Impact of Parameter Estimation on the Performance of the FSU Global Spectral Model Using its Full-Physics Adjoint

Yanqiu Zhu, *I. M. Navon

Geophysical Fluid Dynamics Institute

*Department of Mathematics and SCRI

Florida State University

Tallahassee, FL 32306

August 7, 1998

Abstract

The full-physics adjoint of the FSU Global Spectral Model at resolution T42L12 is applied to carry out parameter estimation using an initialized analysis data set. The three parameters, i.e., the bi-harmonic horizontal diffusion coefficient, the ratio of the transfer coefficient of moisture to the transfer coefficient of sensible heat, and the Asselin filter coefficient, as well as the initial condition, are optimally recovered from the data set using adjoint parameter estimation.

The fields at the end of the assimilation window starting from the retrieved optimal initial conditions and the optimally identified parameter values successfully capture the main features of the analysis fields. A number of experiments are conducted to assess the effect of carrying out 4D VAR assimilation on both the initial conditions and parameters, versus the effect of optimally estimating only the parameters. A positive impact on the ensuing forecasts due to each optimally identified parameter value is observed, while the maximum benefit is obtained from the combined effect of both parameter estimation and initial condition optimization. The results also show that during the ensuing forecasts, the model tends to "lose" the impact of the optimal initial condition first, while the positive impact of the optimally identified parameter values persists beyond 72 hours. Moreover, we notice that their regional impacts are quite different.

1 Introduction

A numerical weather forecast model involves a number of parameters which are determined empirically. Some of these parameters, which are very common in physical parameterization schemes, contain information about the flow's properties and characteristics or originate from the simplification of the physical parameterizations. Other parameters are introduced due to numerical stability considerations. The values of the parameters directly or indirectly impact upon the performance of the model. Generally, the values of some parameters are determined by trial and error, i.e., there is no objective criterion to choose "optimal" values of the parameters. In this study, we will focus on the optimal estimation of several parameters which are known to impact the performance of a numerical weather prediction (NWP) model. Namely, we wish to identify the "optimal" bi-harmonic horizontal diffusion coefficient κ, the ratio γ of the transfer coefficient of moisture to the transfer coefficient of sensible heat, and the Asselin filter coefficient ϵ using adjoint parameter estimation and study how such optimized parameters impact on the ensuing model forecasts.

The parameter estimation procedure is aimed at choosing an "optimal" parameter in an admissible parameter set, so that the model solutions corresponding to this parameter fit the observations as a minimization problem of an output-error criterion (Chavent, 1974). Research on adjoint parameter estimation has been carried out using relatively simple models in the last twenty years, first on topics such as aquifer behavior under transient and steady state conditions in the field of hydrology (Carrera and Neumann, 1986a,b,c), bottom drag coefficient identification in a tidal channel (Panchang and O'Brien, 1989), wind stress coefficient estimation along with the estimation of the oceanic eddy viscosity profile (Yu and O'Brien, 1991), and nudging coefficient estimation in the NMC adiabatic version of the spectral model (Zou, et al., 1992a), et al. The issue of the adjoint parameter estimation was also addressed by LeDimet and Navon (1988). Wergen (1992) recovered both the initial conditions and a set of parameters from observations using a 1-D shallow-water equation model. His results showed that even with noisy observations, the parameters were recovered to an acceptable degree of accuracy. For a detailed survey of the state-of-the-art of parameter estimation in meteorology and oceanography, see Navon (1998). However, very few numerical experiments have been conducted using sophisticated full-physics model, and very little attention has been paid to the impact of the parameter estimation on the the model forecast although the performances of the physical parameterization and/or numerical schemes are also a main factor in determining the model forecast skill. Recently, more and more effort has been focused on improving the initial condition (Thépaut, et al., 1991; Navon, et al., 1992), and its important impact on reducing the forecast error has been demonstrated. What is the effect of the parameter estimation on the model compared to that of the initial condition? In this study, we intend to explore this question, and expect that the performances of the physical parameterization and/or numerical schemes can be improved by tuning the parameters involved.

This paper is organized as follows: the model used and the characteristics of the parameters to be optimally identified are described in section 2. The methodology of adjoint parameter estimation is presented in section 3. The detailed parameter estimation procedure, along with the minimization algorithm and the "optimal" parameter values, are provided in section 4. Section 5 provides the description of the forecast experiments and results using both the "optimal" initial condition and the optimally identified parameter values. Section 6 describes the impact of the optimal parameters alone on ensuing 24h forecasts. The model's "memory" (in terms of ensuing forecast period results) of the impact of either the optimal initial condition or the optimal parameter values is discussed in section 7. Summary and future research are finally presented in section 8.

2 The model and characteristics of the parameters

The full-physics FSU Global Spectral Model (FSU GSM) (Krishnamurti, et al., 1988) and its adjoint model are employed to recover both the optimal initial condition and optimal parameter values from the observations. The FSU GSM has been successfully applied for numerical weather forecasts, especially for the tropics. The model is a multi-level (12 vertical levels) primitive equation model with σ coordinate. All variables are expanded horizontally in a truncated series of spherical harmonic functions (at resolution T42) and the transform technique is applied which makes it possible and efficient to perform the horizontal operators and to calculate the physical processes in real space. The finite difference schemes in the vertical and semi-implicit scheme for time integration are employed. The full physical packages include orography, planetary boundary layer, dry adjustment, large-scale precipitation, moist convection, horizontal diffusion and radiation processes.

The effort to establish a 4D variational (4D VAR) data assimilation system for the FSU GSM was initiated in 1993. The adjoint code of the dry adiabatic version of the FSU GSM was developed by Wang (1993). Later, Tsuyuki (1996) incorporated moisture variables, the smoothed parameterization of moist processes, horizontal diffusion and a simplified surface friction. We continued their effort by incorporating radiation, planetary boundary layer (including vertical diffusion) processes into the 4D VAR data assimilation system due to the important roles they play in simulating various large-scale and mesoscale phenomena, especially in tropical weather systems, thus obtaining the full-physics adjoint model (Zhu and Navon, 1997). This effort renders the adjoint model consistent with the nonlinear forecast model. If the two models (the forward model and its adjoint) are inconsistent, this may imply a negative impact on the process of adjoint optimal parameter estimation.

The horizontal diffusion term is usually incorporated in a numerical weather prediction model to parameterize the effects of motions on the unresolved scales and to inhibit spectral blocking, i.e., the growth of small scales in the dynamic model variables due to the accumulation of energy at high wave numbers. It is also employed to eliminate the aliasing effect (Phillips 1959, and Hamming 1973). The presence of any dissipation, physical or computational, can attenuate the amplitude of the short wavelengths very significantly; In this case, the errors introduced by aliasing are minimal. MacVean's study (1983) showed that, without dissipation, integrations exhibited grossly physically unrealistic features after about several days even with very high mesh resolution, indicating the crucial role played by the dissipation in non-linear baroclinic development. Considerable effort has been made by various groups in tuning the dissipation parameterizations in their general circulation or forecast models. For instance, Phillips (1956), Lilly (1965), Leith (1965) and Richard (1965) etc. applied the eddy diffusion terms in their numerical models; Navon (1969) included the lateral viscosity in a two level general circulation model and carried out a 62 day integration, comparing its impact to that of using the Matsuno dissipative scheme. Kanamitsu, et al. (1983, 1989), Gordon, et al. (1982) employed a bi-harmonic horizontal diffusion in their experiments. Presently, the bi-harmonic horizontal diffusion is used worldwide in NWP models due to its better scale selectivity. Its limitation is the lack of a sound physical foundation. Some other horizontal diffusion schemes are also used, such as the scheme developed by Leith (1971) based on turbulence theory which is now implemented in the NCEP (NMC) spectral model (Kanamitsu, et al., 1991).

The bi-harmonic horizontal diffusion $\kappa\ \nabla^4$ is used for vorticity, divergence, dewpoint depression and virtual temperature in the FSU GSM to selectively control small-scale noise without affecting large scales. The e-folding diffusive decay time at total wavenumber n is given by

$$\tau(n) = \frac{1}{\kappa}\,\frac{a^4}{n^2(n+1)^2} \tag{1}$$

where a is the radius of the earth. The physical significance of the diffusion coefficient κ is not directly intuitive. Its effect may be understood in terms of the time scale τ at the smallest spatial scale resolved. The model dissipation should remove energy from the end of the spectrum at a rate sufficient to prevent a spurious accumulation of energy there, while not affecting the medium and large scales. In the GFDL spectral model, the coefficients of the eddy diffusion for ∇^4 are determined by trial and error, using the quality of the medium range 500 mb geopotential height forecast as a criterion (Gorden, et al. 1982). The value they used is $1\times10^{16}m^2s^{-1}$ or $2.5\times10^{16}m^2s^{-1}$ for T30, i.e., a value of τ of about 53 or 21 hours (based on Equation 1). The value used in FSU Global Spectral Model T42 is $6\times10^{15}m^2s^{-1}$, i.e., yielding a value of τ corresponding to about 23 hours.

In addition to the spatial diffusion, filters are also commonly used in numerical models to remove high frequency noise that cannot be resolved at the given model resolution. In the FSU GSM, a supplementary time filter of the form

$$\hat{F}(t) = F(t) + \epsilon\,[\hat{F}(t-1) - 2F(t) + F(t+1)]$$

is used. The characteristics of this filter are described in detail by Asselin (1972). With an adequately chosen coefficient ϵ, this filter damps the spurious computational frequencies and a significant part of the spectrum of the external and internal gravity waves. The Rossby motions of comparable horizontal dimensions are much less affected due to their low frequency. This filter with $\epsilon = 0.25$ removes or filters $2\Delta t$ waves and reduces the amplitude of

$4\Delta t$ waves by one-half, but has little effect on longer-period waves, i.e. it acts as a low-pass filter in time. However, despite the advantage of immediately suppressing the $2\Delta t$-wave, values of ϵ less than 0.25 are preferable since the stability condition requires a progressively smaller Δt as ϵ increases. Moreover, repeated use of even a weak filter eventually dampens the lower frequencies (Haltiner and Williams 1980). Skamarock and Klemp (1992) discussed how the Asselin filter affected stability for a nonhydrostatic model. Robert (1966) used this filter in a general circulation primitive-equation spectral model with centered differences to control instability. The values he used are $\epsilon = 0.02$ and $\Delta t = 20min$. A number of experiments carried out by Krishnamurti showed that $\epsilon = 0.05$ is the best value for the Asselin filter for the model forecast. We will carry out an optimal parameter estimation experiment to obtain the optimal value of ϵ in this study. The lower bound of ϵ is set to be zero, while its upper bound is set to be 0.1.

The third parameter we consider is the ratio γ of the transfer coefficient of moisture to the transfer coefficient of sensible heat, which arises from the parameterization of the surface fluxes in the boundary layer. Accurate heat and moisture flux parameterizations are very important since the surface heat and moisture fluxes exert a strong influence on the surface energy budget and precipitation rate. Similarity theory has successfully provided a framework for the description of the atmosphere surface layer, the lowest 50 m or so of the boundary layer where the Coriolis force can be ignored and the fluxes can be assumed to be constant with height. The flux profiles and other properties of the flow are reasonably parameterized via this theory. However, certain empirical parameters or constants evolving from the theory need to be determined by experiments, for instance, the von Kármán constant and constants associated with the stability dependence of the flux-profile formulations. In this study, however, our goal is not to determine the parameter values from physical soundness. We will instead aim to obtain an optimal value of γ for this particular model via a variational parameter identification approach in order to improve the model forecast skill. Dyer (1967) found that, over the range (of the ratio of height z to the Monin-Obukhov length L) $0.02 < |z/L| < 0.6$ both the transfer coefficient of sensible heat ϕ_H and the transfer coefficient of moisture ϕ_Q varied approximately as $|z/L|^{-\frac{1}{3}}$; For $|z/L| > 0.2$, ϕ_H is found to vary as $|z/L|^{-\frac{1}{2}}$, but insufficient data limited the value of the corresponding analysis for ϕ_Q. Since reasonable agreement is found between the ϕ_H and ϕ_Q data, many numerical weather prediction models adopt a simple relationship of the type $\phi_Q = \phi_H$, i.e., $\gamma = 1.0$. Recently, Smith et al. (1996) reviewed the 25 years of progress on air-sea fluxes. Theory and observations indicate that the ratio should depend on the surface characteristics. However, since the ratio is considered as a constant in the original FSU GSM and also numerical experience with the model shows that the model tends to produce larger precipitation, we will still take the parameter as a constant and set its upper and lower bounds to be 1.8 and 0.3 respectively in this study. Its optimal value will be estimated via numerical fitting of the model to the observations. The specification of the γ value directly impacts the moisture flux. We expect to improve the performance of its corresponding physical process by tuning this parameter.

3 The methodology of parameter estimation

In this study, we aim to perform optimal parameter estimation via a variational approach, i.e. to obtain an optimal value α^o of the parameter vector α such that

$$J(\alpha^o) < J(\alpha) \;\; \forall \alpha$$

where J is a cost function which measures the discrepancy between the observations and the corresponding model forecast variables. Hence, the optimal parameter can be retrieved by fitting the model forecast fields to the observations.

Given constrained parameters whose values vary between certain bounds, for instance, when the parameter α satisfies $\alpha \in [a, b]$, where a and b denote the lower and upper bounds, respectively, the cost function for parameter estimation may assume the following form:

$$J(\mathbf{X}, \alpha) = \frac{1}{2} \int_{t_0}^{t_R} < W(\mathbf{X} - \mathbf{X}^{obs}), (\mathbf{X} - \mathbf{X}^{obs}) > dt + \lambda^T \; g(\alpha) \tag{2}$$

where λ is the penalty coefficient vector, W is the diagonal weighting matrix defined as in Navon, et al., 1992, $< >$ denotes spatial integration, $\mathbf{X}$ represents the state variable vector, $\mathbf{X}^{obs}$ the observation vector, and t_0 and t_R denote the assimilation window. The second term consists of a penalty function. The value of λ is determined such that the penalty term is of the same order of magnitude as the other terms in the cost function, and $g(\alpha)$ is defined as:

$$g(\alpha) = \begin{cases} \frac{1}{2}\,(\alpha - b)^2 & \text{if } \alpha \geq b \\ 0 & \text{if } a < \alpha < b \\ \frac{1}{2}\,(\alpha - a)^2 & \text{if } \alpha \leq a \end{cases} \tag{3}$$

where $g(\alpha)$ is a function only of the violated constraints. The first derivative of this function is:

$$\frac{\partial g}{\partial \alpha} = \begin{cases} \alpha - b & \text{if } \alpha \geq b \\ 0 & \text{if } a < \alpha < b \\ \alpha - a & \text{if } \alpha \leq a \end{cases} \tag{4}$$

Another type of penalty effective in transforming a constrained optimization problem into an unconstrained one is the barrier method, which imposes a penalty for reaching the bound of an inequality constraint. Typically, a logarithmic barrier function is of the form:

$$J(\mathbf{X}, \alpha) = \frac{1}{2} \int_{t_0}^{t_R} < W(\mathbf{X} - \mathbf{X}^{obs}), (\mathbf{X} - \mathbf{X}^{obs}) > dt - \mu^T \; \log\left(h(\alpha)\right) \tag{5}$$

where μ is the barrier coefficient and h is the constraint function. The barrier methods are strictly feasible methods, i.e., the iterates lie in the interior of the feasible region and create a "barrier" to keep iterates away from boundaries of the feasible region (Nash and Sofer, 1996).

There are two different purposes for the inclusion of the second term in Equations 2 or 5. One is to ensure that the retrieved parameter lies within the prescribed bounds. A penalty term of the form in Equation 2 or a logarithm of the box constraints as in Equation 5 is efficient to achieve this. The other purpose is to increase the convexity of the cost function by adding a positive value λ to the Hessian matrix, thereby increasing its positive-definiteness.

Suppose that the forward model is perfect and is given in the form

$$\frac{\partial \mathbf{X}}{\partial t} = \mathbf{F}(\mathbf{X}, \alpha, t) \tag{6}$$

Its corresponding tangent linear model is defined as

$$\frac{\partial \delta \mathbf{X}}{\partial t} = \left(\frac{\partial F(\mathbf{X}, \alpha, t)}{\partial \mathbf{X}} \right) \delta \mathbf{X} + \left(\frac{\partial F(\mathbf{X}, \alpha, t)}{\partial \alpha} \right) \delta \alpha \tag{7}$$

Then the adjoint model can be expressed in the form

$$-\frac{\partial P}{\partial t} - \left(\frac{\partial F(\mathbf{X}, \alpha, t)}{\partial \mathbf{X}} \right)^T P = W(\mathbf{X} - \mathbf{X}^{obs}) \tag{8}$$

where P represents the adjoint variables. The gradients of the cost function with respect to the initial condition and the parameter α are, respectively,

$$\nabla_{\mathbf{X}_0} J = P(0) \tag{9}$$

$$\nabla_{\alpha} J = \int_{t_0}^{t_R} \left[(\frac{\partial F}{\partial \alpha})^T P \right] dt + \lambda^T \frac{\partial g}{\partial \alpha} \tag{10}$$

The adjoint model is of the same form as that where only the initial condition is considered as the control variable. Hence, the problem of parameter estimation via the adjoint method does not result in an additional computational effort when the number of parameters to be estimated is small. We may expect that the parameter estimation process will provide us with both optimally-determined parameters and initial condition simultaneously. The gradient of the cost function with respect to both the initial condition and parameters is written as:

$$\nabla J = (\nabla_{\mathbf{X}_0} J, \nabla_{\alpha} J)^T \tag{11}$$

4 Parameter estimation procedure and results

In this study, the FSU Global Spectral Model and its full-physics adjoint model are employed to optimally identify the aforementioned three parameters separately along with initial condition retrieval. An arbitrary initialized ECMWF analysis dataset is used in our study. A 6h assimilation window is used, i.e., from 0000UTC to 0600UTC 3 Sept. 1996. The initialized ECMWF analysis data at the beginning and end of the assimilation window are taken as the observations during the 6h assimilation window period.

The parameter estimation procedure is carried out as follows:

- 1. The 6h forecast starting from the initialized analysis at 1800UTC 2 Sept. 1996 is taken as the initial guess of the initial condition. Given a positive definite initial approximation to the inverse Hessian matrix H_0 (generally taken as the identity matrix I), we integrate the full-physics FSU GSM 6 hours and then calculate the cost function using Equation 2. The initial guess of α, the penalty coefficient value λ and the upper and lower bounds where the parameters may vary are specified here.

- 2. Integrate the full-physics adjoint model of the FSU GSM backward in time with null initial condition to obtain the gradient of J with respect to the control variable $\mathbf{Y} = (\mathbf{X}_0, \alpha)^T$,

$$g_0 = g(\mathbf{Y}) = (\nabla_{\mathbf{X}_0} J, \nabla_{\alpha} J)^T$$

and the search direction

$$d_0 = -H_0 g_0$$

Generally, physical processes involve many on-off switches which may introduce discontinuities in the parameterization schemes. In the full-physics adjoint model of the FSU GSM, since we hope to keep the original parameterization characteristics unchanged as much as possible, only the discontinuities which impact the tangent linear approximation and the convergence of the minimization algorithm most are considered for removal by using smooth function and cubic spline interpolation (Zhu and Navon, 1997).

- 3. For k=0,1,2,..., minimize $J(\mathbf{Y}_k + \beta_k d_k)$ with respect to $\beta \geq 0$ to obtain $\mathbf{Y}_{k+1}$ as

$$\mathbf{Y}_{k+1} = \mathbf{Y}_k + \beta_k d_k$$

where β_k is a positive scalar, the step-size being obtained by a line search so as to satisfy a sufficient rate of decrease (see Gill, et al., 1981).

- 4. Compute the new gradient and search direction, respectively

$$g_{k+1} = \nabla J(\mathbf{Y}_{k+1})$$

$$d_{k+1} = -H_{k+1} g_{k+1}$$

- 5. Check whether the solution converges. If the convergence criterion

$$\| g_{k+1} \| \leq \epsilon' \; Max \; (1, \| \mathbf{Y}_{k+1} \|)$$

is satisfied, where ϵ' is a user supplied small number, then the algorithm terminates with $\mathbf{Y}_{k+1}$ as the optimal solution; otherwise go back to Step 3.

The 6h forecast from the initialized analysis at 1800UTC 2 Sept. 1996, which serves as the initial guess of the initial condition, shows important underestimates of the tropical divergence field, particularly in the Pacific (Fig. 1).

[Fig. 1 near here, please]

The minimization procedure is terminated after 60 iterations. Although the convergence rate is expected to be much slower than that of the adiabatic version, due to the high nonlinearity of the model including the full physical processes, a sufficient decrease in the cost function and its gradient for each experiment is achieved after 60 iterations. At the end of the minimization procedure, both the optimal initial condition and optimal value of the parameter are recovered.

The initial guess (estimated value) of the horizontal diffusion coefficient κ is taken to be 6×10^{15} as used routinely in the FSU GSM T42L12. The penalty parameter λ for this parameter is taken to be 1.0×10^{-12}. The upper and lower bounds where κ varies are taken to be 3.0×10^{16}, i.e., five times the value used in the original forecast model, and 2.0×10^{15}, respectively. The variations of the cost function and the gradient norm with respect to the number of iterations are presented in Fig. 2. We see that the cost function decreases from an initial value of 3643 to 575.3, i.e., it decreases to about 15.8% of its original value, while the norm of the gradient value decreases to 23.6% of its original value. Fig. 3 displays the evolution of the horizontal diffusion coefficient during the minimization procedure. This coefficient experiences a rapid increase until the 35-th iteration, reaching its peak at the 44-th iteration, then experiences a slight decrease. The optimal value obtained for the horizontal diffusion coefficient is 1.1124×10^{16}. This value is almost twice the value used in the original forecast model. We also notice that the optimal parameter value is not very sensitive to the value of λ as long as the penalty term is of the same order of magnitude as the other terms in the cost function.

The issue of the sensitivity of the optimal parameters values to the values of the penalty coefficients deserves some attention. Research work by Craven and Wahba (1979) and by Lewis and Grayson (1972) has shown strong dependence of the analysis results on the specification of the penalty coefficients.

In our case the optimal parameter values are not found to be very sensitive to the value of the penalty coefficient as long as the penalty parameter attains a certain threshold value.

To understand why this happens one should consider the basics of the penalty technique as a method to solve a sequence of unconstrained optimization problems whose solution is usually infeasible to the original constrained problem. A penalty for violation of the constraints is incurred and as this penalty is increased, the iterates are forced towards the feasible region.

The experience of Zou et al. (1992b), Zou et al. (1993) in applying a penalty method for controlling gravity oscillations in variational data assimilation shows that the effect of the penalty parameter is ineffective until it attains a threshold value. Thereafter it is effective in constraining the minimizers toward the feasible region and filtering the gravity waves. Increasing the penalty parameters by an order of magnitude or more over the threshold value does not change this effect, but slows down the minimization due to the ill-conditioning of the related Hessian of the cost functional (see Nash and Sofer, 1996; Gill at al. 1981; Fletcher 1987)

A higher sensitivity of minimization results to the value of the penalty parameter is typical for cases when from the start the iterates are rather far from the feasible region and reaching the effective threshold value requires several iterations.

In the present case we are from the starting point very near to the feasible region -

hence the perceived lack of sensitivity of the optimal parameter values to the values of the penalty coefficients when these attain their threshold value. When we increase the penalty parameters further (i.e. by an order of magnitude or more) there is no perceived change except for a serious slow-down in the rate of convergence of the minimization due to the aforementioned ill-conditioning.

Similar variations of the cost function and the gradient norm with respect to the iteration number are obtained when the ratio γ and the Asselin filter coefficient ϵ are retrieved from the observations, respectively. The initial guess (estimated value) of the Asselin filter coefficient ϵ is set to be 0.05, which is the value used in the original forecast model. The upper and lower bounds for the variation of this parameter are taken as 0.1 and 0.0, respectively, and the penalty parameter λ for this parameter is set to be 1.0×10^8. The optimal ϵ value obtained is 0.0487 which is very close to the initial guess, which means that either the initial guess is a fairly good guess or the model is not very sensitive to the parameter. Further study with the initial guess away from 0.05 still yields the same result, which indicates that the initial guess is fairly good. The initial guess (estimated value) of the ratio γ of the transfer coefficient of moisture to the transfer coefficient of sensible heat is initially set to be 1.0 which is the value used in the original FSU GSM T42L12. The penalty parameter λ for γ is set to be 2.0×10^5, while the upper and lower bounds where γ may vary are taken to be 1.8 and 0.3, respectively. The optimal parameter estimation for γ yields a value of 0.4974, which is only half of the initial guess value, due to numerical fitting of the model to the observations. We should be aware that the optimal value retrieved here includes the possibility of the compensating for errors in the physical parameterization schemes of the forecast model. It doesn't represent the true physical value, it is only the optimal value for this particular model for improving the model forecast.

[Figs. 2 & 3 near here, please]

5 Forecast experiments using both the retrieved initial condition and parameters

5.1 Assessing the combined impact of the optimal initial condition and parameter estimation

Three experiments are carried out to integrate the model for 6 hours in order to obtain the analysis fields at the end of the assimilation window for the bi-harmonic horizontal diffusion coefficient κ, namely: 1) a control experiment (i.e., experiment C1 in Table 1) which is integrated from the initial guess fields, i.e., 6h forecast from the 1800UTC 2 Sept. 1996 analysis, and uses the estimated parameter $\kappa = 6 \times 10^{15}$, 2) the "optimal experiment" (i.e., experiment O2 in Table 1), so called since we are using both the optimal initial condition and the optimal parameter value κ, and 3) a simulation experiment using the estimated parameter $\kappa = 6 \times 10^{15}$ where the model is integrated from the initialized 0000UTC 3 Sept. 1996 analysis to serve as the best simulation without using variational data assimilation. The differences between the control experiment and the optimal experiment reflect the combined impact of the optimal initial condition and the optimal bi-harmonic horizontal diffusion coefficient parameter value.

The RMS errors of the fields at the end of the assimilation window are calculated for the aforementioned experiments and are provided in Fig. 4. The results obtained show that the optimal experiment yields the best results throughout all of the vertical levels, especially for the divergence field. The RMS errors of the logarithm of the surface pressure are 0.140E-

2, 0.939E-3, and 0.517E-3 for the control experiment, the simulation experiment and the optimal experiment, respectively.

[Figs. 4 & 5 near here, please]

Fig. 5 presents the divergence analysis fields at 200mb for the aforementioned experiments and the observation at the end of the assimilation, i.e., 0600UTC 3 Sept. 1996. Only the optimal experiment successfully provides a high quality analysis, capturing the main features of the divergence field; Both the control experiment and the simulation experiment fail to simulate the strong divergence field corresponding to a heavy precipitation event over Indonesia and overestimate the divergence field to the west of Africa. The control experiment also overestimates the divergence field over the northern Africa and underestimates the divergence field around 35°S, 20°W, while the simulation experiment overestimates the divergence field to the east of Australia.

The difference fields of the specific humidity at 500 mb and of the temperature at 850 mb between the results of the aforementioned experiments and the observation field at the end of the assimilation are displayed in Figures 6 – 7. The improvement obtained from the optimal experiment over the results obtained from the other experiments is obvious, particularly over areas with large errors.

[Figs. 6 & 7 near here, please]

Fig. 8 displays the RMS errors of the meteorological variables for the 24h forecast control experiment and optimal experiment (i.e., experiment O4 in Table 1) for the bi-harmonic horizontal diffusion coefficient κ. The use of both the optimal initial condition and the parameter value improves the 24h forecast fields, however, compared with Fig. 4, the differences between the control experiment and the optimal experiment decrease as the forecast period increases.

[Fig. 8 near here, please]

Similar results are also obtained for the other two parameters.

5.2 Assessing the impact of optimal parameter estimation

In the previous sections, we recovered three pairs of the optimal initial conditions and optimal values of several model parameters, and discussed their combined impact on the model forecast. An important question remains to be clarified: how much of the improvement obtained is directly attributable to the optimal initial condition, and how much of it originates from the optimal values of the identified parameters. In order to provide a closer look at this issue, three additional experiments are performed to compare with the above-mentioned 6h forecast control experiment (referred to as C1) and the optimal experiment (referred to as O2) as well as with the 24h forecast optimal experiment (referred to as O4) for κ. The additional experiments are: experiment C2 where the initial guess of the initial condition and the optimally identified parameter value are used, experiment O1 where the optimal initial condition and estimated parameter value are used, and experiment O3 where the optimal initial condition and estimated parameter value are used (Table 1).

[TABLE 1 near here, please]

To assess the impact of the optimal parameter estimation, we use the differences between the RMS errors at all vertical levels for vorticity, divergence, virtual temperature and dew-point depression fields. The difference between experiments C1 and C2 reflects the impact of the optimal parameter values on the forecast when the optimal initial condition is not used; While the differences between experiments O1 and O2, and between experiments O3 and O4 reflect the impacts of the optimal parameter value on the forecast when the optimal initial condition is applied for the 6h and 24h forecasts, respectively. If the difference is negative, it means that the RMS error of the experiment where the optimally identified parameter value is used is smaller than the RMS error of the experiment where the estimated parameter value is used. The differences of the RMS errors of the vorticity, divergence, virtual temperature and the dewpoint depression at the end of the forecast between the experiment pairs C2 and C1, O2 and O1, O4 and O3 are shown in Fig. 9. For the pair C2 and C1, the RMS differences are negative except for the dewpoint depression at the lowest two levels. This indicates a relatively small (compared to the differences of the RMS errors between C1 and O2) but positive impact of the optimal parameter value on the model simulation. Experiment O1 is comparable to O2 in terms of RMS error, with the impact of the optimal initial condition dominating that of the optimal parameter value at the end of the assimilation window. This is very reasonable, since in the first few hours of the forecast, the optimal initial condition may reconstruct the dynamic structure while the optimal bi-harmonic horizontal diffusion coefficient value can only adjust it. However, the lower level divergence fields are largely improved in all of the pairs of experiments when the optimal parameter value κ is used. As the length of the forecast period increases, the impact of the optimal parameter value on the model forecast becomes more pronounced for the pair O3 and O4. Compared with Fig.8, this result implies the known fact that the effect of the optimal initial condition decays as the forecast period becomes longer.

[Fig. 9 near here, please]

An additional experiment is conducted with the fixed initial condition 0000UTC 3 Sept. 1996 and where the parameter κ is considered as the sole control variable. An identical cost function and initial parameter value to those used before are employed. The cost function reaches a constant value while the norm of gradient decreases by three orders of magnitude in five iterations. The optimal bi-harmonic horizontal diffusion coefficient value obtained is

1.1946×10^{16}, which is very close to that obtained when we recovered both the optimal initial condition and the optimal parameter value. A similar positive impact is observed when comparing the forecasts using the estimated parameter value and the optimal parameter value, respectively.

Similar experiments (related to the experiments detailed in Table 1) are also conducted separately for the parameters ϵ and γ. Since the optimal value of Asselin filter coefficient ϵ is very close to its estimated value, the impact of this parameter on the ensuing forecasts is marginal. A positive impact of the optimal ratio γ on the forecasts is also observed while the impact of the optimal initial condition is found to be dominant during the first several hours of the forecast.

6 Impact of the optimal parameters alone on ensuing 24h forecasts

A number of 24h forecast experiments beginning from the 0000UTC 3 Sept. 1996 analysis are performed in order to examine the impact of each parameter separately as well as the combined impact of all the three parameters on the ensuing forecast fields. The experiments are: (a) experiment with the estimated parameters, i.e., $\kappa = 6.0 \times 10^{15}$, $\epsilon = 0.05$ and $\gamma = 1.0$; (b) experiment with the estimated ϵ and γ but the optimally retrieved κ value; (c) experiment with the estimated ϵ and κ but the optimally retrieved γ; (d) experiment using optimally retrieved κ, ϵ and γ simultaneously. Since the optimal value of ϵ is very close to its estimated value, the impact of this parameter is rather marginal.

The RMS errors of the vorticity, divergence, dewpoint depression and the virtual temperature fields at all vertical levels are calculated between the above experiments and the analysis at 0000UTC 4 Sept. 1996. Fig. 10 displays the differences of the RMS errors between experiments (b) and (a), (c) and (a), (d) and (a), which indicate the impacts of optimal values of κ, γ, and their combined impact, respectively. Negative values of the RMS differences indicate that the RMS errors are less than those of the experiment (a) with the estimated parameter values, i.e., the optimal parameter value has a positive impact on the forecast.

[Fig. 10 near here, please]

The results show that the optimal horizontal diffusion coefficient κ impacts mainly the vorticity and divergence fields, and also has a positive impact on the virtual temperature and dewpoint depression at upper and middle levels but a negative impact at lower levels. On the contrary, in experiment (c) when only the optimal value of γ is employed, the vorticity and divergence fields are only slightly improved, but the middle and lower levels of the virtual temperature and the lower levels of the dewpoint depression experience a large improvement. The effect of γ is mainly confined to the lower levels of the model. The best forecasts are obtained by experiment (d) where all the optimal parameter values of κ and ϵ as well as γ are used simultaneously. In this sense, experiment (d) combines all of the advantages of experiments (b) and (c).

7 A study of the model's "memory" of impact of optimal initial condition and identified parameter values

So far, we have studied either the combined impact of both the optimal initial conditions and the optimal parameter values or the impact of only the optimally identified parameter values on the ensuing forecast. In this section, we discuss the persistence (memory) of the combined impact of the above-mentioned three optimally identified parameter values as well as the optimal initial condition obtained by variational data assimilation on the ensuing forecast.

[TABLE 2 near here, please]

Three sets of experiments are carried out by integrating the model for 24h, 48h and 72h respectively from 0000UTC 3 Sept. 1996. The first set of experiments (C1, C2 and C3) are control experiments which are integrated from the initial guess of the initial condition (i.e., the 6h forecast from 1800UTC 2 Sept. 1996) using the estimated parameter values. The second set of "optimal parameter" experiments (P1, P2 and P3) start from the initial guess of the initial condition using the optimally identified values of the above-mentioned three parameters. The third set of "optimal initial condition" experiments (I1, I2 and I3) are integrated from the variationally derived optimal initial condition using estimated parameter values (Table 2).

[Fig. 11 near here, please]

The RMS errors of vorticity, divergence, virtual temperature and dewpoint depression fields are calculated for each experiment, then the percentages of the differences of the RMS errors between C1 and P1, C2 and P2, C3 and P3, C1 and I1, C2 and I2 and C3 and I3 are computed. The percentages of the decrease of RMS errors due to the use of the three optimally identified parameter values are displayed in Fig.11. The results show that all of the experiments using optimally identified parameter values exhibit smaller RMS errors than those using estimated parameter values through all of the vertical levels except for the virtual temperature at the top vertical level of the model, i.e., the combined impact of the three identified parameter values still persists for the 72h forecast, and probably even further beyond. The largest improvement in the 24h forecast occurs at the low levels of the dewpoint depression field where the RMS error decreases by up to 17%. The overall forecast RMS errors further decrease for the 48h forecast. The impacts of the optimally identified parameter values on the vorticity field and on the middle levels of the divergence, virtual temperature fields as well as the lower levels of the dewpoint depression field decay for the 72h forecast compared to the results obtained with optimal parameters for the corresponding 48h forecast, however, the RMS errors still decrease by up to 10% compared to the experiment where estimated parameters are used for the same 72h forecast.

[Fig. 12 near here, please]

The percentages of the decrease in RMS errors for vorticity, divergence, virtual temperature and dewpoint depression fields by using optimal initial condition are presented in Figure 12. A very clear trend is observed, namely, the impact of the optimal initial condition decays as the forecast time increases, especially the impact on the divergence field decays very rapidly. This is in agreement with the results obtained in section 4. The improvement of the forecast (in terms of RMS errors) due to the combined impact of the three optimally identified parameter values exceeds that obtained due to the impact of the optimal initial condition in the ensuing 72h forecast. Even for the 24h forecast, the impact of the three optimal parameter values on the lower level of the dewpoint depression is larger than that of the optimal initial condition. The model tends to "lose" first the impact of the optimal initial condition in the ensuing forecast, while the impact of using optimal parameters on above-mentioned forecast fields lingers even after 72 hours.

Figures 13 and 14 show how the RMS errors of vorticity and dewpoint depression evolve with forecast time, respectively, in which black bars are for the optimal parameter experiments (P1, P2, and P3) and grey bars are for the optimal initial condition experiments (I1, I2, and I3). The results indicate that the errors of the optimal initial condition experiment become larger than those of the optimal parameter experiment at 72 hour forecast.

[Figs. 13 - 14 near here, please]

[Figs. 15 - 17 near here, please]

We also would like to point out that the mechanisms of the impacts resulting from optimal initial condition and from the optimal parameter values are quite different. The impact of the optimally identified parameter values is effective throughout the entire integration period via the corresponding physical parameterization or numerical schemes, while the impact of the optimal initial condition obtained via variational data assimilation is to enhance the model forecast skill by reducing the errors in the initial condition which might lead to poor forecast. Therefore, their impacts are quite different. Fig. 15 displays the difference fields between experiments C1, P1 and the analysis for specific humidity at 850 mb, respectively, while the difference field of specific humidity at 850 mb between experiments C1 and P1 is displayed in Fig. 16. Comparing Fig. 15 (a) and (b), we notice that they have almost the same pattern, except that the positive differences between the forecast and the analysis become smaller while the negative differences between the forecast and the analysis become larger over most of the forecast area in experiment P1 when the optimally identified parameter values are used. That is to say, the use of the optimally identified parameter values tends to produce a smaller specific humidity forecast over most of the global domain, since the optimal ratio of the transfer coefficient of the moisture to the transfer coefficient of the sensible heat is only about half of its estimated value. Larger specific humidity forecasts in experiment P1 than those in experiment C1 are observed only in a few regions, such as central South America, which might have been caused by the impact of the optimal bi-harmonic horizontal diffusion coefficient or by interactions between the impacts of the three optimal parameter values. In this case, the improvement due to the use of the optimally identified parameter values is observed mainly over the overestimated areas with respect to the analysis (in experiment C1) which spread over the land. Fig. 17 displays the difference fields of the specific humidity at 500 mb between experiments C1, I1 and the analysis, respectively. Experiment I1 where the optimal initial condition is used, however, does not exhibit such features as mentioned above, rather the improvement is obtained both in underestimated and overestimated areas compared with the result of experiment C1, especially over areas of large errors.

It is also known that forecasts starting from the variationally derived optimal initial conditions are not as good as the forecasts starting from the latest available analysis (Pu, et al., 1997). However, we compare the forecasts starting from 0600UTC 3 Sept. 1996 analysis (the latest available analysis in this study) using both the estimated values of the above-mentioned three parameters and the previously optimally identified parameter values, respectively. Similar improvements as shown in Fig. 11 are observed. The optimally identified values of the bi-harmonic horizontal diffusion coefficient and the ratio of the transfer coefficient of the moisture to the transfer coefficient of the sensible heat actually improve the performance of the corresponding physical parameterization and/or numerical schemes, so that their impacts on the model forecast persist under similar large-scale atmospheric environment.

[Fig. 18 near here, please]

Finally, we present the anomaly correlation results for the specific humidity forecasts at 850 mb for the control experiments, "optimal parameter" experiments, and "optimal initial condition" experiments as well as the experiments using both optimal initial conditions and optimally identified parameter values. These anomaly correlations are measured both for the tropical belt (defined here to be 40°S - 40°N where physical initialization is usually carried out (Krishnamurti, 1991)) and for the Northern Hemisphere (Fig. 18). The specific humidity field, which is mainly distributed in the lower troposphere, is very important in the tropical system. Comparing Fig. 18 (a) and (b), we see that the anomaly correlation for the Northern Hemisphere consistently is higher than that for the tropical belt for the entire 5-day integration period. The results of "optimal initial condition" experiments appear to be more skillful than the "optimal parameter" experiments for the first 24h forecast in the tropical belt and for the first 48h forecasts in Northern Hemisphere, respectively. However, later in the forecast (i.e., for the period 48 - 120 h), the "optimal parameter" experiments yield better skill results than the "optimal initial condition" experiments both in the tropical belt and Northern Hemisphere. The experiments using the optimal initial condition and optimally identified parameter values simultaneously have the highest anomaly correlations, where for the 120-h forecast, the anomaly correlation increases from 65.1% in the control experiment to 70.7% for the tropical belt and from 72.5% in the control experiment to 77.8% for the Northern Hemisphere.

8 Summary and future research

In this study, we recovered both optimal initial conditions and optimal values of the biharmonic horizontal diffusion coefficient, the Asselin filter coefficient and the ratio of the transfer coefficient of moisture to the transfer coefficient of sensible heat using the full-physics adjoint of the FSU Global Spectral Model. The results obtained are very encouraging. The fields at the end of the assimilation window starting from the retrieved optimal initial condition and the optimal parameter values successfully capture the main features of the analysis fields. Although the impact of optimal initial conditions dominates that of the optimal parameter values at early stages of the forecast, a positive impact due to each optimally estimated parameter value is observed. The ensuing 24h forecast experiments where only the optimal parameter values are used further indicate the positive impact of using optimal parameter values. The bi-harmonic horizontal diffusion coefficient improves vorticity and divergence fields as well as the upper levels of virtual temperature and of dewpoint depression fields, while the ratio of the transfer coefficient of moisture to the transfer coefficient of sensible heat has a large positive impact on the lower levels of the model, especially those of the virtual temperature and the dewpoint depression fields. By combining the three optimal parameter values, we obtain the best forecast results. Further studies of ensuing forecasts using the optimally identified parameter values and the optimal initial condition respectively show that the model tends to "lose" first the impact of the optimal initial condition while the impact of optimally identified parameter values persists beyond 72 hours. The optimally identified values of the bi-harmonic horizontal diffusion coefficient and the ratio of the transfer coefficient of the moisture to the transfer coefficient of the sensible heat improve the performance of the corresponding physical parameterization schemes. The experiments using the optimal initial condition and optimally identified parameter values simultaneously yield the best performance. However, we should be aware that these results are obtained for a single case study. Further studies should be conducted for different initial conditions in order to draw a general conclusion.

In this study, the nonlinear forecast model is assumed to be perfect, i.e., it is used as a strong constraint. Further studies should be conducted to take into account the model forecast error, i.e., the forecast model is regarded as a weak constraint, and its effect on the retrieved initial conditions and optimal parameter values should be examined.

The parameters studied in this paper are assumed to be constant in both time and space. In future research, we should address the issue of retrieving values of κ and γ that vary both in space and time. In particular, we would like to focus on the values of γ for different stages of a tropical system and for different regions, and study their impact on the forecasts, especially on the precipitation field. Moreover, for a parameter estimation to be properly specified, the parameter's seasonal variation should also be taken into account. The issue of the period of validity of optimally estimated parameters and the frequency with which they should be refreshed in an operational model would necessitate further studies. Another question that deserves further study is the issue of the feedbacks between the effects of various optimally identified parameters.

Acknowledgments

Special thanks go to Dr. T. N. Krishnamurti for his encouragement and insightful comments throughout this study. The first author also wishes to thank Drs. H. S. Bedi, X.-L. Zou and W. Han for their stimulating discussions. Thanks are also due to Dr. W. Han for providing all the data used in this study and Dr. H. S. Bedi for providing the anomaly correlation subroutine. The second author would like to acknowledge private communication with Dr. J. Sela (NCEP) who suggested using bi-harmonic horizontal diffusion coefficients for parameter identification. Stimulating discussions with Dr. Eugenia Kalnay on the topic of parameter estimation are also acknowledged. The authors also wish to thank the two anonymous reviewers for their useful and stimulating comments and suggestions.

The support provided by NSF grant ATM-9413050 managed by Dr. Pamela Stephens is gratefully acknowledged. This work is part of the satisfying the requirements of a Ph.D. Thesis directed by Prof. I.M. Navon and supported fully by the above grant. Additional support is provided by Supercomputer Computations Research Institute at Florida State

University which is partially funded through contract DE-FC0583ER250000.

The computer facilities are provided by the Special SCD Grant 35111089 in NCAR and National Science Foundation 35111100. Additional computer support is provided by Supercomputer Computations Research Institute at Florida State University.

References

Asselin, R., 1972: Frequency filter for time integrations. *Mon. Wea. Rev.*, **100**, 487-490.

Carrera, J., and S.P. Neumann, 1986a: Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information, *Water Resources Research*, **22**, No. 2, 199-210.

Carrera, J., and S.P. Neumann, 1986b: Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability and solution algorithms, *Water Resources Research*, **22**, No. 2, 211-227.

Carrera, J., and S.P. Neumann, 1986c: Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data, *Water Resources Research*, **22**, No. 2, 228-242.

Chavent, G., 1974: Identification of functional parameters in partial differential equations. In *Identification of parameters in distributed systems, ASME conference on Automatic Control*, Austin, Texas, June 17-21, 1974. Goodson R.E. and Polis M. Eds., 31-48.

Craven P, and Grace Wahba, 1979: Smoothing Noisy Data with Spline functions. *Numer. Math. Vol.*, **31**, 377-403

Dyer, A.J., 1967: The turbulent transport of heat and water vapor in an unstable atmosphere. *Q. J. R. Meteorol. Soc.*, **93**, 501-508.

Fletcher, R, 1987: Practical Methods of optimization, Second Edition, *John Willey & Sons*, New York, 436pp

Gill P.E., Walter Murray and Margaret H. Wright, 1981: Practical Optimization. *Academic Press*, London, 401pp.

Gordon, C.T. and W.F.Stern, 1982: A description of the GFDL global spectral model. *Mon. Wea. Rev.*, **110**, 625–644.

Haltiner George J. and R.T. Williams, 1980: Numerical prediction and dynamic meteorology. *John Wiley & Sons, Inc.*, **Second Edition**, 477pp.

Hamming, R. W., 1973: Numerical methods for scientists and engineers. 2nd ed.,*McGraw-Hill*, New York, 505pp.

Kanamitsu, M., K. Tada, T. Kudo, N. Sato and S. Isa, 1983: Description of JMA Operational Spectral Model. *J. Meteorol. Sci. of Japan*, **61**, 812-827.

Kanamitsu, M., 1989: Description of the NMC Global Data Assimilation and Forecast System. *Weather and Forecasting*, **4**, 335–342.

Kanamitsu M., J.C. Alpert, K.A. Campana, P.M. Caplan, D.G. Deaven, M.Iredell, B. Katz, H.-L. Pan, J. Sela and G.H. White, 1991: Recent changes implemented into the global forecast system at NMC. *Weather and Forecasting*, **6**, 425-435.

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

Krishnamurti, T.N. and N. Dignon, 1988: The FSU global spectral model. *Dept. of Meteorology, Florida State University.*

LeDimet F. X. and I. M. Navon, 1988: Variational and optimization methods in meteorology: A review. Technical Report - SCRI - 144. Supercomputer Computations

Research Institute, Florida State University. Tallahassee, FL 32306, 83pp.

Leith, C. E., 1965: Numerical simulation of the earth's atmosphere. *Math. in Comp. Physics*, **4**, New York, Academic Press.

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.*, **28**, 145-161.

Lewis J.M. and T.H. Grayson, 1972: The adjustment of Surface wind and Pressure bu Sasaki's Variational Matching Technique. *J. Applied Meterorology*, **11**, 586-597.

Lilly, D. K., 1965: On the computational stability of numerical solutions of time dependent non-linear geostrophical fluid dynamics problem. *Mon. Wea. Rev.*, **93**, 11-26.

MacVean, M.K., 1983: The effects of horizontal diffusion on baroclinic development in a spectral model. it. Q. J. R. Meteorol. Soc., **109**, 771–783.

Nash, S. G. and A. Sofer, 1996: Linear and nonlinear programming. *McGraw Hill Series on Industrial and Management Science*, 692pp.

Navon, I. M., 1969: Inclusion of lateral viscosity and the application of a Matsuno scheme in a two level model of the general circulation of the atmosphere. *M.Sc. Thesis*, Dept. of Meteorology, Hebrew University, Jerusalem, 77pp plus 40 figures.

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

Navon, I.M., 1998: Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *In Special Issue of Dynamics of Atmosphere and Ocean in honor of Prof. Richard Pfeffer.* 747pp.

Panchang, V. G. and J. J. O'Brien, 1989: On the determination of hydraulic model parameters using the strong constraint formulation. *Modeling Marine Systems*, **I**, Editor: A. M. Davies, CRC Press, Inc., 5-18.

Phillips, N.A., 1956: The general circulation of the atmosphere: a numerical experiment. *Q. J. R. Meteorol. Soc.*, **82**, 123-164.

Phillips, N.A., 1959: An example of non-linear computational instability. *The atmosphere and the sea in motion.* Ed. Bert Bolin, The Rockefeller Institute Press, New York, 501-504.

Pu, Z. X., E. Kalnay., J.C. Derber, and J.G. Sela, 1997: Using forecast sensitivity patterns to improve future forecast skill. *Q. J. R. Meteorol. Soc.*, **123**, 1035-1053.

Richard, J.R. and K.E. German, 1965: A contribution to the problem of stratospheric diffusion by large-scale mixing. *Mon. Wea. Rev.*, **93**, 313-321.

Robert, A.J., 1966: The integration of a low order spectral form of the primitive meteorological equations. *J. Meteorol. Sci. of Japan*, **44**, 237-245.

Skamarock, W.C. and J.B. Klemp, 1992: The stability of time-split numerical methods for the hydrostatic and the nonhydrostatic elastic equations. *Mon. Wea. Rev.*, **120**, 2109-21127.

Smith, S.D., C.W. Fairall, G.L. Geernaert, and L. Hasse, 1996: Air-sea fluxes: 25 years of progress. *Boundary-Layer Meteorol.*, **78**, 247-290.

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

Tichonov, A.N. and V. Y. Arsenin, 1977: Solutions of ill-posed problems. Halsted Press, New York, 258pp.

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

Yu, L. and J. J. O'Brien, 1991: Variational estimation of the wind stress drag coefficient and the oceanic eddy viscosity profile. *J. Phys. Oceanogr.*, **21**, 709-719.

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

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, College of Arts and Sciences, The Florida State University, 235 pp.

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.

Zou, X., I.M. Navon, and F.X. LeDimet, 1992a: An optimal nudging data assimilation scheme using parameter estimation. *Q. J. R. Meteorol. Soc .*, **118**, 1163-1186.

Zou, X., I.M. Navon and F.X. LeDimet, 1992b: Incomplete Observations and Control of Gravity Waves in Variational Data Assimilation, *Tellus*, vol. bf 44A, No. **4** , 273-296.

Zou, X., I.M. Navon and J. Sela, 1993: Control of Gravity Oscillations in Variational Data Assimilation, *Mon. Wea. Rev.*, vol. **121**, No. **1**, 272-289.

TABLE CAPTIONS

Table 1. Experiments designed to assess the impact of optimal parameter estimation.

Table 2. Experiments carried out by integrating the model for 24h, 48h and 72h from 0000UTC 3 Sept. 1996, respectively.

Experiment	Initial Condition	Parameter Value	Integration
C1	initial guess	estimated value	6 h
C2	initial guess	optimal value	6 h
O1	optimal initial condition	estimated value	6 h
O2	optimal initial condition	optimal value	6 h
O3	optimal initial condition	estimated value	24 h
O4	optimal initial condition	optimal value	24 h

Exp.	Initial condition	Parameter values	Length of forecast
C1	initial guess	estimated values	24 h
C2	initial guess	estimated values	48 h
C3	initial guess	estimated values	72 h
P1	initial guess	optimal values	24 h
P2	initial guess	optimal values	48 h
P3	initial guess	optimal values	72 h
I1	optimal initial condition	estimated values	24 h
I2	optimal initial condition	estimated values	48 h
I3	optimal initial condition	estimated values	72 h

FIGURE CAPTIONS

Figure 1. The divergence fields at 200 mb at 0000UTC 3 Sept. 1996: (a) the observation, (b) the initial guess. The contour interval is $5.0 \times 10^{-6}\ s^{-1}$.

Figure 2. The variation of the cost function and the gradient norm with respect to the iteration number when both the optimal initial condition and the bi-harmonic horizontal diffusion coefficient κ are retrieved.

Figure 3. The evolution of the horizontal diffusion coefficient.

Figure 4. The RMS errors of the vorticity, divergence, virtual temperature and dewpoint depression at the end of the assimilation window. solid line: the control experiment; dashed line: the simulation experiment; dotted line: the optimal experiment.

Figure 5. The divergence fields at 200 mb at the end of the assimilation window: (a) the observation, (b) the control experiment, (c) the simulation experiment and (d) the optimal experiment, respectively. The contour interval is $8.0 \times 10^{-6}\ s^{-1}$.

Figure 6. The difference fields of specific humidity at 500 mb between the results of the following three experiments and the observation at the end of the assimilation window: (a) the control experiment, (b) the simulation experiment and (c) the optimal experiment. The contour interval is 0.0005 g/g.

Figure 7. The difference fields of temperature at 850 mb between the results of the following three experiments and the observation at the end of the assimilation window: (a) the control experiment, (b) the simulation experiment and (c) the optimal experiment. The contour interval is $1K$.

Figure 8. The RMS errors of the vorticity, divergence, virtual temperature and dewpoint depression at the end of the 24h forecast. solid line: the control experiment; dotted line: the optimal experiment.

Figure 9. The differences of the RMS errors of the vorticity, divergence, virtual temperature and the dewpoint depression between the three pairs at the end of the forecast. solid line: C2 and C1; dashed line: O2 and O1; dot dashed line: O4 and O3.

Figure 10. The differences of the RMS errors of the 24h forecast vorticity, divergence, virtual temperature and the dewpoint depression between the following pairs. dashed line: experiments (b) and (a); dot dot dashed line: experiments (c) and (a); solid line: experiments (d) and (a).

Figure 11. The percent differences of the RMS errors of the ensuing forecast fields vorticity, divergence, virtual temperature and the dewpoint depression between when the optimally identified parameter values are used and when the estimated parameter values are used. solid line: 24h forecast; long, short dashed line: 48h forecast; dot dot dashed line: 72h forecast.

Figure 12. The percent differences of the RMS errors of the ensuing forecast fields vorticity, divergence, virtual temperature and the dewpoint depression between when the optimal initial condition is used and when the initial guess of initial condition is used. solid line: 24h forecast; long, short dashed line: 48h forecast; dot dot dashed line: 72h forecast.

Figure 13 The histogram of vorticity RMS errors for the optimal parameter experiments (black bar) and optimal initial condition experiments (grey bar).

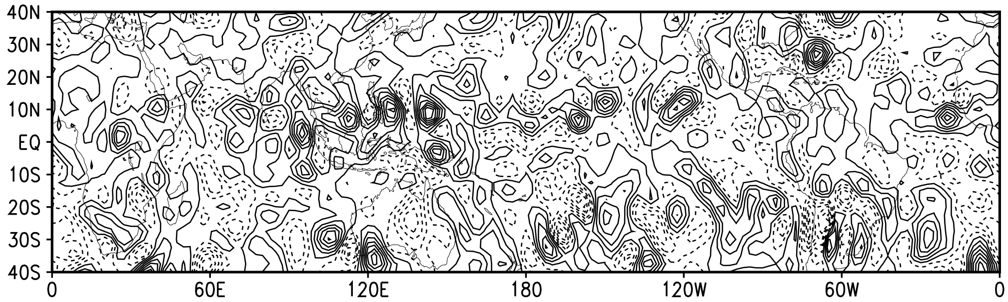
Figure 14 The histogram of dewpoint depression RMS errors for the optimal parameter experiments (black bar) and optimal initial condition experiments (grey bar).

Figure 15. The difference fields between experiments (to be enumerated below) and the analysis for the specific humidity at 850 mb as follows: (a) the control experiment C1, (b) experiment P1. The contour interval is 0.003 g/g.

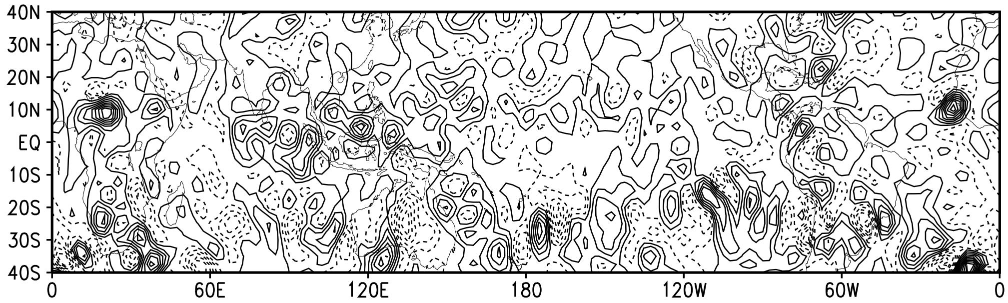
Figure 16. The difference field of specific humidity at 850 mb between experiments C1 where the estimated parameter values are used and P1 where the optimally identified parameter values are used. The contour interval is 0.0008 g/g.

Figure 17. The difference fields between experiments (to be enumerated below) and the analysis for the specific humidity at 500 mb as follows: (a) the control experiment C1, (b) experiment I1. The contour interval is 0.0008 g/g.

Figure 18. Anomaly correlation of the specific humidity forecasts at 850 mb for the control experiments (solid line), "optimal parameter" experiments (dot dot dashed line), "optimal initial condition" experiments (dotted line), and the experiments using both the optimal initial condition and optimally identified parameter values (dashed line). Upper panel: tropical belt; Lower panel: Northern Hemisphere.



(a)



(b)

Figure 1: The divergence fields at 200 mb at 0000UTC 3 Sept. 1996: (a) the observation, (b) the initial guess. The contour interval is $5.0 \times 10^{-6}\ s^{-1}$.

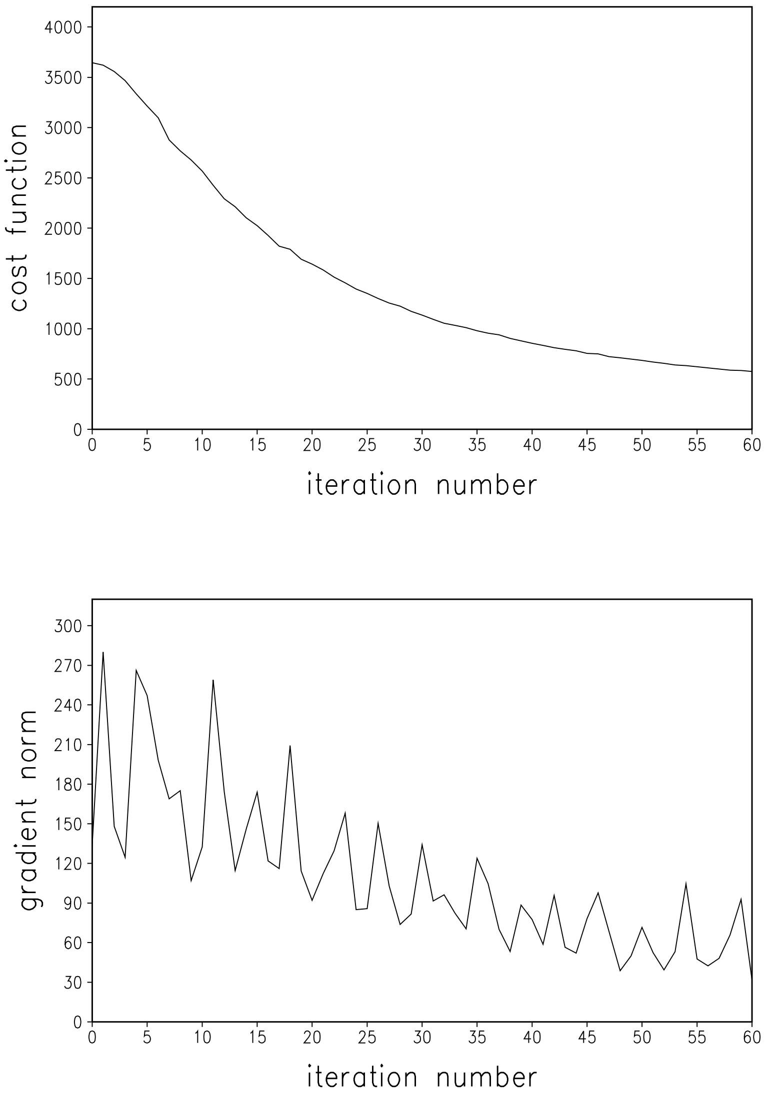


Figure 2: The variation of the cost function and the gradient norm with respect to the iteration number when both the optimal initial condition and the bi-harmonic horizontal diffusion coefficient κ are retrieved.

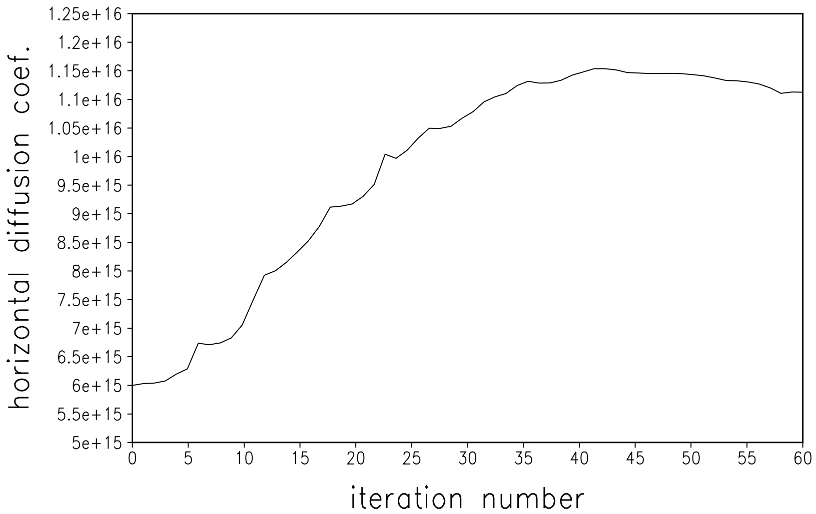


Figure 3: The evolution of the horizontal diffusion coefficient.

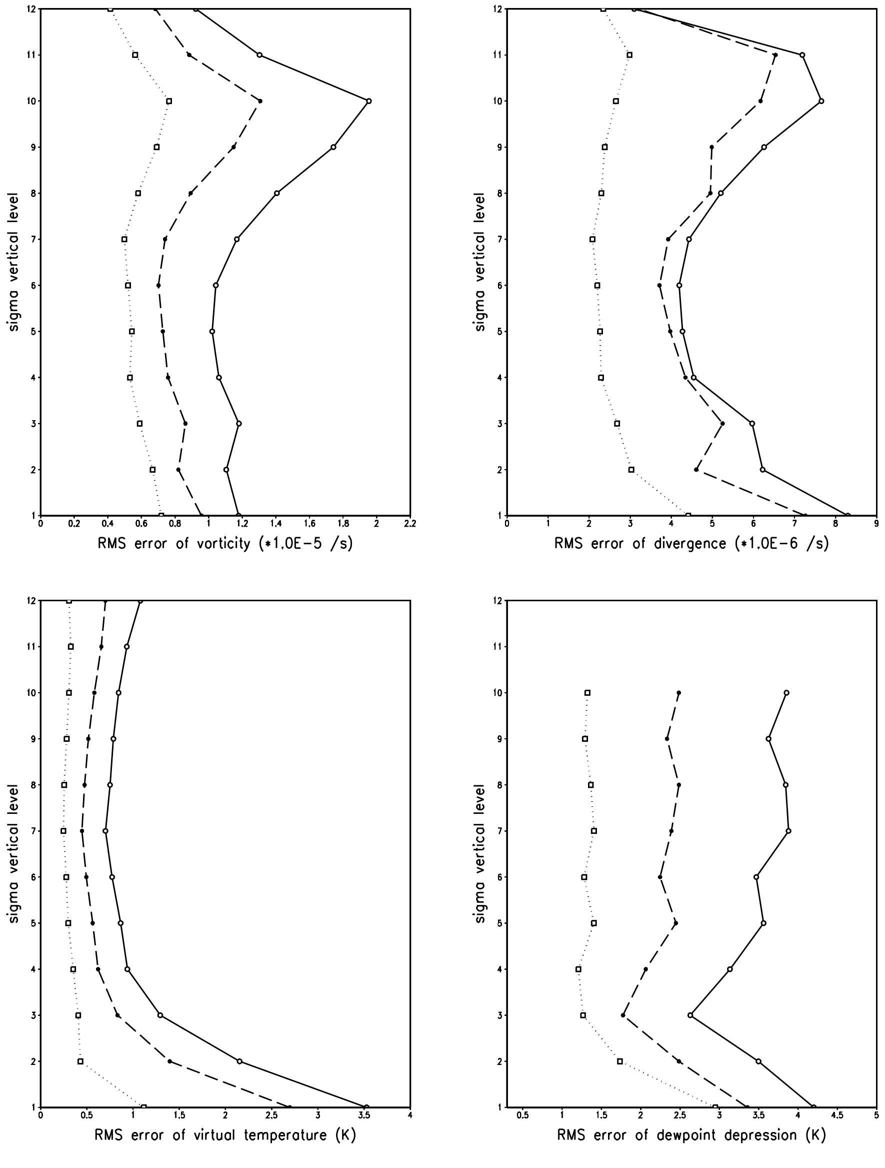


Figure 4: The RMS errors of the vorticity, divergence, virtual temperature and dewpoint depression at the end of the assimilation window. solid line: the control experiment; dashed line: the simulation experiment; dotted line: the optimal experiment.

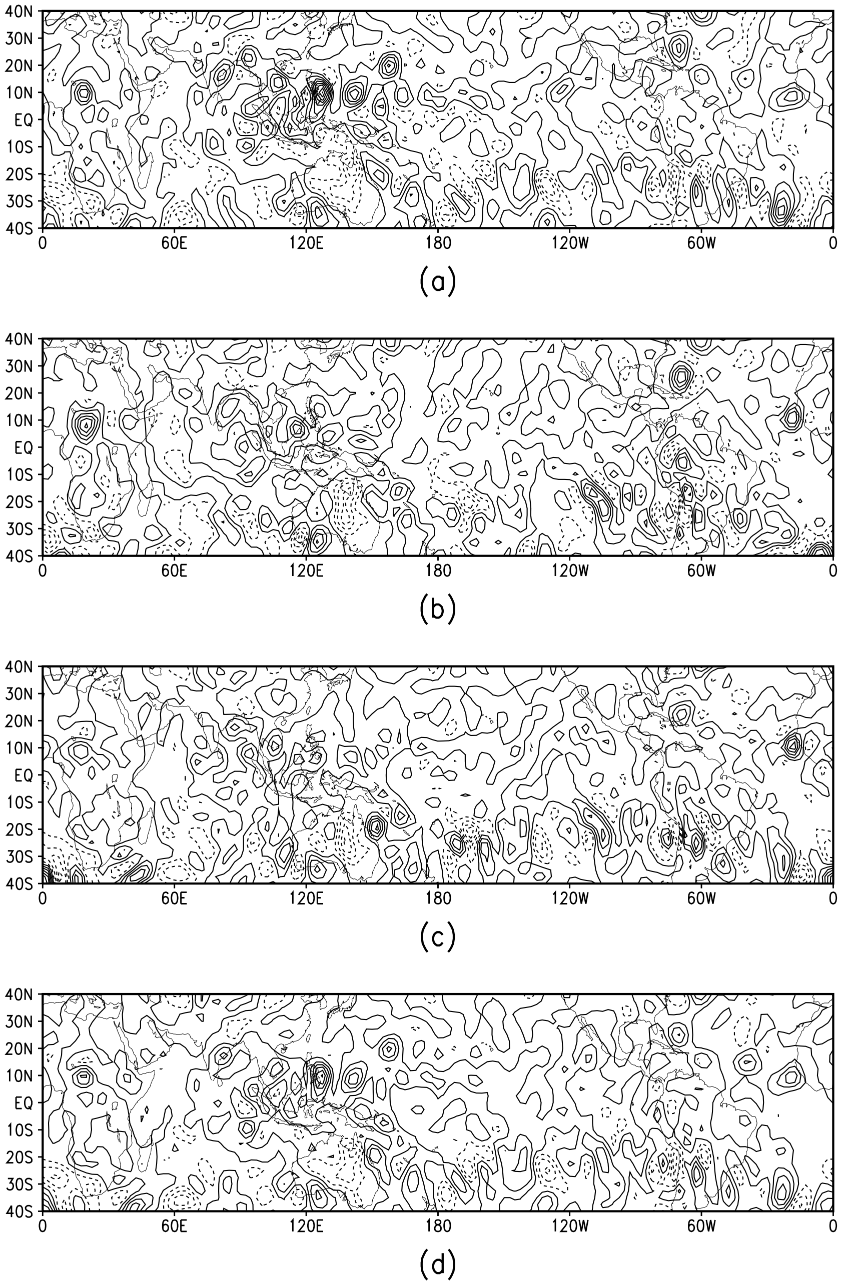


Figure 5: The divergence fields at 200 mb at the end of the assimilation window: (a) the observation, (b) the control experiment, (c) the simulation experiment and (d) the optimal experiment, respectively. The contour interval is $8.0 \times 10^{-6}\ s^{-1}$.

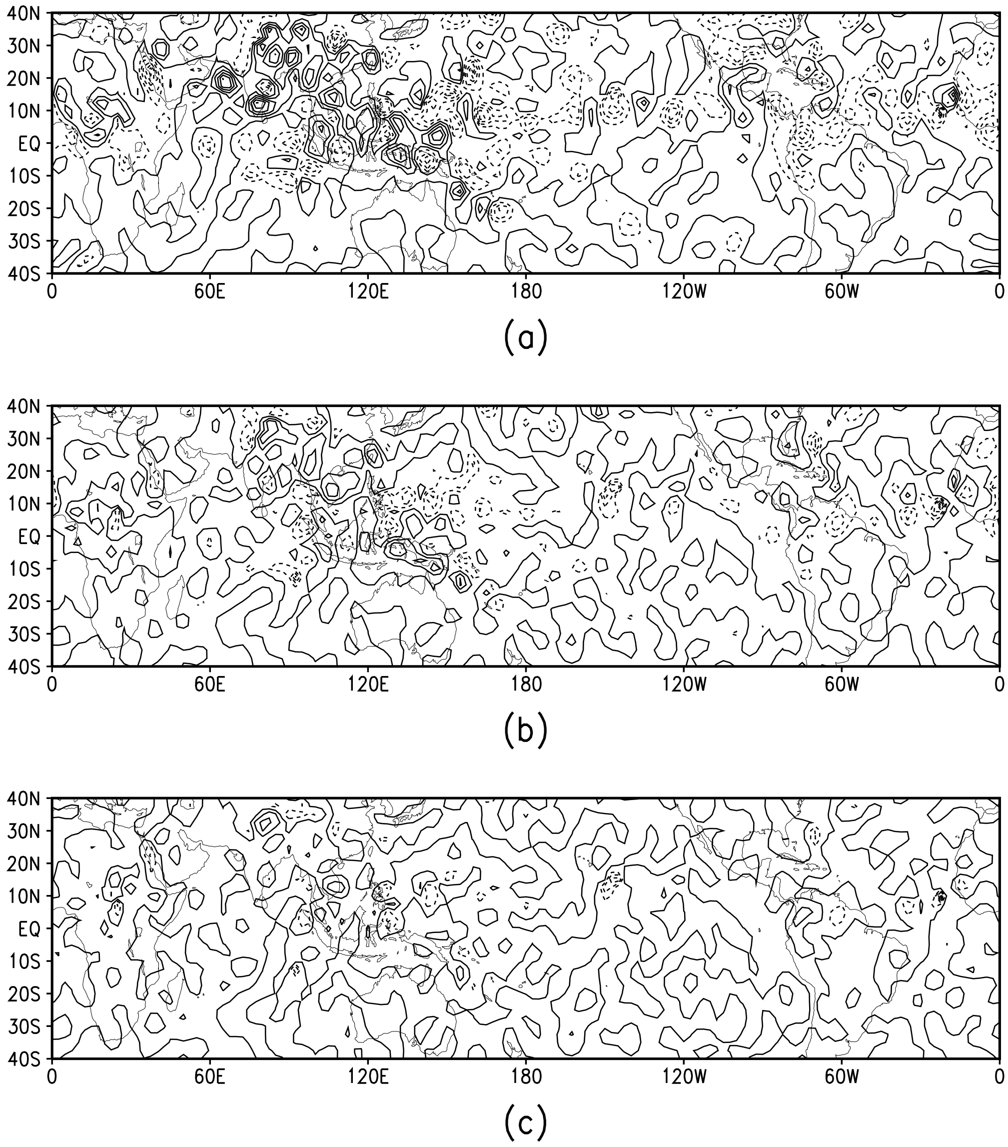


Figure 6: The difference fields of specific humidity at 500 mb between the results of the following three experiments and the observation at the end of the assimilation window: (a) the control experiment, (b) the simulation experiment and (c) the optimal experiment. The contour interval is 0.0005 g/g.

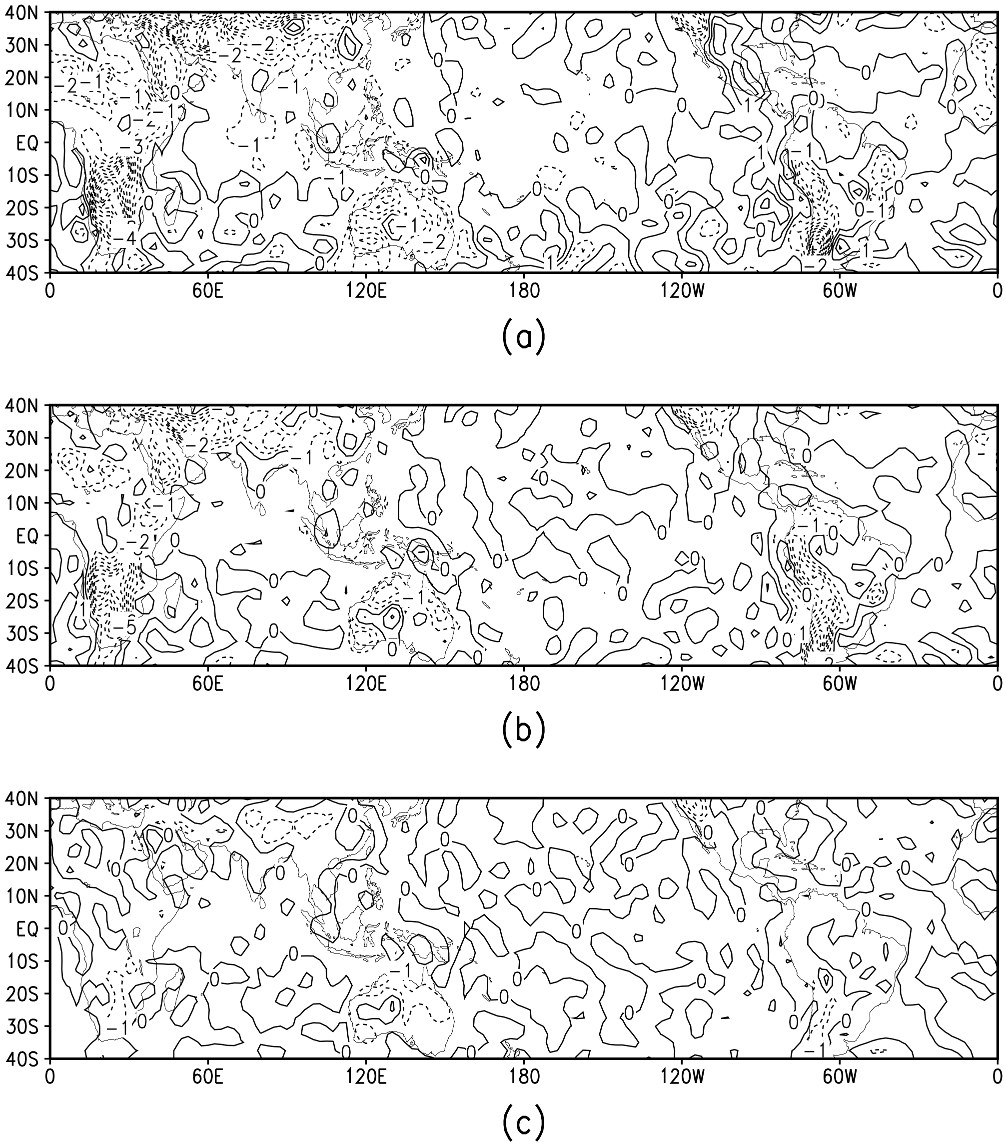


Figure 7: The difference fields of temperature at 850 mb between the results of the following three experiments and the observation at the end of the assimilation window: (a) the control experiment, (b) the simulation experiment and (c) the optimal experiment. The contour interval is $1K$.

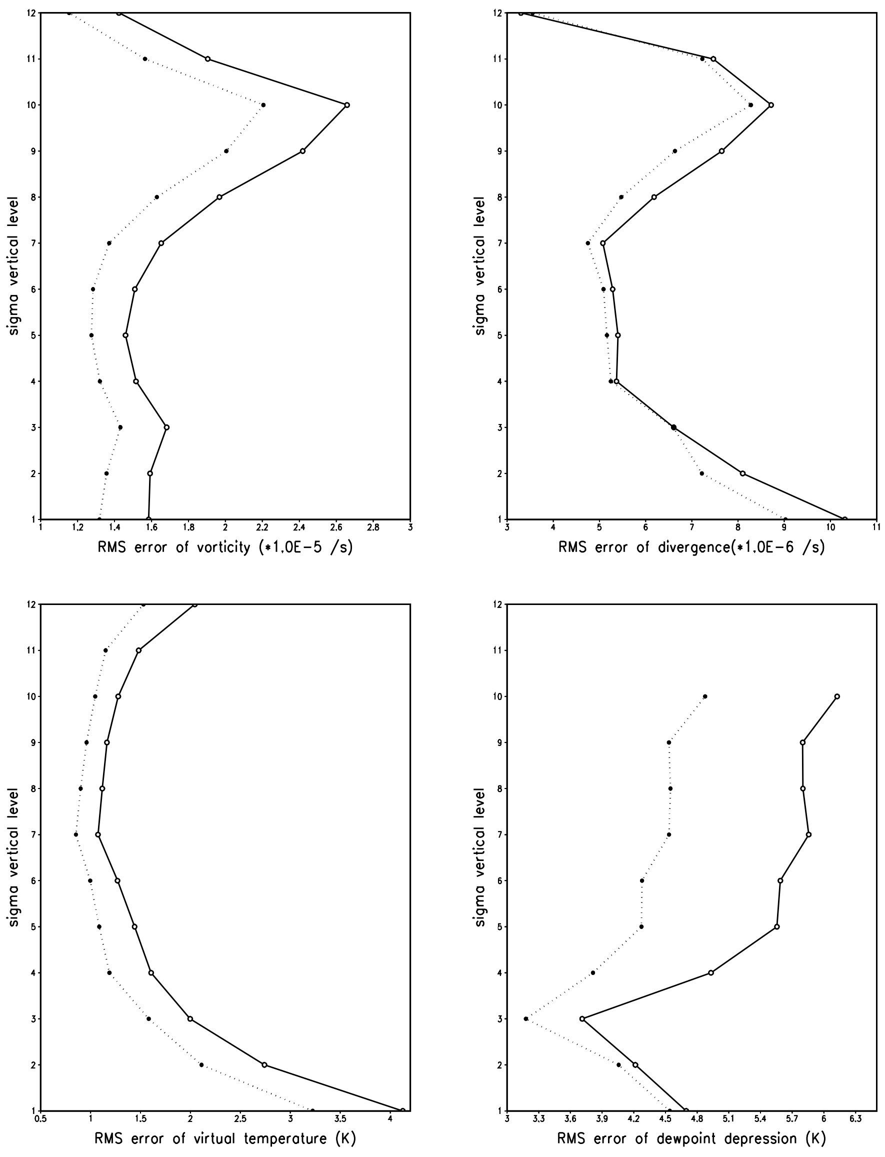


Figure 8: The RMS errors of the vorticity, divergence, virtual temperature and dewpoint depression at the end of the 24h forecast. solid line: the control experiment; dotted line: the optimal experiment.

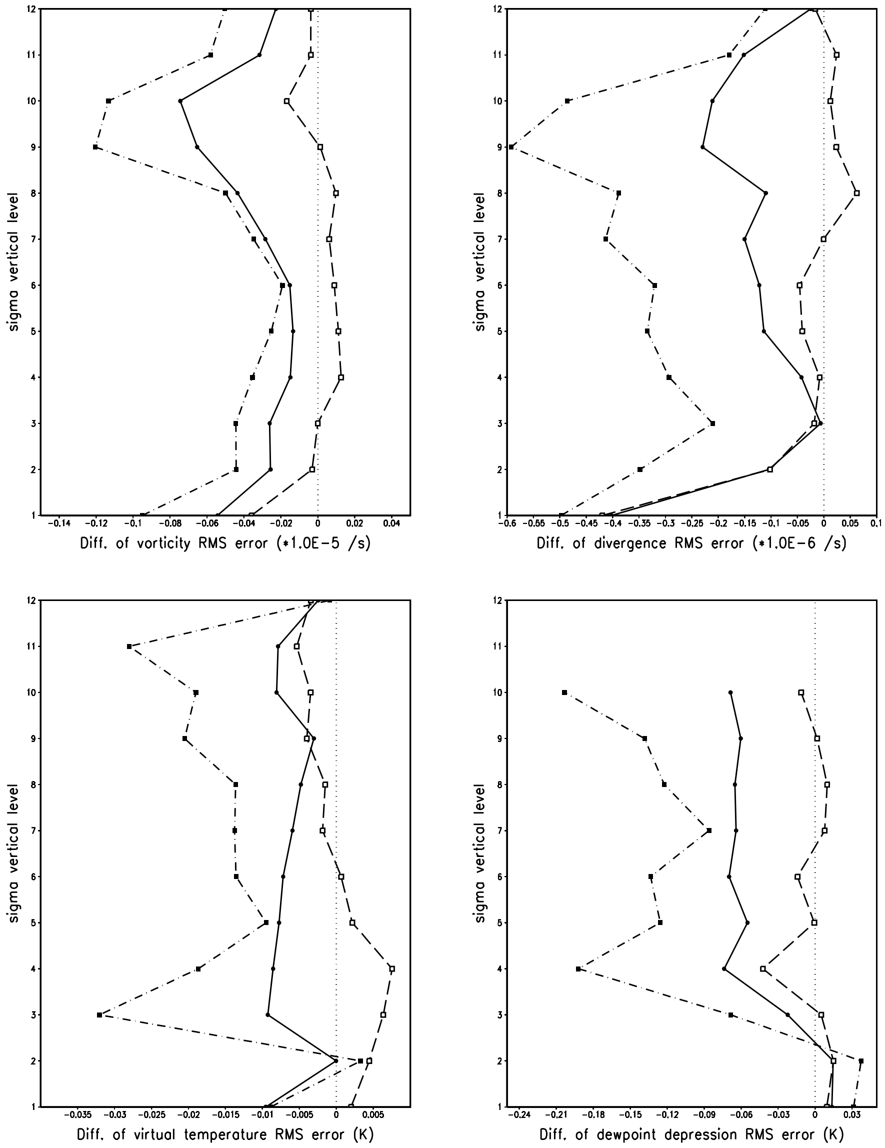


Figure 9: The differences of the RMS errors of the vorticity, divergence, virtual temperature and the dewpoint depression between the three pairs at the end of the forecast. solid line: C2 and C1; dashed line: O2 and O1; dot dashed line: O4 and O3.

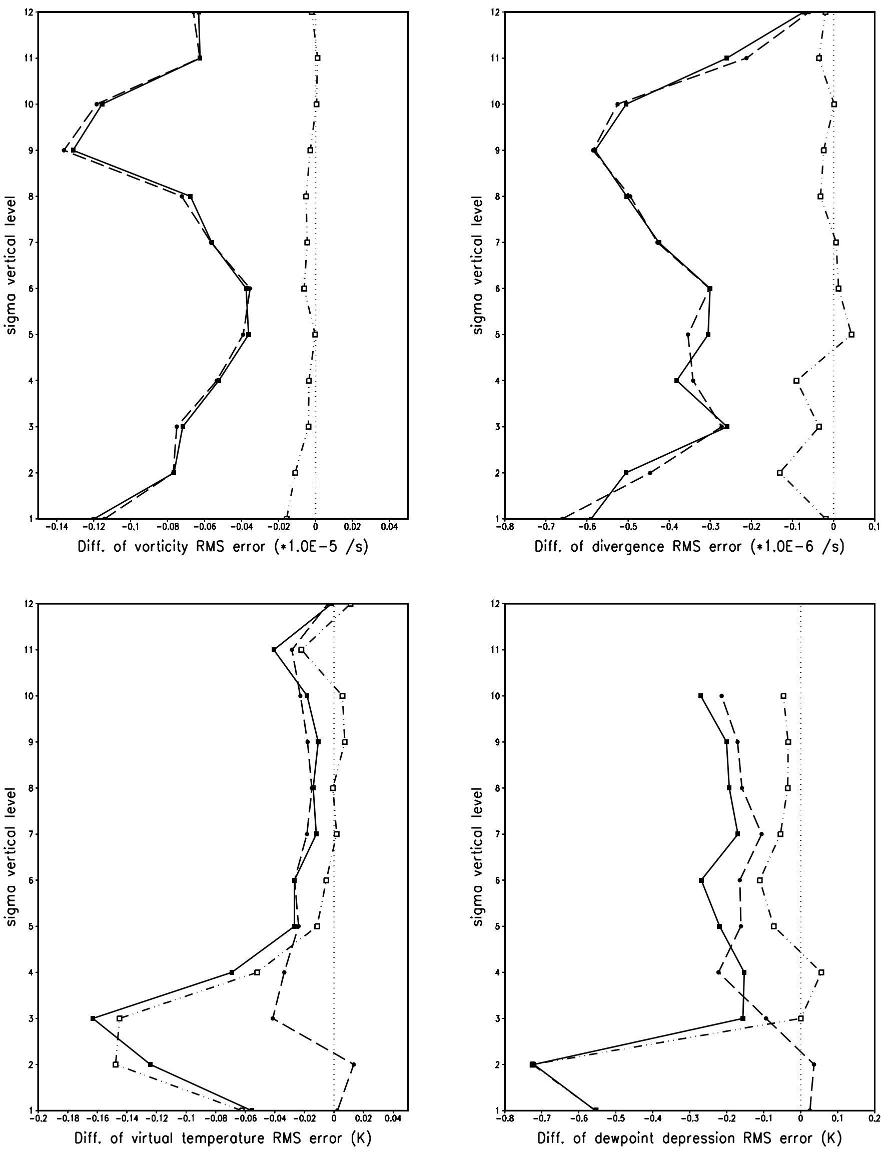


Figure 10: The differences of the RMS errors of the 24h forecast vorticity, divergence, virtual temperature and the dewpoint depression between the following pairs. dashed line: experiments (b) and (a); dot dot dashed line: experiments (c) and (a); solid line: experiments (d) and (a).

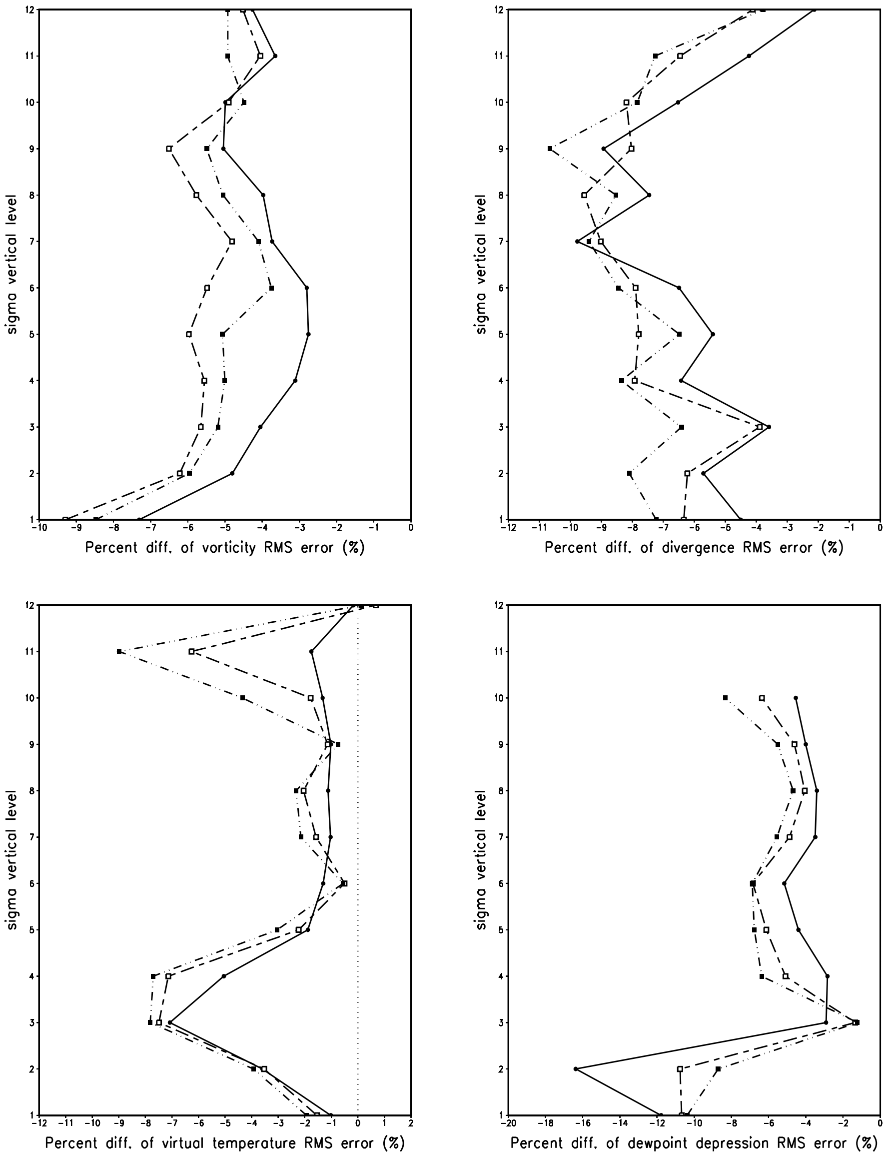


Figure 11: The percent differences of the RMS errors of the ensuing forecast fields vorticity, divergence, virtual temperature and the dewpoint depression between when the optimally identified parameter values are used and when the estimated parameter values are used. solid line: 24h forecast; long, short dashed line: 48h forecast; dot dot dashed line: 72h forecast.

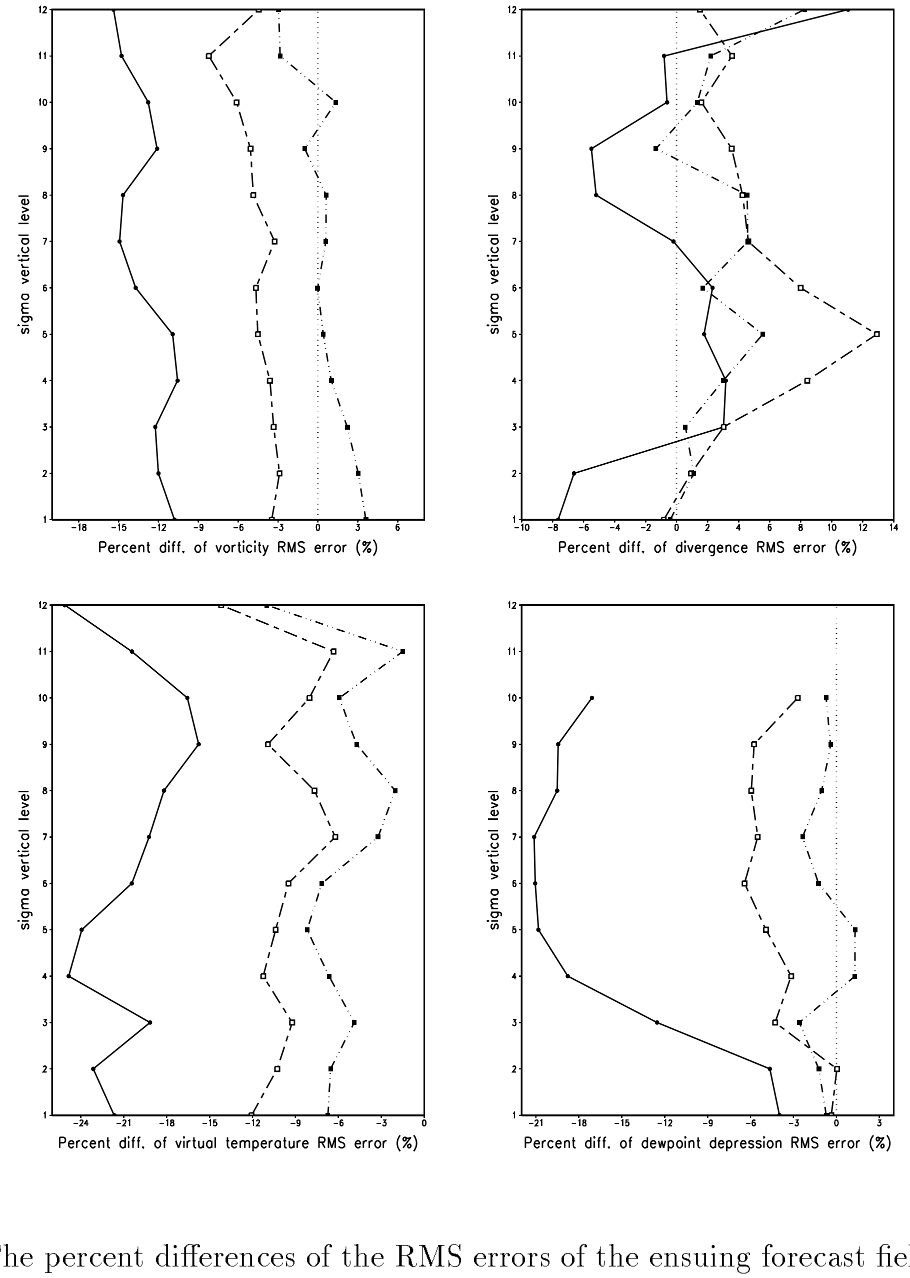


Figure 12: The percent differences of the RMS errors of the ensuing forecast fields vorticity, divergence, virtual temperature and the dewpoint depression between when the optimal initial condition is used and when the initial guess of initial condition is used. solid line: 24h forecast; long, short dashed line: 48h forecast; dot dot dashed line: 72h forecast.

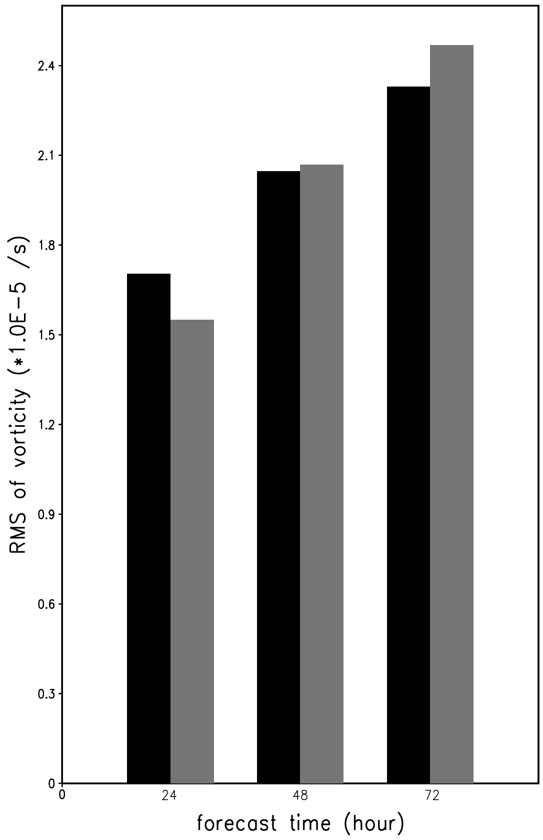


Figure 13: The histogram of vorticity RMS errors for the optimal parameter experiments (black bar) and optimal initial condition experiments (grey bar).

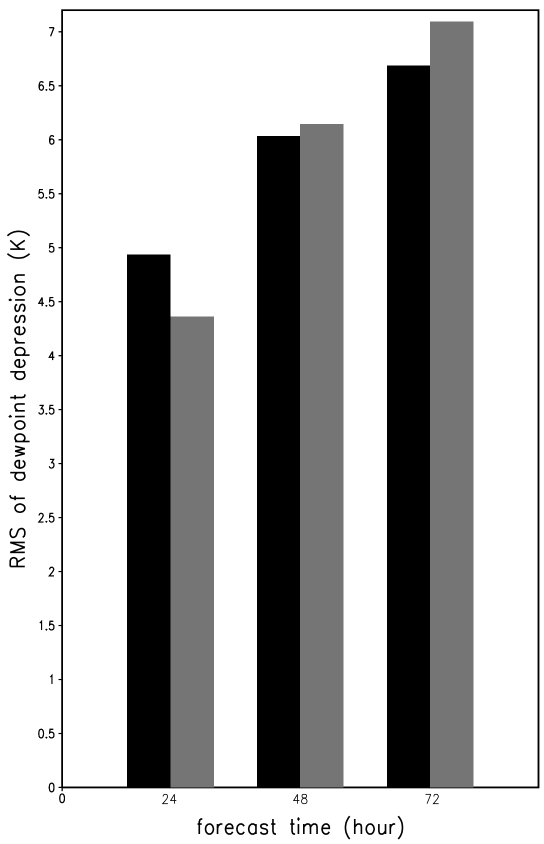
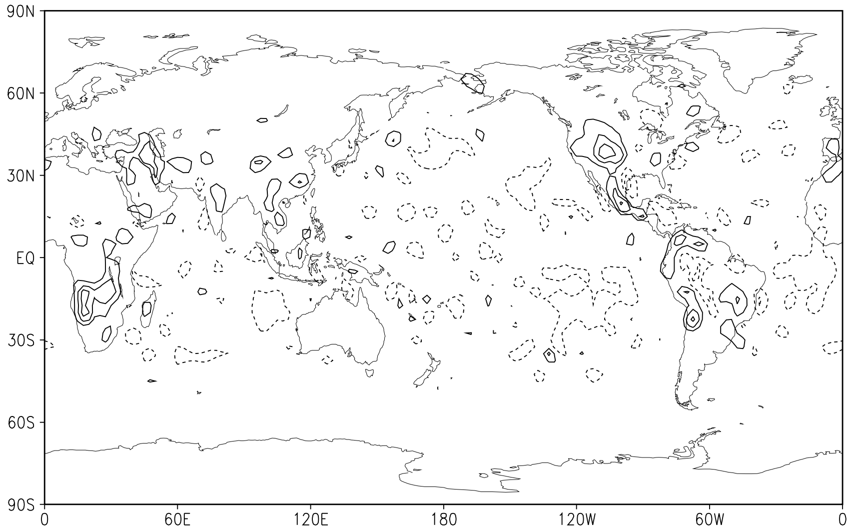
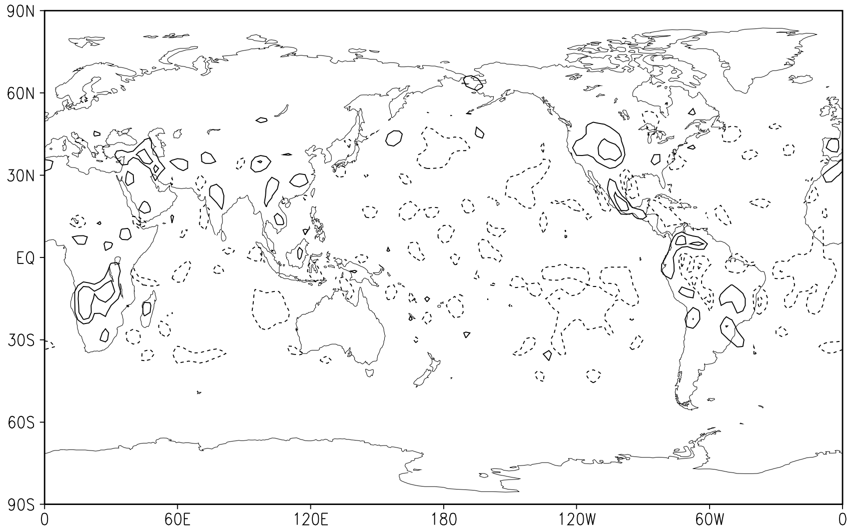


Figure 14: The histogram of dewpoint depression RMS errors for the optimal parameter experiments (black bar) and optimal initial condition experiments (grey bar).



(a)



(b)

Figure 15: The difference fields between experiments (to be enumerated below) and the analysis for the specific humidity at 850 mb as follows: (a) the control experiment C1, (b) experiment P1. The contour interval is 0.003 g/g.

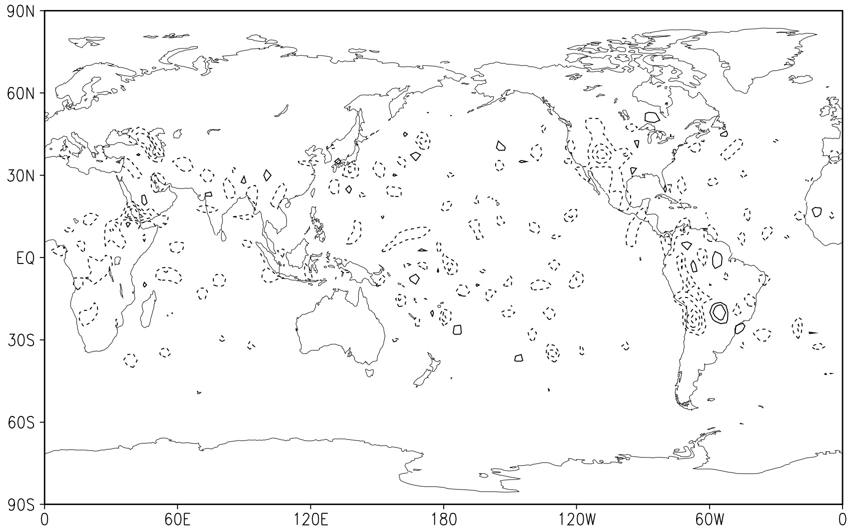
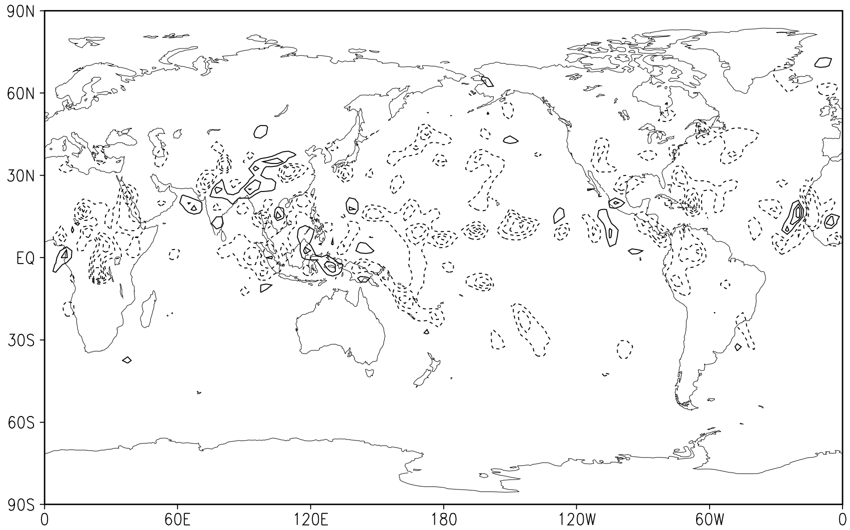
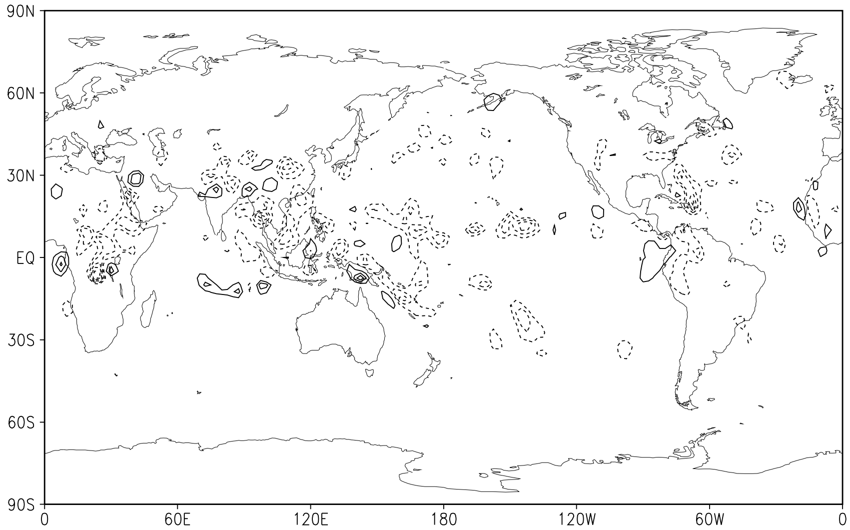


Figure 16: The difference field of specific humidity at 850 mb between experiments C1 where the estimated parameter values are used and P1 where the optimally identified parameter values are used. The contour interval is 0.0008 g/g.



(a)



(b)

Figure 17: The difference fields between experiments (to be enumerated below) and the analysis for the specific humidity at 500 mb as follows: (a) the control experiment C1, (b) experiment I1. The contour interval is 0.0008 g/g.

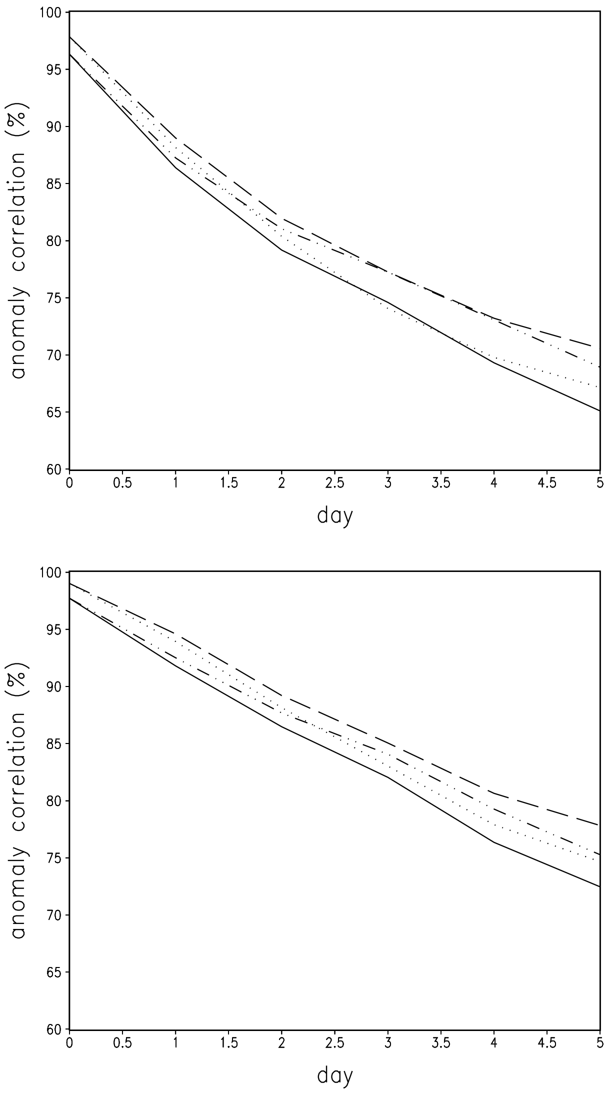


Figure 18: Anomaly correlation of the specific humidity forecasts at 850 mb for the control experiments (solid line), "optimal parameter" experiments (dot dot dashed line), "optimal initial condition" experiments (dotted line), and the experiments using both the optimal initial condition and optimally identified parameter values (dashed line). Upper panel: tropical belt; Lower panel: Northern Hemisphere.