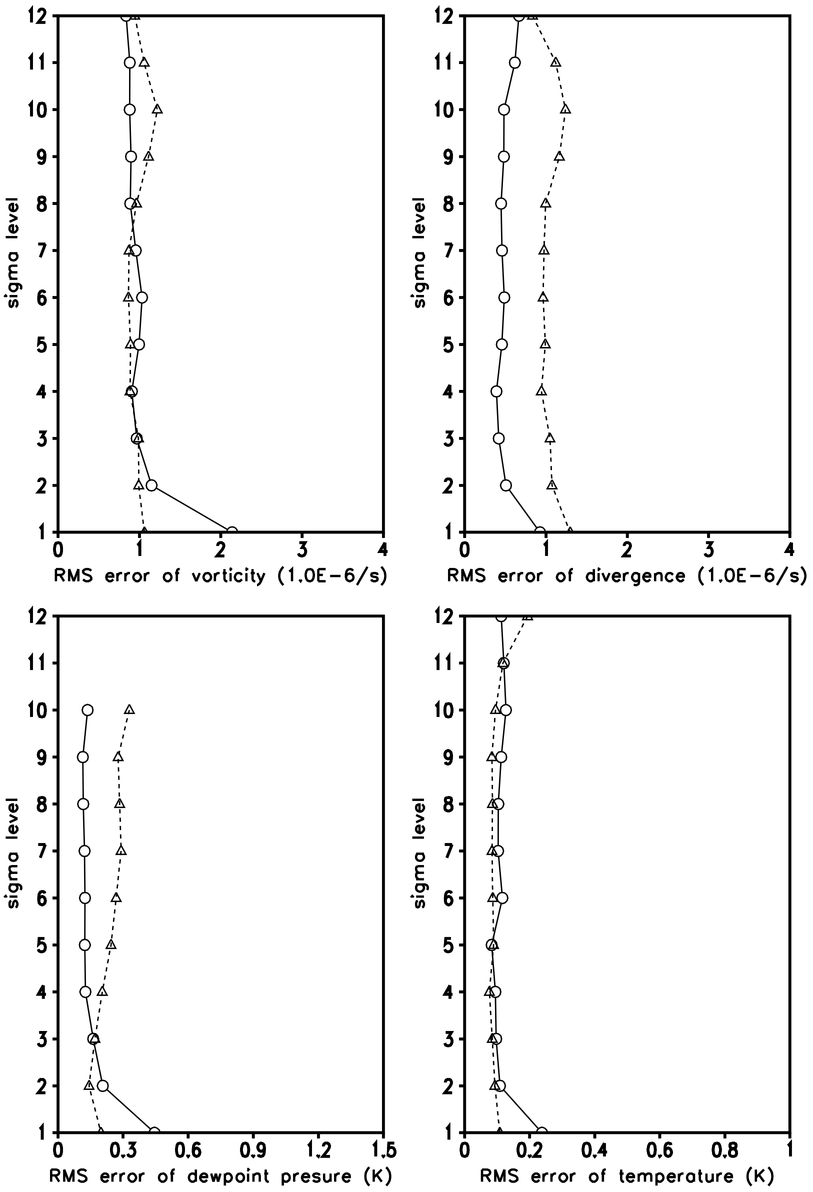

Figure 19: Same as in Fig. 5 but for the one-cycle truncated Newton-like incremental method at the end of the minimization processes.

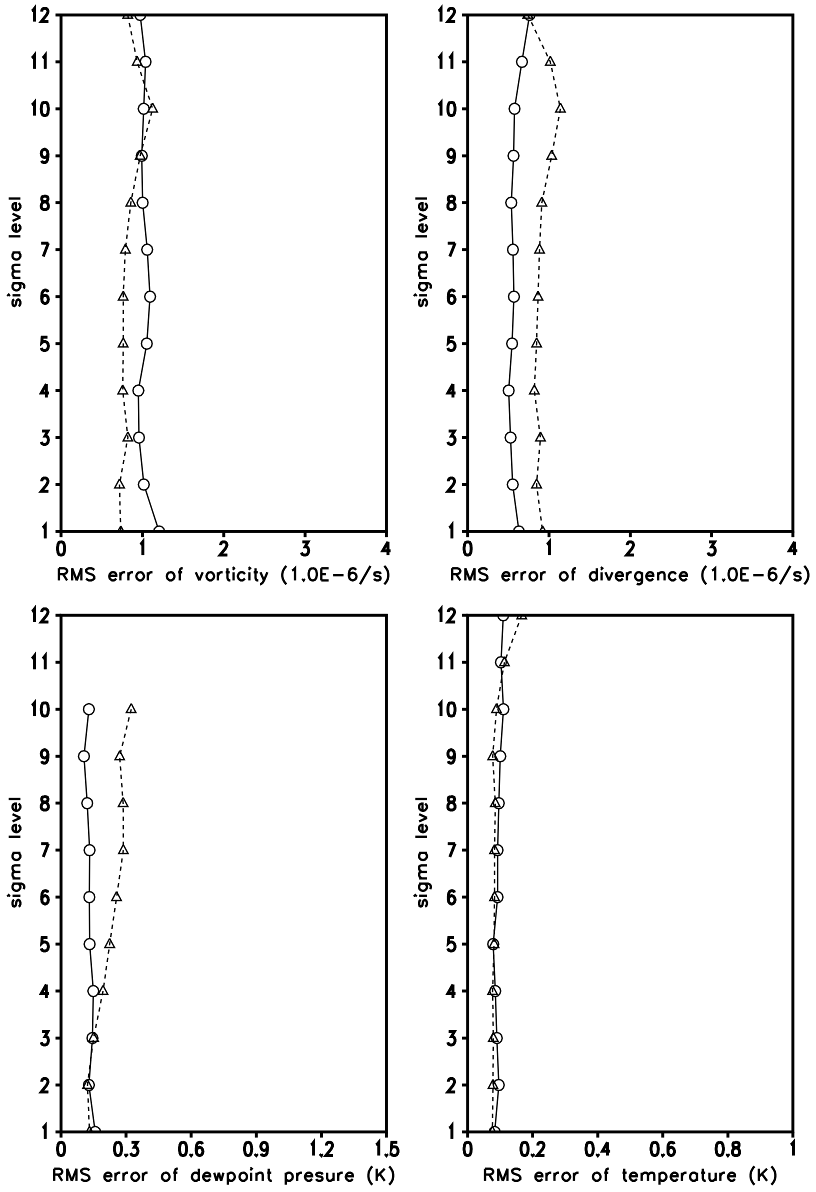

Figure 20: Vertical distributions of the RMS errors for the two-cycle truncated Newton-like incremental method at the end of the minimization processes. In the first cycle, the incremental approach is used for the first 25 minimization iterations where no physics is included into the adjoint model, and the standard 4D-Var approach is used for the last 10 minimization iterations where the full physics adjoint model is used. In the second cycle, the incremental approach is used for the first 15 minimization iterations, while the standard 4D-Var minimization approach is used for the last 15 minimization iterations.

were always associated with the absence of the boundary layer physics during the entire minimization procedure. Unlike the precipitation physics, the boundary layer physics in the adjoint model impacted positively the 4D-Var during the entire minimization procedure. The importance of including the boundary layer physics in the adjoint model has been recognized in several studies (e.g. Buizza 1994; Rabier et al. 1997).

The influence of precipitation physics warrants a detailed discussion. Assimilation errors in the standard 4D-Var using full physics in the adjoint model did not exhibit a reduction over precipitation regions at least prior to iteration 21. The reduction in assimilation errors over precipitation regions occurred at the latest stages of the minimization process. A further explanation can be provided in terms of the eigenvalue spectrum of the Hessian of the cost function. We conjecture that precipitation physics is related to eigenvectors of the Hessian associated with smaller eigenvalues. Since the minimization process generally deals with smaller eigenvalues at later stages of the minimization process in the framework of the L-BFGS minimization algorithm, assimilation errors can only be reduced at the later stages of the minimization. This also means that if this mechanism prevails for the 4D-Var with full physics, then in order to beneficially extract the additional information provided by the full physics 4D-Var, a larger number of minimization iterations must be carried out than in the adiabatic case. In order to alleviate this situation, an efficient preconditioning approach is mandatory.

The influence of the physical processes in the adjoint model on 4D-Var was found to exhibit a complex behavior. The precipitation physics was detrimental in terms of assimilation analysis errors to the 4D-Var at the beginning stage of the minimization process, but became beneficial in the later stages of the minimization process. This behavior appears to be related to the presence of strong nonlinearities in the precipitation physics. As mentioned in the introduction, the linearization approximation may lose its validity due to relatively large sizes of the perturbations at the starting stages of the minimization process. As the minimization process proceeded, the sizes of the perturbations became smaller, and as a consequence the linearization approximation became better. At this stage the advantage of the consistency between the nonlinear forecasting model and the adjoint model started playing a dominant role. However, the boundary layer physics in the adjoint model turned out to be beneficial during the entire minimization process. This may be attributed in part to the relatively weak nonlinearities present in the boundary physics.

The results obtained suggest some issues which should be taken into account in order to maximize the beneficial impact when 4D-Var uses full physics in the adjoint model. At the starting stages of the minimization process, the 4D-Var would be better implemented using the incremental method which does not include full physics in the adjoint model. 4D-Var experiments have shown that a minimization process generally acts first on the larger scales (Thepaut and Courtier 1991; Navon et al. 1992; Tanguay et al. 1995). Thus the small scales can be dealt with only during the last stage of the minimization process as in Versee and Thepaut (1998). We suggest to avoid using full physics in the adjoint model at the starting stages of the minimization process in order to avoid the detrimental effect of the strong nonlinearities, which does not mean that physics has little impact on the 4D-Var at this stage. To take into account the impact of full physics, we should use full physics in the adjoint model repeatedly after the starting stages of the minimization. Here we have proposed to avoid using continuously the full physics package in the adjoint model during the entire minimization process, a suggestion justified theoretically by the recognition that physical processes are controlled to a large degree by dynamical processes. This observation is validated by the numerical experiments presented in this research.

The above mentioned requirements have been fulfilled using the new truncated Newton-like incremental method. The algorithm is a variant of the incremental method using a sequence of cost functions. It consisted of an inner loop and an outer loop. The incremental method including no physics in the adjoint model comprised the inner loop, while the outer loop consisted of the standard 4D-Var using the full physics adjoint model. In the standard truncated Newton method a line search is performed when the minimization process moves from the inner loop to the outer loop, while preconditioning is used in the inner loop. We used the L-BFGS algorithm for both the outer and inner loops. This allowed updating the Hessian continuously at every minimization iteration in both the outer and inner loops. A two-cycle truncated Newton-like incremental experiment was performed. In the first cycle, 25 minimization iterations were conducted in the inner loop using the adjoint model without physics and 10 minimization iterations in outer loop using the full physics adjoint model. In the second cycle, 20 minimization iterations were conducted in the inner loop and 15 minimization iterations in outer loop. The results obtained showed the quality of the assimilation analyses to be better than that obtained from either the standard 4D-Var or the incremental 4D-Var after 70 iterations in all the aspects examined. This two-cycle truncated Newton-like incremental method accomplishes the aim of alleviating the negative effects of the nonlinearities in the physics along with satisfactorily taking into account the impact of physical processes. The CPU time required increased by 35% compared with that required by the incremental method which involves no physics at all in the adjoint model.

We recognize that the way in which various physical mechanisms operate is highly model-dependent. Some of the conclusions derived in this study may be restricted to the particular model used here, namely the FSUGSM. There are still many other open problems to consider. We used only a simple scaling method to precondition the minimization. Better preconditioning methods are available such as those of Courtier et al (1994), Yang et al (1996), and Zupanski (1997) to cite but a few. We do not know if better preconditioning methods available will reduce the detrimental effect of the strong nonlinearities present in the physics at the beginning stages of the minimization process. In this research, we only used observations of wind, temperature, dewpoint depression, and surface pressure fields in a twin experiment frame. Observations sensitive to physics were not used, such as precipitation amount and total precipitable water, or other observations from satellite platforms. These problems require further investigation.

Acknowledgements

We acknowledge the support from the NSF grant number ATM-9731472 managed by Dr. Pamela Stephens whom we would like to thank for her support and from the Supercomputer Computations Research Institute, which is partially funded by the Department of Energy through Contract DE-FG05-85ER250000. The computer facilities were provided by the Special SCD Grant XXXXXX in NCAR and National Science Foundation 35111100.

REFERENCES

Businger, J.A., J.C. Wyngard, Y. Izumi and E.F. Bradley, 1971: Flux profile relationships in the atmospheric surface layer. *J. Atmos. Sci.*, **28**, 181-189.

Buizza., R. 1994: Sensitivity of optimal unstable structures. *Q. J. R. Meteorol. Soc.*, **120**, 429-451.

Chao, W. C., and Chang L.-P., 1992: development of a four-dimensional variational analysis system using the adjoint method at GLA. Part 1: Dynamics. *Mon. Wea. Rev.*, **120**, 1661-1674.

Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II. Numerical results. *Q. J. R. Meteorol. Soc.*, **113**, 1129-1347.

Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart J. Roy. Meteor. Soc.*, **120**, 1367-1388.

Dembo, R.S., and T. Steihaug, 1983: Truncated-Newton algorithms for large-scale unconstrained optimization. *Math. Programming*, **26** 190-212.

Fillion, L., and R. Errico, 1997: Variational assimilation of precipitation data using moist convective parameterization schemes: A 1 D-var study. *Mon. Wea. Rev.*, **125**, 2917-2942.

Harshvardan, and T. G. Corsetti, 1984: Long-wave parameterization for the UCLA/GLAS GCM. *NASA Tech. Memo. 86072, Goddard Space Flight Center*, Greenbelt, MD 20771, 52 pp.

Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation: Operational sequential and variational. In Data Assimilation, Meteorology and Oceanography: Theory and Practice, *J. Meteorol. Soc Japan*, **75, 1B**, 71-79.

Kanamitsu, M., K. Tada, K. Kudo, N. Sato, and S. Isa, 1983: Description of the JMA operational spectral model. *J. Met. Soc. Japan*, **61**, 812-828.

Krishnamurti, T. N., H. S. Bedi, and K. Ingles, 1993: Physical initialization using the SSM/I rain rates. *Tellus*, **45A**, 247-269.

Krishnamurti, T. N., J. Xue, H. S. Bedi, K. Ingles, and D. Oosterhof, 1991: Physical initialization for numerical weather prediction over the tropics. *Tellus*, **43A**, 53-81.

Lacis, A. A. and J. E. Hansen, 1974: A parameterization of the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118-133.

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

Lewis, J., and J. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309-327.

Li, Z., and I. M. Navon, 1997: Adjoint sensitivity of the Earth's radiation budget in the NCEP/MRF model. *J. Geophy. Res.*, **103, D4**, 3801-3814.

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

Lorenc A.C., 1986: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194.

Louis, J.F., 1979: A parametric model of vertical eddy fluxes in the atmosphere. *Boundary Layer Meteorology*, **17**, 187-202.

Mahfouf, J.-F., 1998: Influence of physical processes on the tangent-linear approximation. *Tellus*, in press.

Mahfouf, J.-F., R. Buizza, and R. M. Errico, 1996: Strategy for including physical processes in the ECMWF variational data assimilation system. *In proceeding of the ECMWF workshop on non-linear aspects of data assimilation*, 9-11 September, Reading, UK.

Navon, I. M., X. Zou, M. Berger, P. K. Phua, T. Schlick, and F. X. Le Dimet, 1992a: Numerical experience with limited memory quasi-Newton and truncated Newton methods. *Optimization Techniques and Applications*, Vol. 1, Edited by K. H. Phua, World Scientific Publishing Co., pp33-48.

Navon, I. M., X. Zou, J. Derber, and J. Sela, 1992b: Variational data assimilation with the NMC spectral model. Part 1: Adiabatic model tests. *Mon. Wea. Rev.*, **120**, 1433-1446.

Nash, S. G., and J. Nocedal, 1991: A numerical study of the limited memory BFGS method and the truncated-Newton method for large scale optimization. *SIAM Journal on Optimization*, 1, 358-372.

Nocedal, J., 1980: Updating quasi-Newton matrices with limited storage. *Math. of Computation*, **35**, 773-782.

Rabier, F., Mahfouf J.-F., M. Fisher, H. Jarvinen, A. Simmons, E. Andersson, F. Bouttier, P. Courtier, H. Hamrud, J. Haseler, A. Hollingsworth, L. Isaksen, P. Klinker, S. Saarinen, C. Temperton, J.-N. Thepaut, P. Unden and D. Vasiljevic, 1997: Recent experimentation on 4D-Var and first results from a simplified Kalman filter. *ECMWF Research Department Tech. Memo. No. 240*, ECMWF, Reading, UK, pp42.

Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart J. Roy. Meteor. Soc.*, **124**, 1861-1887.

Schlick, T., and A. Fogelson, 1992: TNPACK - A truncated Newton package for large-scale problems: I. Algorithms and usage. *ACM Transactions on Mathematical Software*, 18, 46-70.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I. Theory. *Q. J. R. Meteorol. Soc.*, **113**, 1131-1328.

Tanguay, M., Bartello, and P. Gauthier, 1995: Four dimensional data assimilation with a wide range of scales. *Tellus*, **47A**, 974-997.

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

Tsuyuki, T., 1996: Variational data assimilation in the tropics using precipitation data. Part II: 3-D model. *Mon. Wea. Rev.*, **124**, 2545-2561.

Tsuyuki, T., 1997: Variational data assimilation in the tropics using precipitation data. Part III: Assimilation of SSM/I precipitation rates. *Mon. Wea. Rev.*, **125**, 1447-1464.

Veerse, F., and J.-N. Thepaut, 1998: Multiple-truncated incremental approach for four-dimensional variational assimilation. *Quart. J. Roy. Meteorol. Soc.*, **124**, 1889-1908.

Vukicevic, T., and J.-W. Bao, 1998: The effect of linearization errors on 4DVAR data assimilation. *Mon. Wea. Rev.*, **126**, 1695-1706.

Wang, Z., 1993: Variational data assimilation with 2-D shallow water equations and 3-D FSU Global Spectral Models, *Ph. D. Dissertation*, Department of Mathematics, The Florida State University, 235 pp.

Wang, Z., I. M. Navon, F. X. Le Dimet, and X. Zou, 1992: The second order adjoint analysis: Theory and Application. *Meteor. Atmos. Phys.*, 50, 3-20.

Wang, Z., I. M. Navon, X. Zou, and F. X. Le Dimet, 1995: A truncated-Newton optimization algorithm in meteorology applications with analytic hessian/vector products. *Computational Optimization and Applications*, 4, 241-262.

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

Xu, Q., 1996b: Generalized adjoint for physical processes with parameterized discontinuities. Part II. Vector formulations and matching conditions. *J. Atmos. Sci.*, **53**, 1143-1155.

Xu, Q., 1997: Generalized adjoint for physical processes with parameterized discontinuities. Part IV. Problems and treatment in time discretization. *J. Atmos. Sci.*, **54**, 2722-2728.

Yang, W., I. M. Navon, and P. Courtier, 1996: A new Hessian preconditioning method applied to variational data assimilation experiments using adiabatic version of NASA/GEOS-1 GCM. *Mon. Wea. Rev.*, **124**, 1000-1017.

Zhu, Y., and I.M. Navon, 1997: Documentation of the tangent-linear and adjoint models of the radiation and boundary layer parameterization packages of the FSU Global Spectral Model T42L12. *Tech. Report*, FSU-SCRI-97-98.

Zhu, Y., and I. M. Navon, 1998a: Impact of parameter estimation on the performance of the FSU Global Spectral Model using its full-physics adjoint. *Mon. Wea. Rev.*, in press.

Zhu, Y., and I. M. Navon, 1998b: FSU-GSM forecast error sensitivity to initial conditions: application to Indian summer monsoon. Meteor. Atmos. Phys., in press.

Zou, X., 1997: Tangent linear and adjoint of "on-off" processes and their feasibility for use in 4-dimensional variational assimilation, *Tellus*, **49A**, 3-31.

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

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

Zou, X., I.M. Navon, M. Berger, K. H. Phua, T. Schlick, and F. X. Le Dimet, Numerical experience with limited-memory quasi-Newton methods and truncated Newton methods.

SIAM J. on Numerical optimization, 3, 582-608.

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

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

Zupanski, M., 1996: A preconditioning algorithm for four-dimensional variational data assimilation. *Mon. Wea. Rev.*, **124**, 2562-2573.

Figure Captions

Figure 1. Accumulated precipitation of the reference forecast ('observation') over the last 3-h within the assimilation window. The contour interval is $1.0mm$. Values larger than 1.5 mm are shaded.

Figure 2. The logarithm (base 10) of the normalized cost function with number of iterations for the standard 4D-Var, in which the full physics adjoint model is used. The cost function is partitioned into the tropical and extratropical parts, which are also displayed. A total of 70 minimization iterations are carried out, while 75 evaluations of the cost function and its gradients are required.

Figure 3. Values of single model variable contributions to both the tropical and extratropical parts of the cost function for the standard 4D-Var. The others are the same as in Fig. 2.

Figure 4. Errors in the assimilation analysis of dewpoint depression for the standard 4D-Var for representative model σ-levels after 70 minimization iterations. Level 1 represents model levels near the ground, level 3 in the lower troposphere and level 9 in the upper troposphere. The contour interval is $0.25K$. Values larger than $0.5K$ are shaded.

Figure 5. Vertical distributions of RMS errors for the standard 4D-Var analyses at minimization iteration 70 and 6-h pre-forecasts starting from the assimilation analyses. The full physics adjoint model is used. Solid lines are for the assimilation analyses, and dashed lines for the 6-h forecasts.

Figure 6. The logarithm (base 10) of of the normalized cost function with the number of iterations for the incremental 4D-Var. The physical processes are not included in the adjoint model. A total of 70 minimization iterations require 75 evaluations of the cost function and its gradients. The cost function is partitioned into the tropical and extratropical parts, which are also displayed.

Figure 7. The logarithm (base 10) of the standard cost function evaluated in terms of the update analyses of each iteration of the incremental 4D-Var corresponding to Fig. 6. The values of the standard cost function are comparable with those obtained via the standard 4D-Var in Fig. 2.

Figure 8. The cost function changes for each minimization iteration for the incremental 4D-

Var. They are defined as the differences of values of the standard cost function between two subsequent iterations k and $k+1$.

Figure 9. Same as in Fig. 8, but for the standard 4D-Var.

Figure 10. Differences in the values of the cost function between the incremental and standard 4D-Var versus the number of minimization iterations.

Figure 11. Vertical distributions of RMS errors for the incremental 4D-Var analyses after minimization iteration 21 and 6-h pre-forecasts starting from the assimilation analyses. The physical processes are not included in the adjoint model. Solid lines are for the assimilation analyses, and dashed lines for the 6-h forecasts.

Figure 12. Same as in Fig. 11, but for the standard 4D-Var, where the full physics adjoint model is used.

Figure 13. Errors in the assimilation analysis of dewpoint depression for the standard 4D-Var at minimization iteration 21. The contour interval is $0.5K$. Values larger than $2.0K$ are shaded. The others are the same as in Fig. 4.

Figure 14. Same as in Fig. 13, but for the incremental 4D-Var.

Figure 15. Same as in Fig. 5 but for the incremental 4D-Var analyses at minimization iteration 70 and the 6-h pre-forecasts starting from the assimilation analyses. The physical processes are not included in the adjoint model.

Figure 16. Same as in Fig. 4 but for the incremental 4D-Var, where no physical process is included into the adjoint model.

Figure 17. Errors in 3-h accumulated precipitation for the pre-forecast of the last 3-h within the 6-h assimilation window for the incremental 4D-Var. The contour interval is $0.5mm$. Values larger than $0.75mm$ are shaded.

Figure 18. The logarithm (base 10) of normalized cost function for the one-cycle truncated Newton-like incremental method versus the number of minimization iterations. The incremental approach is used for the first 40 minimization iteration, where no physics is included in the adjoint model. The standard 4D-Var minimization approach is used for the last 30 minimization iterations, where the full physics adjoint model is used. For the solid line, the information on the Hessian is jointly updated at every iteration in both the inner and outer

loops. For the dotted line, the minimization process is restarted when the minimization process switches from one to another loop.

Figure 19. Same as in Fig. 5 but for the one-cycle truncated Newton-like incremental method at the end of the minimization processes.

Figure 20. Vertical distributions of the RMS errors for the two-cycle truncated Newton-like incremental method at the end of the minimization processes. In the first cycle, the incremental approach is used for the first 25 minimization iterations where no physics is included into the adjoint model, and the standard 4D-Var approach is used for the last 10 minimization iterations where the full physics adjoint model is used. In the second cycle, the incremental approach is used for the first 15 minimization iterations, while the standard 4D-Var minimization approach is used for the last 15 minimization iterations.