

Sensitivity of Forecast Errors to Initial Conditions with a Quasi-Inverse Linear Method

ZHAO-XIA PU

Lanzhou University, Lanzhou, People's Republic of China, and UCAR Visiting Scientist, National Centers for Environmental Prediction, Washington, D.C.

EUGENIA KALNAY, JOSEPH SELA, AND ISTVAN SZUNYOGH*

Environmental Modeling Center, National Centers for Environmental Prediction, Washington, D.C.

(Manuscript received 19 September 1996, in final form 9 January 1997)

ABSTRACT

A quasi-inverse linear method has been developed to study the sensitivity of forecast errors to initial conditions for the National Centers for Environmental Prediction's (NCEP) global spectral model. The inverse is approximated by running the tangent linear model (TLM) of the nonlinear forecast model with a negative time step, but reversing the sign of friction and diffusion terms, in order to avoid the computational instability that would be associated with these terms if they were run backward. As usually done using the adjoint model integrations, the quasi-inverse TLM is started at the time of the verified forecast error and integrated backward to the corresponding initial time.

First, a numerical experiment shows that this quasi-inverse linear estimation is able to trace back the differences between two perturbed forecasts from the NCEP ensemble forecasting system and recover with good accuracy the known difference between the two forecasts at the initial time. This result shows that both the linear estimation and the quasi-inverse linear estimation are quite close to the nonlinear evolution of the perturbation in the nonlinear forecast model, suggesting that it should be possible to apply the method to the study of the sensitivity of forecast errors to initial conditions. The authors then calculate the perturbation field at the initial time (estimate the initial error) by tracing back a 1-day forecast error using the TLM quasi-inverse estimation. As could be expected from the previous experiment, when the estimated error is subtracted from the original analysis, the new initial conditions lead to an almost perfect 1-day forecast. The forecasts beyond the first day are also considerably improved, indicating that the initial conditions have indeed been improved.

In the remainder of the paper, this quasi-inverse linear method is compared with the adjoint sensitivity method (Rabier et al., Pu et al.) for medium-range weather forecasting. The authors find that both methods are able to trace back the forecast error to perturbations that improve the initial conditions. However, the forecast improvement obtained by the quasi-inverse linear method is considerably better than that obtained with a single adjoint iteration and similar to the one obtained using five iterations of the adjoint method, even though each adjoint iteration requires at least twice the computer resources of the quasi-inverse TLM estimation. Whereas the adjoint forecast sensitivities are closely related to singular vectors, the quasi-inverse linear perturbations are associated with the bred (Lyapunov) vectors used for ensemble forecasting at NCEP (Toth and Kalnay). The features of the two types of perturbations are also compared in this study. Finally, the possibility of the use of the sensitivity perturbation to improve future forecast skill is discussed, and preliminary experiments encourage further testing of this rather inexpensive method for possible operational use.

The model used in this study is the NCEP operational global spectral model at a resolution of T62/L28. The corresponding TLM, and its adjoint, are based on an adiabatic version of the model but include both horizontal and vertical diffusion.

1. Introduction

In the last two decades, the skill of numerical weather prediction has improved enormously and has become

the essential guidance in most weather forecast centers. These improvements are due to three main factors: 1) the use of finer spatial resolution made possible by substantial increases in computational power and more efficient numerical techniques, 2) more comprehensive and accurate representation of the physical processes within the models, and 3) improved methods for data assimilation and use of new types of observations resulting in better initial conditions for the atmospheric models. Recent experience with data assimilation and forecast experiments suggest that large forecast errors usually arise from errors in the initial conditions rather

* Current affiliation: Visiting Scientist, National Center for Atmospheric Research, Boulder, Colorado.

Corresponding author address: Eugenia Kalnay, EMC/NCEP2, WWB/Rm. 207, 5200 Auth Road, Camp Springs, MD 20746.
E-mail: wd23ek@sun1.wwb.noaa.gov

than from errors in model formulation, at least in the extratropics (Reynolds et al. 1994; Simmons 1995; Rabier et al. 1996). Predictability studies such as those by Simmons et al. (1995) suggest that improvements in the estimation of the initial state offer the most promising path to more accurate individual (deterministic) forecasts, although there is still scope for benefits from model improvement and from ensemble forecasting.

In recent years a considerable effort has been placed on the use of advanced data assimilation methods in order to improve the forecasts initial conditions. The 3D variational techniques have become operationally feasible and were implemented in 1991 at the National Centers for Environmental Prediction (NCEP, formerly the National Meteorological Center) and in 1996 at the European Centre for Medium-Range Weather Forecasts (ECMWF), replacing the optimal interpolation schemes (Derber et al. 1991; Parrish and Derber 1992; Andersson et al. 1996). The more advanced 4D variational data assimilation method remains very expensive and still on the edge of feasibility for operational implementation (Courtier et al. 1994), although Zupanski and Zupanski (1996) have shown excellent convergence properties for the NCEP regional Eta Model. Efforts to develop computationally feasible applications of Kalman filtering to data assimilation are also under way (Cohn 1994). There are also studies suggesting improvement of the initial conditions by using either the forecast error itself, or the estimation of the growing modes of the atmosphere to decrease the uncertainty in the initial conditions (Kalnay and Toth 1994; Rabier et al. 1996).

Since it is not easy to exactly separate the errors due to the initial conditions from those due to model deficiencies, there has been considerable interest in the investigation of the sensitivity of forecast errors to initial conditions. Recent studies with adjoint models in numerical weather prediction (Rabier et al. 1996; Pu et al. 1996, 1997) have shown that the gradient of the short-range forecast error taken with respect to the initial conditions, commonly referred to as a "sensitivity pattern," can be an effective means of identifying structures in the initial conditions that might cause large forecast errors. Rabier et al. calculated a small perturbation (forecast error sensitivity) in the initial conditions that minimized the observed 2-day forecast error over the Northern Hemisphere using the adjoint method; the perturbation is obtained as the gradient of an error function with respect to the initial condition, multiplied by a fixed step size. Their experiments showed that the sensitivity forecasts from the adjusted (perturbed) initial conditions were better than the forecasts from the original (unperturbed) initial conditions at the same starting time (2 days old), but *not* better than the forecasts from the latest available operational initial conditions. Pu et al. (1997) extended this idea, by first calculating the initial perturbation of forecast sensitivity with a single iteration of conjugate-gradient method for NCEP's global spec-

tral forecast model, and then using the improved 2-days-old initial conditions in a second iteration of the NCEP three-dimensional variational analysis cycle until the latest initial conditions are reached. Their results demonstrated that the method enhances the future medium-range weather forecast skill. Since the adjoint minimization method is an iterative process, Zupanski (1995) suggested that the sensitivity pattern should also be performed iteratively. Using the NCEP regional model Eta Model and its adjoint, he showed considerable improvement in the sensitivity patterns using up to 10 iterations.

In the present study we develop a quasi-inverse linear estimation of the perturbation in the initial conditions leading to observed forecast errors as an alternative to the adjoint (transpose) method used so far. The quasi-inverse is calculated by integrating the tangent linear model (TLM) backward and thus tracing back the forecast errors to the initial time. The differences between the quasi-linear error estimation and the adjoint sensitivity pattern are then presented.

The TLM has been used in the context of sensitivity studies (Errico and Vukicevic 1992; Lacarra and Talagrand 1988; Errico et al. 1993), since it describes the *forward* evolution of small perturbations in a forecast model. In most cases, however, the TLM has been used to determine the evolution of small perturbations of fields in a model forecast and, most importantly, in the development and assessment of the adjoint model. This latter application has received much attention in recently years, since the adjoint model has been extensively applied in four-dimensional variational data assimilation (Le Dimet and Talagrand 1986; Derber 1987, 1989; Navon et al. 1992; Zupanski 1993; Zou et al. 1992; Zupanski and Mesinger 1995) and in sensitivity analyses (Vukicevic 1991; Zou et al. 1992; Zupanski 1995; Pu et al. 1997; Rabier et al. 1996). The TLM is always used to evaluate the level of accuracy of applications of its corresponding adjoint, because the accuracies of both models are strongly related, and it is usually easier to think in terms of a forward model of perturbation evolution (in a TLM) than a backward model of sensitivities (in an adjoint). In this situation, the TLM is intended to approximately describe the evolution of the small differences between two nonlinear model solutions, where one solution begins from perturbed initial or boundary conditions, or perturbed model parameters. A TLM may be considered to be accurate as long as the solution of TLM integration from the perturbation is a good estimate of the differences between two nonlinear model integrations. There are some studies that examined the accuracy of specific TLMs in the process of creating the adjoint (transpose) model (e.g., Errico et al. 1993, etc.). For this reason, although the development of a TLM is not strictly necessary for many of the adjoint applications, the TLM and the adjoint are usually developed together.

So far, only the adjoint model has been used in the analysis of the sensitivity of forecast error to initial con-

ditions, and although this application does not require the use of the TLM, it still contains the basic assumption of the TLM, namely, that the error dynamics are linear. The studies of the sensitivity of forecast errors to the initial conditions have been done by defining an objective forecast error function, and trying to find a solution that minimizes this function, and therefore required the adjoint model to calculate the gradient of the function. However, in studies of the behavior of initial perturbations growing in a forecast model, there has been evidence that the TLM can be directly used to estimate the amplitude of this growing perturbation up to 1–2-day forecasts. Lacarra and Talagrand (1988) experimentally showed that the barotropic time evolution of a small perturbation (with amplitude comparable to analysis errors) can be described by its linear approximation if the time interval is not longer than 2–3 days. Vukicevic (1991) investigated the linearity of initial error evolution using a primitive equation limited-area model and demonstrated that the major portion of initial forecast error (with magnitude comparable with analysis errors) can be described by the tangent model solutions for periods of about 1.5 days. Buizza (1994), comparing subjectively the time evolution of integrations, started adding and subtracting the same structure to the control initial condition with amplitude comparable to (optimal interpolation) analysis error estimates, and concluded that nonlinear effects are small up to forecast day 2 but they can be quite large after forecast day 4. From these experiments we can conclude that we can estimate the initial error evolution by the time integration of the TLM. If this is the case, we should consider whether we can use directly the TLM to estimate the initial error from the forecast error.

In this paper we address the following two questions. 1) Is it possible to estimate the initial errors from observed (perceived) short-range forecast errors through a backward integration of the TLM, and 2) can we use this method to improve operational forecasts? We develop a quasi-inverse linear method to estimate the initial errors leading to observed forecast errors. The numerical experiments are performed using the NCEP operational global spectral model with full physics, with T62 (horizontal triangular truncation of 62 waves) and 28 vertical sigma levels, and the simplified corresponding TLM based on an adiabatic version (Navon et al. 1992) but including horizontal diffusion and vertical mixing (Pu et al. 1997).

The paper is organized as follows. In section 2, we describe the mathematical formulation and the projection operator used to obtain the quasi-inverse linear estimation. In section 3, the nonlinear, linear, and adjoint models are described. A numerical experiment with two members of the NCEP operational forecast ensemble is performed in order to assess the accuracy of the linear forward model (propagator) and of its quasi-inverse operator. Section 4 contains numerical experiments showing the impact of the linear estimate of the initial error

(inverse of the perceived forecast error) on the forecast and a comparison of this impact with that obtained with the adjoint sensitivity patterns. In section 5, we further compare the differences of the estimates of the initial error of the quasi-inverse TLM estimation and adjoint method. Their respective relationship to the bred (Lyapunov) and singular vectors is also discussed in section 5. In section 6, the possibility of improving future forecast skill by using the initial error estimate is tested. Section 7 is a summary and discussion. The appendix discusses some properties of growing and decaying perturbations in Hamiltonian systems such as the inviscid adiabatic dynamics of the model.

2. Mathematical formulation of the quasi-inverse method

Consider a nonlinear forecast model \mathbf{M} that computes the evolution of the model atmosphere from its initial state \mathbf{X} at time $t = 0$, to its state at time t :

$$\mathbf{X}_t = \mathbf{M}_t(\mathbf{X}_0). \quad (1)$$

The corresponding tangent linear model \mathbf{L} propagates a perturbation of the initial state $\delta\mathbf{X}_0$ forward in time:

$$\mathbf{M}_t(\mathbf{X}_0 + \delta\mathbf{X}_0) = \mathbf{M}_t(\mathbf{X}_0) + \mathbf{L}_t\delta\mathbf{X}_0 + O(\delta\mathbf{X}_0)^2. \quad (2)$$

Then, given a finite initial perturbation $\delta\mathbf{X}_0$, and its evolution through any finite time interval t , the TLM approximation can be considered reasonably accurate for as long as

$$\mathbf{M}_t(\mathbf{X}_0 + \delta\mathbf{X}_0) - \mathbf{M}_t(\mathbf{X}_0) \approx \mathbf{L}_t\delta\mathbf{X}_0. \quad (3)$$

As indicated in the introduction, for realistic atmospheric models, and for initial perturbations with amplitudes characteristic of the estimated atmospheric analysis errors, past research experience indicates that the approximation (3) remains acceptable for about 1–3 days.

If the atmospheric model \mathbf{M} contained only adiabatic frictionless dynamics, it would be a reversible Hamiltonian system (see the appendix). In that case, the linear tangent model would also be adiabatic and nondissipative and, therefore, reversible. A reversible model \mathbf{M} can be easily inverted by running it backward in time:

$$\mathbf{X}_0 = \mathbf{M}^{-1}(\mathbf{X}_t), \quad (4)$$

which is simply carried out by starting \mathbf{M} from the “final” conditions \mathbf{X}_t and changing the sign of the time step Δt . The same exact inversion (running backward in time) can be applied to a reversible TLM.¹

¹ A model written with a time-centered scheme can be inverted exactly. If the time scheme is noncentered, it is not exactly reversible, and the inversion will introduce errors of the order of the time step even for nondiffusive dynamics. For example, the NCEP global model is based on a semi-implicit leapfrog time scheme (centered in time), but with a Robert time filter, which slightly damps the highest frequencies. The latter effect will be acting both in the forward and in the backward integrations but is a small effect that does not affect the meteorologically meaningful components of the integration.

In reality, of course, comprehensive atmospheric models contain heating and frictional terms, and, like the real atmosphere, are not reversible. Nevertheless, the successful experience of early numerical weather prediction, which was based on quasigeostrophic, reversible dynamics, suggests that, at least for short-range forecasts, the evolution of the atmosphere is dominated by the reversible atmospheric dynamics. In fact, the linear tangent models (and their adjoints) successfully used at the ECMWF in the development of their ensemble forecasting system originally only contained quasigeostrophic reversible dynamics (Molteni and Palmer 1993). Later, the primitive equations were adopted for the TLM and its adjoint, again containing only reversible dynamics with the exception of a simple linear surface friction and vertical diffusion, which were added in lieu of the full parameterization of irreversible physical processes (Buizza et al. 1993). This simplified adjoint was also used by Rabier et al. (1996) in their forecast sensitivity studies. A similar TLM with just the linearized reversible atmospheric dynamics of the NCEP global model, but including a simple linear surface friction and vertical diffusion as irreversible processes, was used by Pu et al. (1997), Pu (1996), and in the present work.

The dominance of the reversible dynamics in the short-range forecasts, and the success of the simple TLM in describing the evolution of small perturbations, suggests that, if it was computationally feasible backward integration in time of the TLM would provide a fairly good approximation of its inverse and, therefore, allow tracing of forecast errors backward in time and approximate determination of the corresponding analysis errors. We know that dissipative terms are computationally unstable if they are integrated backward in time, so we have a simple choice: either not to include them at all in the approximate inverse of the TLM, or, if the backward integration without friction becomes too noisy, to include them with the sign reversed. Note that if their sign is reversed, dissipative terms would be handled exactly as in an adjoint model. In any case we expect the effect of these terms to be small, except perhaps near the surface.

In summary, we approximate the inverse of the TLM \mathbf{L}_t^{-1} by integrating the TLM backward in time, that is, with a negative time step, without the dissipative terms (which are small except near the surface), or by changing the sign of these terms during the backward integration. If this approximate inverse is accurate, we should be able to approximately recover the initial perturbation $\delta\mathbf{X}_0$ from two model solutions at time t :

$$\delta\mathbf{X}_0 \approx \mathbf{L}_t^{-1}[\mathbf{M}_t(\mathbf{X}_0 + \delta\mathbf{X}_0) - \mathbf{M}_t(\mathbf{X}_0)], \quad (5)$$

where \mathbf{L}_t^{-1} represents the approximate or quasi inverse of \mathbf{L}_t .

Similarly, we can consider the perceived forecast error at time t given by

$$\mathbf{E}_t = \mathbf{M}_t(\mathbf{X}_0) - \mathbf{X}_t^a, \quad (6)$$

where \mathbf{X}_t^a is the verifying analysis valid at time t . Note that this is the perceived, and not the true, error, because the verifying analysis also contains errors, but beyond 12 h these are generally much smaller than the forecast errors. We would like to find a perturbation $\delta\mathbf{X}_0$ that would correct the forecast

$$\mathbf{M}_t(\mathbf{X}_0 + \delta\mathbf{X}_0) \approx \mathbf{M}_t(\mathbf{X}_0) + \mathbf{L}_t\delta\mathbf{X}_0 \approx \mathbf{X}_t^a, \quad (7)$$

so that we use the approximate inverse of the TLM to obtain

$$\delta\mathbf{X}_0 \approx \mathbf{L}_t^{-1}[\mathbf{X}_t^a - \mathbf{M}_t(\mathbf{X}_0)]. \quad (8)$$

Here $\delta\mathbf{X}_0$ will be denoted as “initial error estimate” from the quasi-inverse linear method. It is the solution obtained when we (approximately) trace the short-range perceived forecast error back to the initial time. Since, as discussed above, the small dissipative terms are irreversible and we cannot invert the TLM exactly, we denote our approximation of the inverse of the TLM as a “quasi inverse.” In section 3, we present numerical experiments that test the accuracy of both the linear tangent model and of its quasi inverse for the particular model used in the present study. We have to address at least two questions: 1) Is the linear evolution of the analysis error in the TLM close to the evolution of the analysis error in the nonlinear model, and 2) how accurate is the quasi-inverse linear error estimate, and, in particular, what is the impact on the quasi inverse of the simplified physical processes, which cannot be integrated backward?

3. The accuracy of the quasi-inverse linear error estimate for the NCEP global spectral model

a. The NCEP global spectral model, its TLM, and adjoint

The nonlinear atmospheric model used in this study is a lower-resolution version of the operational NCEP global spectral model, with horizontal triangular truncation of T62 and T28 vertical sigma levels. This model is based on the primitive equations formulated with a spectral discretization in the horizontal and an Arakawa quadratic conserving finite differencing in the vertical (Sela 1980; Sela et al. 1988; Kanamitsu et al. 1991). In order to take advantage of the spectral technique in the horizontal, a vorticity and divergence representation of the momentum equations is used to eliminate the difficulties associated with the spectral representation of vector quantities on a sphere. A semi-implicit time-integration scheme is applied to the divergence, temperature, and surface pressure equations. The vorticity equations are integrated explicitly except for zonal advection, which is treated implicitly. The model has a full set of physical parameterizations. New formulations of the cumulus convection and PBL parameterization were recently implemented (Pan and Wu 1995; Hong

and Pan 1996). The model used in this study is the T62/L28 version of the T126/L28 operational model implemented in January 1995, which is also used in the ensemble forecasting system and in the NCEP–National Center for Atmospheric Research 40-yr reanalysis (Kalnay et al. 1996).

The tangent linear model is a simplified adiabatic version (Navon et al. 1992), but it includes surface friction, horizontal diffusion, and vertical mixing. With these dissipative processes the TLM was shown to represent well the evolution of small perturbations in the nonlinear model with full physics (Pu et al. 1997). The full nonlinear model was used in the computation of the trajectory (basic state) used for all the linear and adjoint model integrations. The adjoint model was developed from this tangent linear model.

b. Test of the accuracy of the linear and quasi-inverse models

Since the analysis errors are not known, we cannot use forecast errors to test the accuracy of the quasi-inverse TLM method. Instead we chose arbitrarily two members of the operational T62 ensemble forecasting system starting from 1200 UTC 28 February 1996 (Toth and Kalnay 1993, 1996) for which we know exactly both the initial and the 24-h forecast differences. These known differences, which we can interpret as “errors,” allow us to test the accuracy of the quasi inverse of the TLM and the success of the method in correcting initial errors.

The forward TLM has been used by Pu et al. (1997), in the context of an adjoint sensitivity study. They showed that the agreement between the perturbation field from the linear and nonlinear integrations is good, and the results indicated that the dry-adiabatic linear

model with surface friction and vertical diffusion can reproduce fairly well the nonlinear perturbations of the model with full physics for short-range forecast.

In order to test the accuracy of the quasi-inverse TLM, the 24-h forecast difference between the two ensemble forecast members (at 1200 UTC 29 February 1996) was taken as the initial condition for the inverse integration and the TLM was integrated backward until the corresponding initial time (1200 UTC 28 February 1996) was reached, as discussed in section 2. Figure 1a illustrates the solution of this quasi-inverse, backward integration. It shows the linear perturbation at the initial time for the temperature and the u and v wind components at sigma level 13 (about 500 mb). Figure 1b shows the corresponding exact differences between the two ensemble members at forecast initial time. The two figures are in very good agreement, indicating that the quasi-inverse method is successful in tracing back the forecast differences to the initial condition differences in this case, at least above the lower boundary layer.

We also carried out the same experiments using the TLM without the diffusive terms, but the results showed that both the forward TLM and the quasi-inverse TLM become unstable without diffusion within a 1-day integration and, therefore, that these terms are needed to maintain computational stability.

Finally, we then took the linear estimate of the initial error obtained from the quasi-inverse method (as in Fig. 1) and integrated it forward 24 h with the TLM: Fig. 2a shows the obtained linear perturbation field at sigma level 13 (about 500 mb), and the corresponding 24-h fully nonlinear forecast difference is shown in Fig. 2b. The agreement between the two figures is still excellent, showing that the quasi inverse is a good approximation of the true inverse of the TLM.

We also compute an energy relative error measure of the combined quasi-inverse and TLM integration:

$$\Theta = \frac{\|\mathbf{LL}^{-1}[\mathbf{M}_t(\mathbf{X}_0 + \delta\mathbf{X}_0) - \mathbf{M}_t(\mathbf{X}_0)] - [\mathbf{M}_t(\mathbf{X}_0 + \delta\mathbf{X}_0) - \mathbf{M}_t(\mathbf{X}_0)]\|}{\|\mathbf{M}_t(\mathbf{X}_0 + \delta\mathbf{X}_0) - \mathbf{M}_t(\mathbf{X}_0)\|}, \quad (9)$$

where the $\|\cdot\|$ denotes the kinetic energy norm [Buizza et al. 1993; also see (10) in section 5a]. Figure 3 shows the vertical variation of the kinetic energy error ratio. There is a relatively large error in the lowest layer (almost 25%), indicating the effect of the friction term, whose sign was changed in the quasi-inverse TLM in order to maintain computational stability. For most of the atmosphere, however, the combined linear and quasi-inverse procedure is accurate to better than about 90% in the 1-day sensitivity test. We also computed the errors using the total energy norm, and the relative errors were very similar to those obtained using the kinetic energy

norm (Fig. 3), indicating that the TLM and quasi inverse maintain a state of quasigeostrophic balance.

4. Sensitivity of forecast error to initial conditions with quasi-linear inverse estimation

In the previous section we tested the accuracy of the linear and quasi-inverse approximations by comparing them with known differences between nonlinear integrations. In this section we test their ability to estimate errors and compare the linear and adjoint forecast sensitivity.

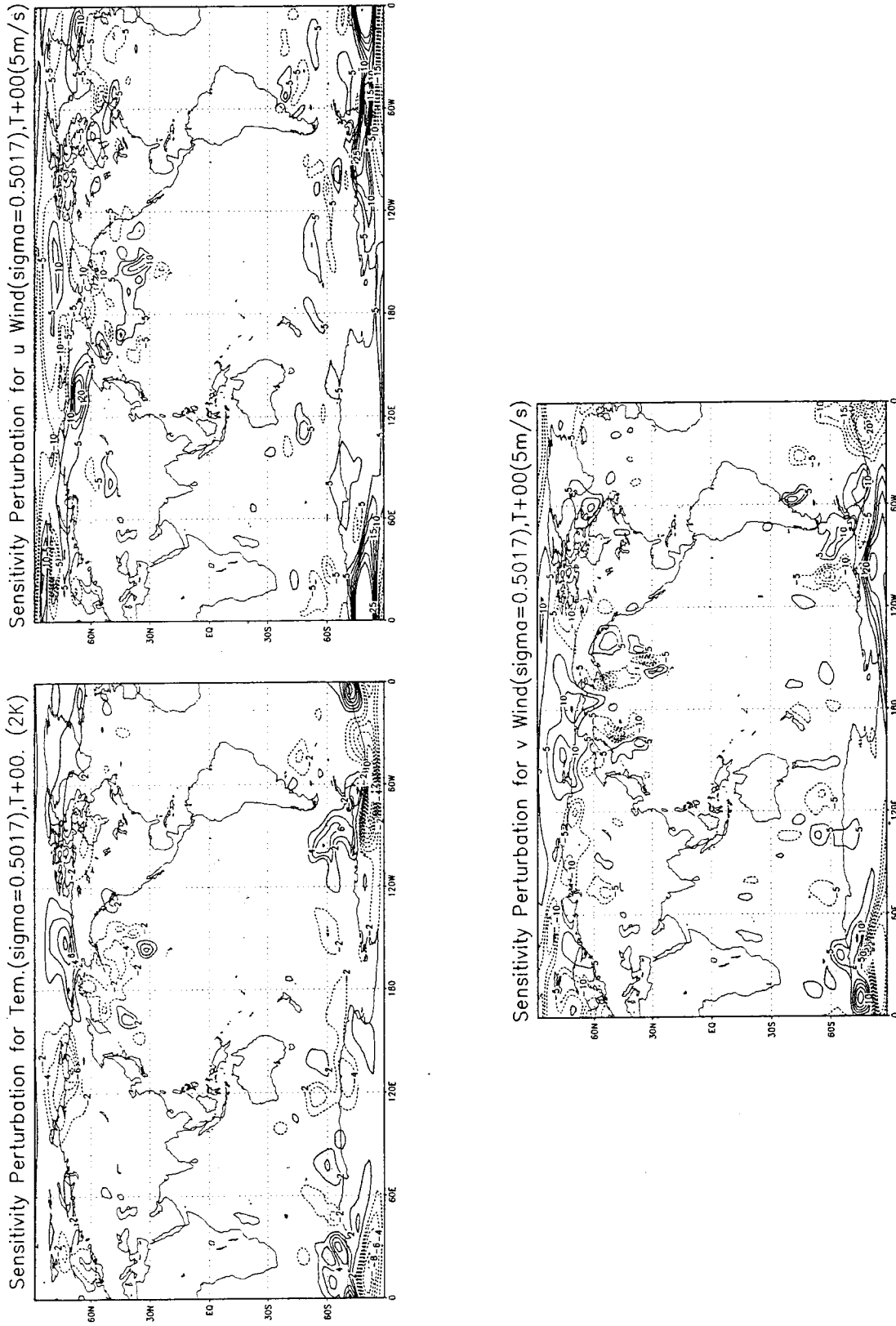


FIG. 1a. The initial perturbation for temperature and wind field at sigma level 13 (about 500 mb), calculated from the 24-h forecast differences between two ensemble members by using the quasi-inverse TLM method.

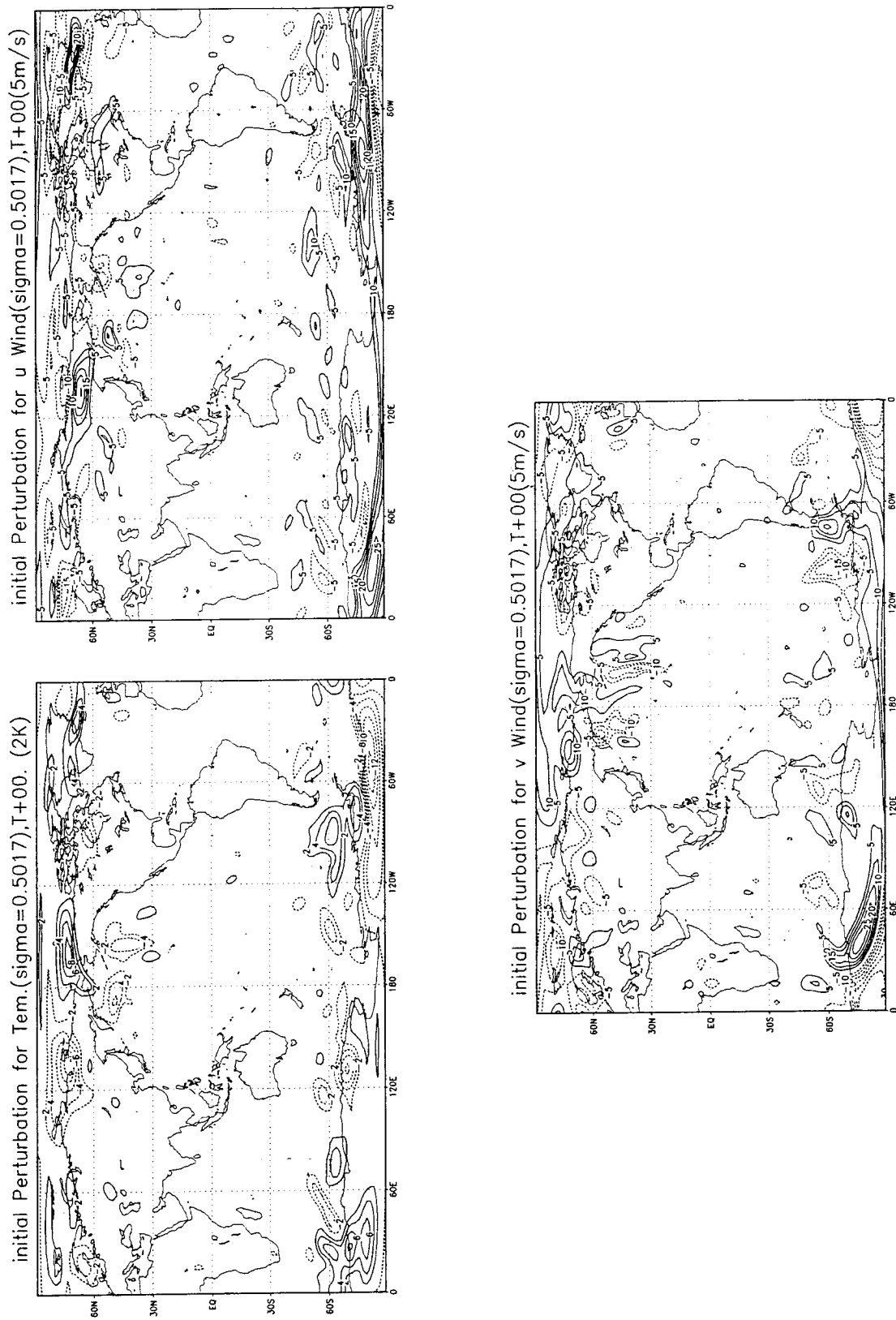


FIG. 1b. Same as Fig. 1a except the real difference between the two ensemble initial conditions.

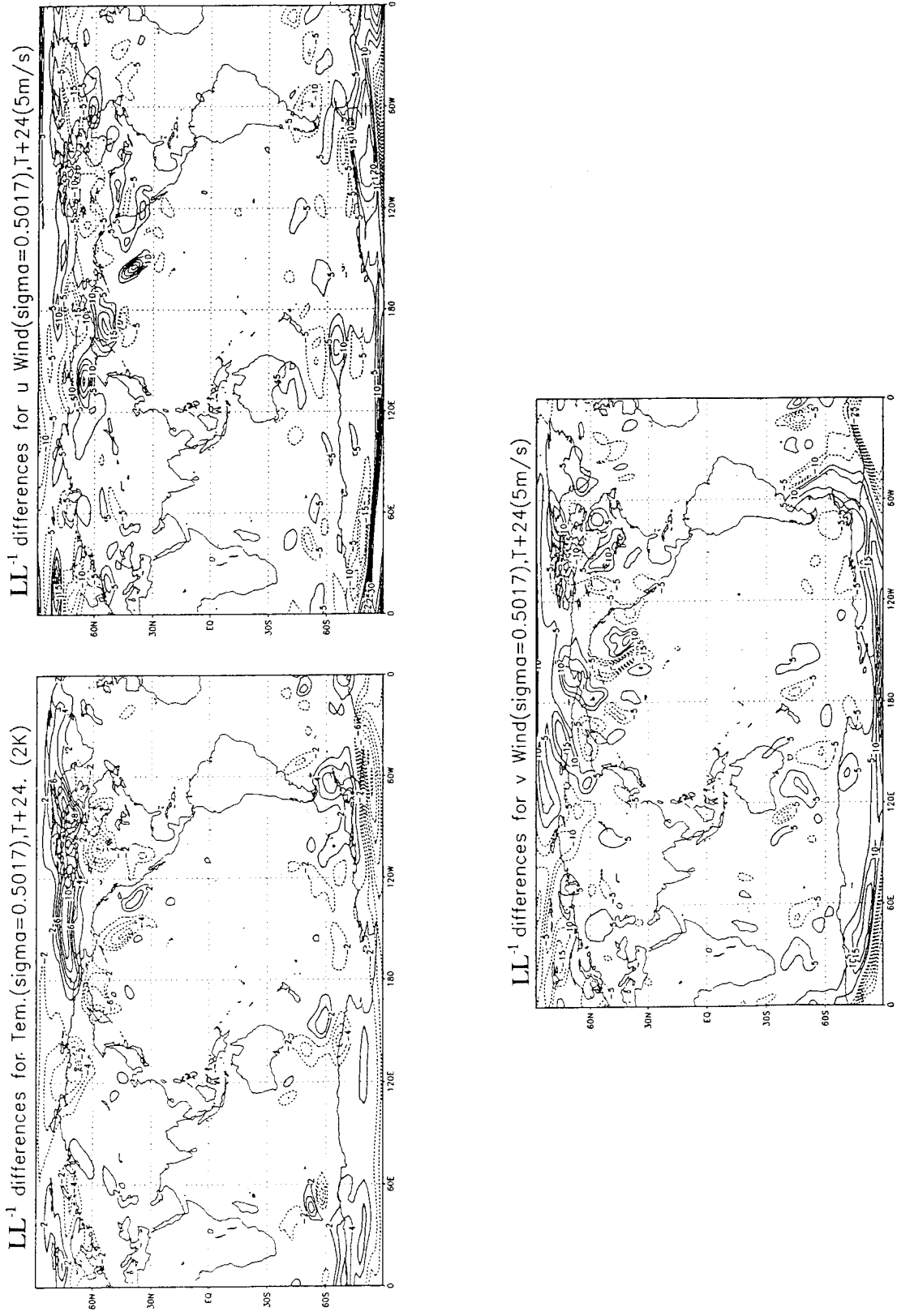


FIG. 2a. The 24-h linear evolution of quasi-inverse linear sensitivity initial perturbation.

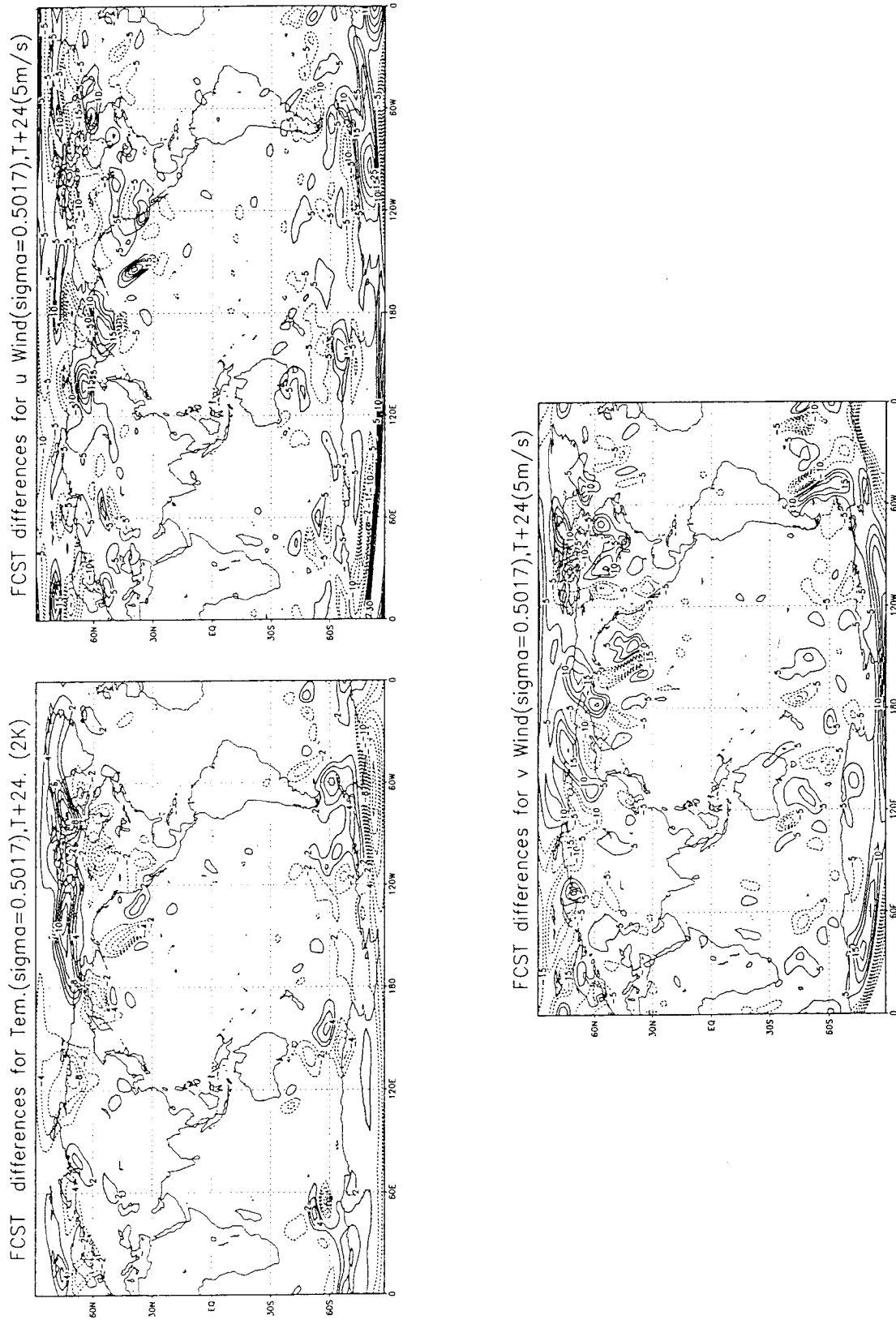


FIG. 2b. The 24-h forecast differences between the two ensemble members.

a. Quasi-linear and adjoint forecast sensitivity

We first tested the use of the quasi-inverse TLM to obtain improved initial conditions starting from an arbitrarily chosen analysis corresponding to 0000 UTC 24 March 1995. The 1-day perceived forecast error field (analysis minus forecast at 0000 UTC 25 March 1995) is used as an initial condition for the TLM backward integration. Note that only one backward integration is needed to obtain the TLM linear initial perturbation and that, unlike the adjoint sensitivity perturbation that provides the gradient of an error cost function, the results are not dependent on either the choice of a norm or on the amplitude that multiplies the gradient field.

In order to compare the inverse TLM estimation with the adjoint sensitivity, we performed the experiment using the two methods for the same case. The adjoint method is used as in Pu et al. (1997): a sensitivity initial perturbation minimizing the norm of the 1-day forecast error. The minimization process is performed iteratively using a conjugate-gradient method. Note that the cost of each adjoint iteration depends on the method used to estimate the optimal step size. If the step size is fixed at a value appropriate for many different cases, as in Rabier et al. (1996), the cost of each adjoint iteration beyond the first one is equivalent to about two times the cost of the quasi-inverse TLM iteration. If an optimal value of the step size is determined for each case (Derber 1987), then each iteration is about three times as costly as the quasi-inverse total computation. In this experiment we have followed the latter procedure.²

After one iteration of the adjoint method the initial error cost function was reduced by about 20%, and after five iterations by about 50%. Figure 4 shows the initial perturbation at 0000 UTC 24 March 1995 for the 500-mb geopotential heights, obtained from one adjoint iteration (Fig. 4a), five adjoint iterations (Fig. 4b), and the quasi-inverse TLM (Fig. 4c). The amplitudes of the adjoint perturbations are considerably larger after five iterations than after one iteration but are even larger for the quasi-inverse TLM perturbation (note that no zero line is plotted and a larger contour interval was used for the TLM perturbation). Close inspection of these fields reveals that there are many areas of the world where the shapes of the perturbations from the five-iteration adjoint and the inverse TLM methods are similar, although the amplitude of the latter tends to be larger (e.g., southeast Australia, southeast of South America, Alaska, Asia north of India, and others). On the other hand, there are other areas where the shape of

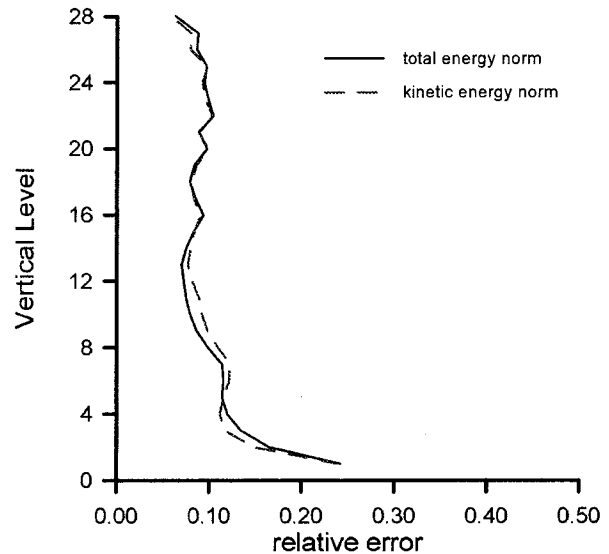


FIG. 3. The variation of the energy relative error with the vertical level.

the five-iteration adjoint and the inverse TLM perturbation are quite different, and even areas where the shape of the one- and five-iteration adjoint perturbations are also different. As discussed later in section 5d, it is not surprising that different perturbation patterns are obtained with the two methods, because the first iteration of the adjoint method retrieves patterns corresponding only to the fastest growing singular vectors, whereas the inverse TLM method recovers both growing and decaying vectors.

To assess the quality of the three different initial error estimates, that is, the extent to which they capture the origins of the forecast errors, we performed nonlinear model integrations from the corresponding perturbed (corrected) initial conditions to 0000 UTC 25 March 1995 and compared them with the original unperturbed (control) forecast. Figure 5 shows the 1-day forecast error (difference between the forecast and analysis field, at 0000 UTC 25 March 1995) for the 500-mb geopotential heights from these nonlinear model integrations starting from (a) the control analysis without corrections, and the analyses corrected with the initial errors estimated by (b) the inverse TLM error estimation, (c) the one-iteration adjoint sensitivity; and (d) the five-iteration adjoint sensitivity. Since all the experiments made use of the 1-day forecast error of Fig. 5a, it is not surprising that they all have achieved a reduction in the 1-day error, which was the original goal. It is clear from the figure that one iteration of the adjoint sensitivity method succeeds in improving the forecast error with respect to the control, and that five iterations are much better than a single adjoint iteration, but that the inverse TLM method gives by far the best results. A comparison

² In a routine operational context, the first adjoint iteration requires only a backward integration of the adjoint (assuming that the amplitude that multiplies the resulting sensitivity pattern is chosen using other information). This is because the 1-day error is already available as part of the operational suite, so that the first adjoint iteration costs about the same as the quasi-inverse method. Successive adjoint iterations require the computation of the updated 1-day error.

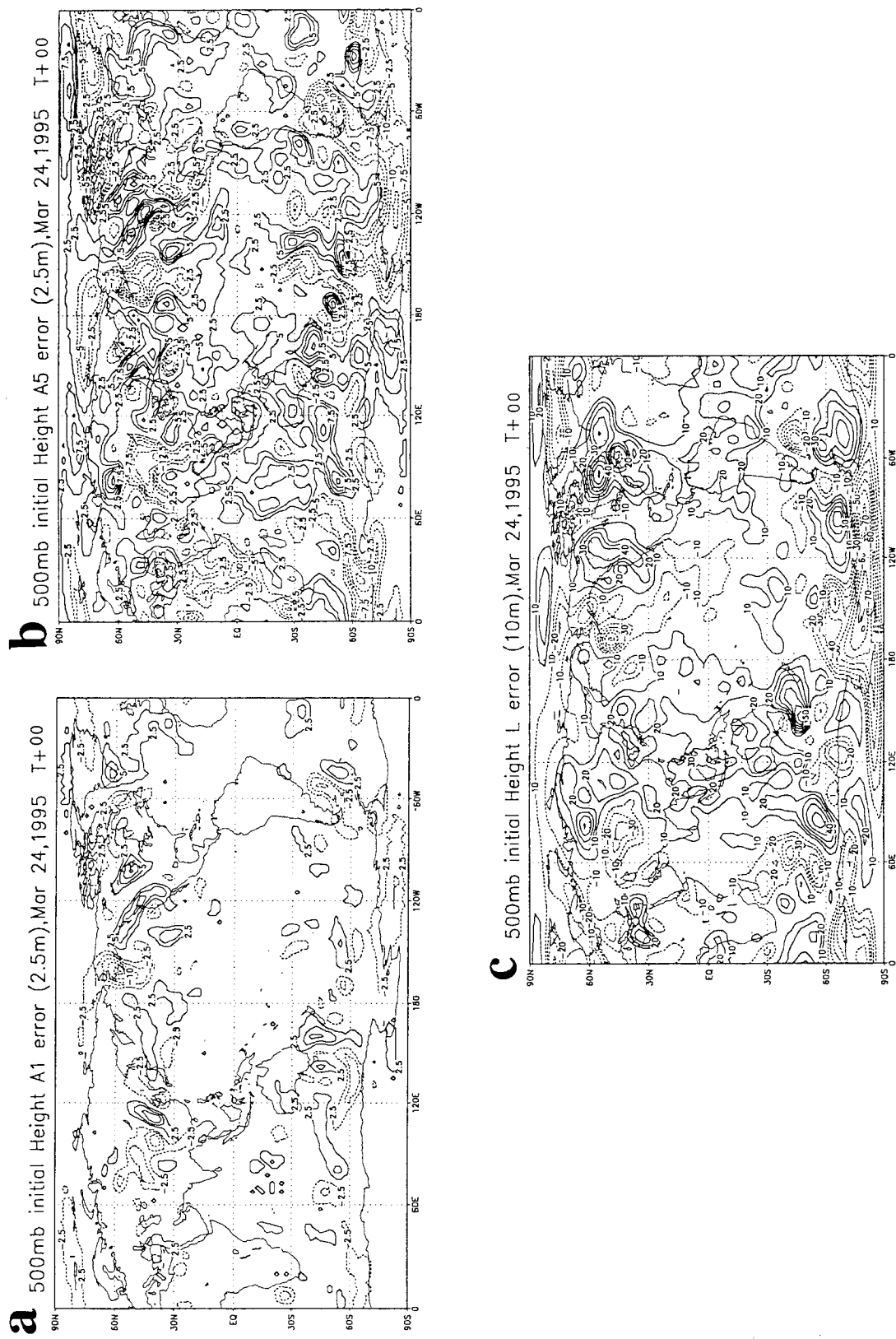


FIG. 4. Initial error estimate for 500-mb geopotential heights: (a) from one iteration of adjoint method, (b) from five iterations of adjoint method, and (c) from quasi-inverse linear estimation. The contour interval is 2.5 m for adjoint sensitivity but 10 m for linear sensitivity.

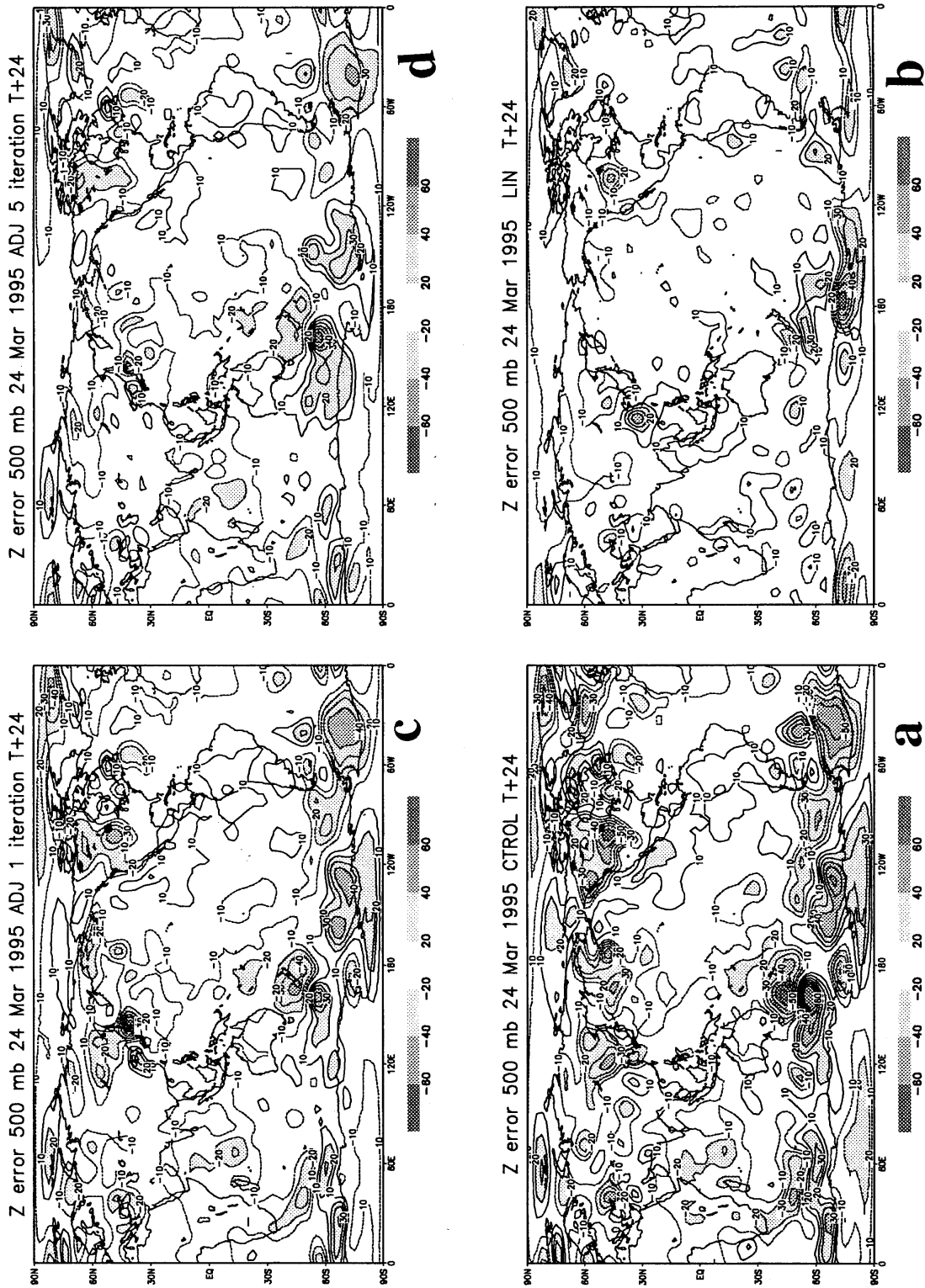


FIG. 5. The 24-h forecast geopotential heights error for 500 mb from the different sensitivity initial condition: (a) for control forecast, (b) for quasi-inverse linear estimation, (c) for one iteration adjoint, and (d) for five iterations adjoint. The forecast started from 0000 UTC 24 March 1995.

of other fields (not shown) yields similar conclusions (see also the discussion of Fig. 10 in section 5c).

Figure 6 shows a vertical east–west vertical cross section at 40°N of the height forecast error field for the control and the three improved initial conditions. It also indicates that the results from the linear sensitivity are substantially better than those of the adjoint method, even after five iterations. An examination of the structure of the error indicates that the quasi-inverse linear TLM perturbation reduces very substantially the original forecast error everywhere except at a few spots near the top of the model where the TLM may have problems associated with very fast growing modes (Kalnay and Toth 1996). The adjoint perturbations are able to reduce the forecast error in some areas, but *they actually increase significantly the original error in other areas* (e.g., near 150°E a very large new error structure is introduced with the one-iteration adjoint perturbation and only partially removed by the five-iteration perturbation). Similar results were observed for the wind forecasts (not shown). The reasons for this error may be due to the adjoint sensitivity method itself: As shown by Rabier et al. (1996), the first iteration in the sensitivity patterns is strongly related to the fastest growing singular vectors, each of which grows during both the backward (adjoint) and the forward integrations, so that they appear with amplitudes proportional to the square of their growth rate. On the other hand, within the adjoint method, which provides the gradient of the cost function, a single optimal step size must be chosen, which Rabier et al. selected to optimize the reduction of error corresponding to the fastest growing singular vector. The use of a single optimal step size cannot be optimal for all the dominant singular vectors and, therefore, may lead to a reduction of errors for some singular vectors but to an increase for others. The adjoint sensitivity will depend on the definition of the function and minimization technique.

Since the 1-day analysis is only an estimate of the true state of the atmosphere, and the perceived 1-day error was used in these calculations, it is necessary to make longer forecasts to test whether the apparent improvement in the forecast error is really meaningful.

b. Impact of estimates of the initial errors on medium-range forecasts

A medium-range weather forecast was performed from each of the perturbed initial conditions discussed above. Table 1 compares the 1–5-day sensitivities with the control forecasts' 500-mb heights anomaly correlation scores, verified against the corresponding analysis fields. We find that the sensitivity forecasts not only improve the 24-h forecast, but also improve the rest of the 5-day forecasts. The quasi-inverse TLM estimation results in the best forecasts, although, with five iterations, the adjoint sensitivity is close to it, es-

pecially in the Southern Hemisphere, where analysis errors are larger, and therefore the perceived forecast error may be less reliable (see also Fig. 10). The impact of the analysis errors on the perceived 24-h forecast errors is probably the reason why the forecast improvements introduced by the TLM are equivalent to about 36 h in the NH and only about 12 h in the SH. The results are similar for the five-iteration adjoint correction, but note that each adjoint iteration requires about 2–3 times the computations required by the TLM in total.

c. Improvements in original forecast quality from the estimates of the initial error

In order to further test the impact of the TLM inverse estimation and to compare the method with the adjoint approach, 14 consecutive cases from 0000 UTC 18 March 1995 to 0000 UTC 31 March 1995 were chosen for a comparison. For each case, we use the 1-day forecast error at 0000 UTC to trace back the error in initial condition at 0000 UTC the day before by quasi-inverse TLM estimation and by the adjoint method. As in Rabier et al. (1996) and in Pu et al. (1997), we performed only one iteration for the adjoint method in order to minimize the computational cost. The 5-day forecasts starting from linear sensitivity initial conditions and from the sensitivity of one iteration adjoint method were compared with the original control forecast. Figure 7a shows the anomaly correlation scores verified against the corresponding control analysis for 500-mb geopotential heights in the extratropics (20°–80°), and Fig. 7b shows the root-mean-square error for 200- and 850-mb wind speeds for the tropical area (20°N–20°S). The results show that in the extratropics the linear estimation method improves the original forecast in all but one case in each hemisphere and is also better than the one-iteration adjoint sensitivity forecast in all but 7 of the 28 cases. In the Tropics, the adjoint forecast tends to be close to the control, because tropical perturbations tend to be slowly growing and, therefore, do not dominate the adjoint perturbation (see Figs. 1–5), and the quasi-inverse TLM perturbation provides the best forecast in the majority of cases.

5. Characteristics of the different estimates of the initial errors

a. Total and kinetic energy

In section 3, the amplitude of the initial estimate of the initial error from quasi-inverse TLM estimation was compared with both one- and five-iterations adjoint initial sensitivity perturbations for 500-mb geopotential height fields. In this section, we compare the amplitude of the different initial perturbations at different vertical sigma levels for a single case. An energy norm is used to measure the amplitude of initial perturbations:

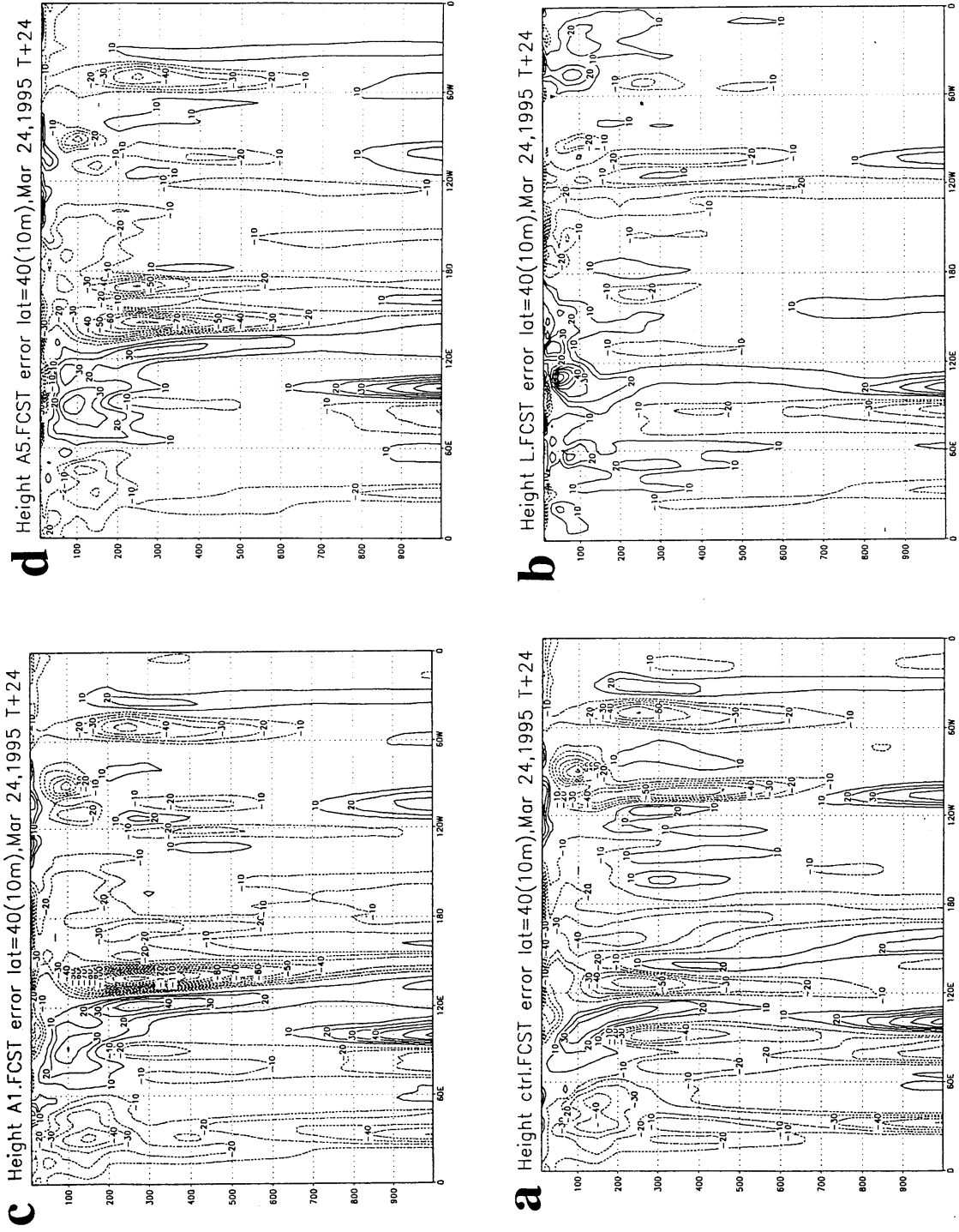


FIG. 6. Same as Fig. 5 except for a west-east vertical cross section at 40°N.

TABLE 1. Comparison of the forecast anomaly correlation scores for geopotential height field (1–20 waves, from 0000 UTC 24 March 1995).

Day	Northern Hemisphere 500 mb				Southern Hemisphere 500 mb			
	Control	Adjoint one iter.	Adjoint five iter.	TLM	Control	Adjoint one iter.	Adjoint five iter.	TLM
1	0.9887	0.9933	0.9953	0.9964	0.9804	0.9884	0.9903	0.9953
2	0.9649	0.9724	0.9823	0.9831	0.9516	0.9595	0.9712	0.9702
3	0.9286	0.9462	0.9656	0.9776	0.9119	0.9316	0.9413	0.9348
4	0.8800	0.9058	0.9449	0.9559	0.8471	0.8687	0.8960	0.8820
5	0.8520	0.8437	0.8988	0.8998	0.7691	0.7514	0.7860	0.7980

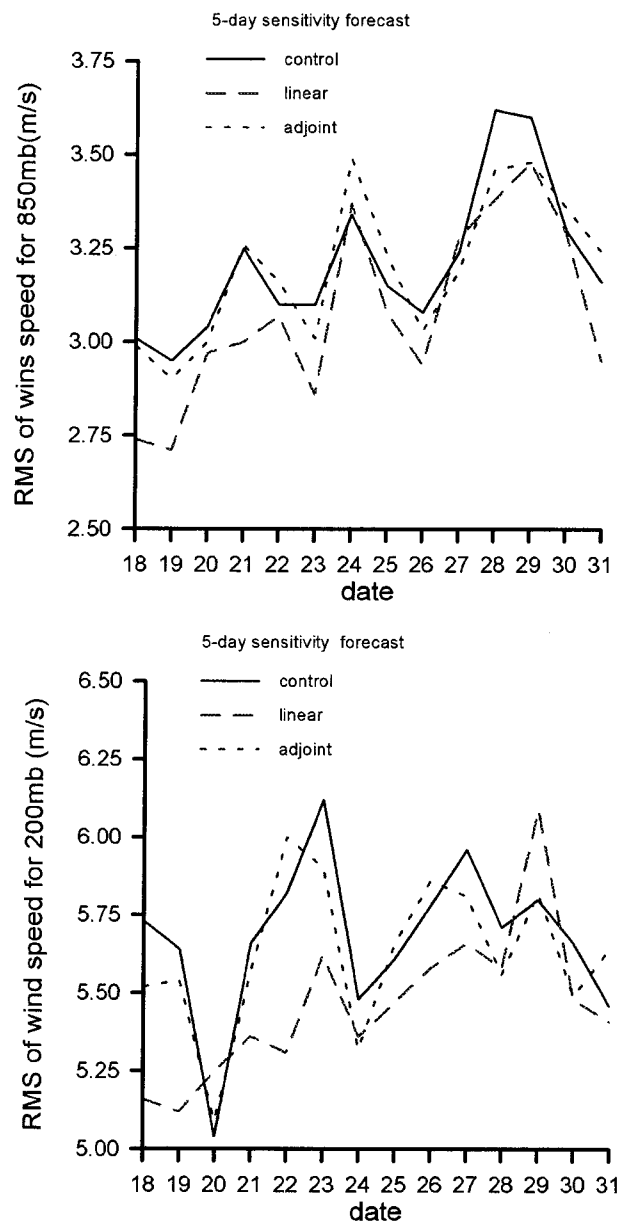
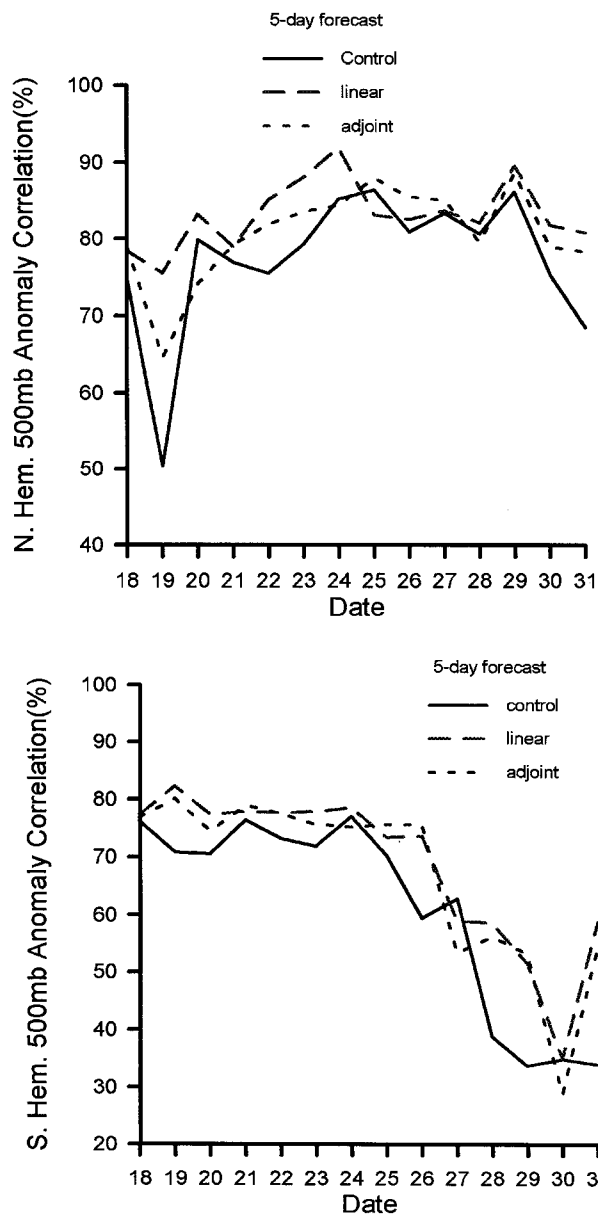


FIG. 7a. The 5-day forecast anomaly correlation scores for 500-mb geopotential heights. For control forecast (solid line), sensitivity forecast from adjoint one iteration (short-dashed line), and sensitivity forecast from inverse linear estimation (long-dashed line). Dates on the horizontal axis denote the starting dates of forecasts.

FIG. 7b. The 5-day forecast root-mean-square error for 850- and 200-mb wind speeds. For control forecast (solid line), sensitivity forecast from adjoint one iteration (short-dashed line), and sensitivity forecast from inverse linear estimation (long-dashed line). Dates on the horizontal axis denote the starting dates of forecasts.

$$E = \int_0^1 \int \int_{\Sigma} \left[\nabla \Delta^{-1} \zeta \nabla \Delta^{-1} \zeta + \nabla \Delta^{-1} D \nabla \Delta^{-1} D \right. \\ \left. + R_p T_r (\ln \Pi)^2 + \frac{C_p T^2}{T_r} \right] d\Sigma \left(\frac{\partial p}{\partial z} \right) d\eta, \quad (10)$$

where ζ , D , T , and Π stand for the perturbations of vorticity, divergence, temperature, and surface pressure; η is the vertical coordinate; T_r is a reference temperature; and R_p and C are thermodynamic constants. Here E is the total energy norm, and the first two terms in square brackets are the rotational and the divergent parts of the kinetic energy norm.

Figure 8a shows the vertical cross section of the total energy norm, and Fig. 8b the kinetic energy norm for the initial perturbation. It indicates that the magnitude of the TLM estimate of the initial error is the largest at all sigma levels. The patterns of the three curves are very similar, showing the maximum amplitude at mid-levels. Note that another large change of kinetic energy appears at the lowest level for the quasi-inverse TLM estimation, presumably due to the inaccuracy of the quasi-inverse at this levels, where the surface friction is most important and its sign has been changed for computational reasons. The one-iteration adjoint does not show this effect, but the five-iteration adjoint also has a substantial increase in both kinetic and total energy near the surface, which may be a result of the poor physics of the TLM and its adjoint.

b. Fit of the perturbed initial conditions and forecasts to rawinsondes

The rms fit and bias of both temperature (K) and vector winds (m s^{-1}) against rawinsonde data are presented in Fig. 9. In each figure, the curves on the left represent the bias, and the curves on the right the rms difference; the dashed curves indicate the fit of the control forecast, and the solid curves the fit of the perturbed forecast. Figure 9a presents the results for the initial time, and Fig. 9b for the 24-h forecast, for both the Tropics and the NH extratropics (the results for the SH, not shown, are similar to those of the NH). At the initial time the fit of the quasi-inverse analysis to the observations is worse than the control analysis. For example, at 300 hPa, the fit to the NH extratropics rawinsonde winds is about 6 and 7.5 m s^{-1} for the control and the quasi-inverse analyses, respectively. The assumed observational error standard deviation for the wind speed at jet level is about 4.6 m s^{-1} . Since the quasi-inverse procedure does not impose constraints to the fit to the data, it is not surprising that the fit is worse than in the control analysis.

In the Tropics, the adjoint method introduces negligible initial differences, even after five iterations. The quasi-inverse linear perturbations slightly improve the

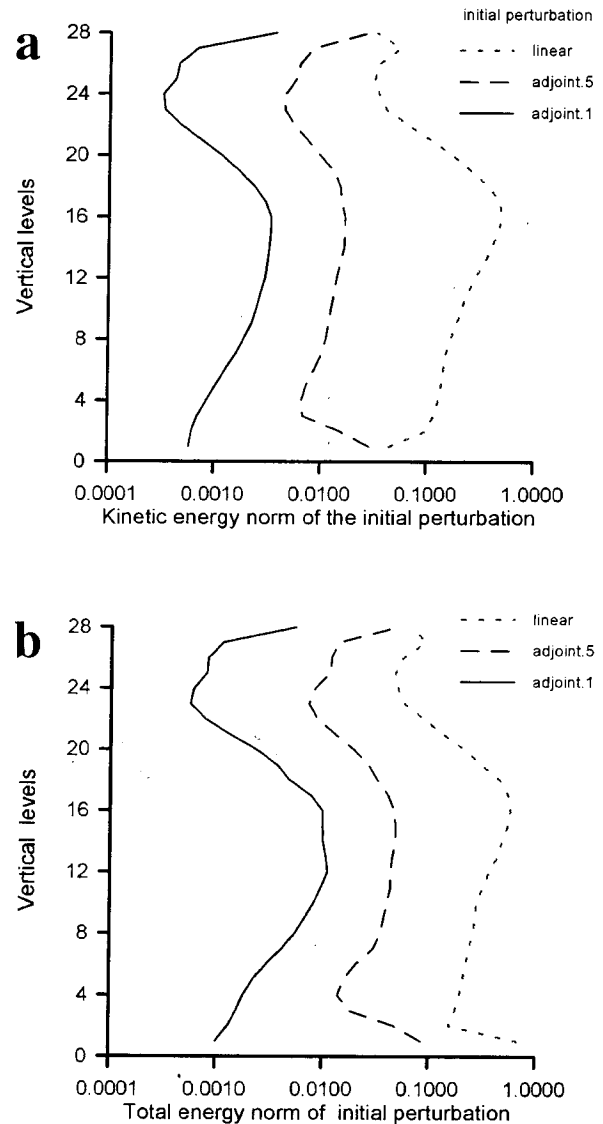


FIG. 8. The vertical cross section of the kinetic energy norm (a) and total energy norm (b) for sensitivity initial perturbation, of one iteration adjoint (solid), five iterations adjoint (long-dashed line), and inverse linear estimation (short-dashed line).

bias in the temperatures and winds but result in a significantly worse fit to the data compared to the control analysis, by about 0.1 K and 1 m s^{-1} in the temperature and wind, respectively. In the extratropics, a single adjoint iteration produces very small changes in the mid-troposphere and essentially no changes in the winds. After five iterations, the adjoint sensitivity increases by up to 0.4 K the fit to the data in the lower troposphere and changes in the wind of the order of 0.5 m s^{-1} but only below 700 hPa. The effect of the quasi-inverse linear sensitivity on the initial bias and rms fit in the extratropics is similar to that observed in the Tropics. It should not be surprising that the perturbed initial con-

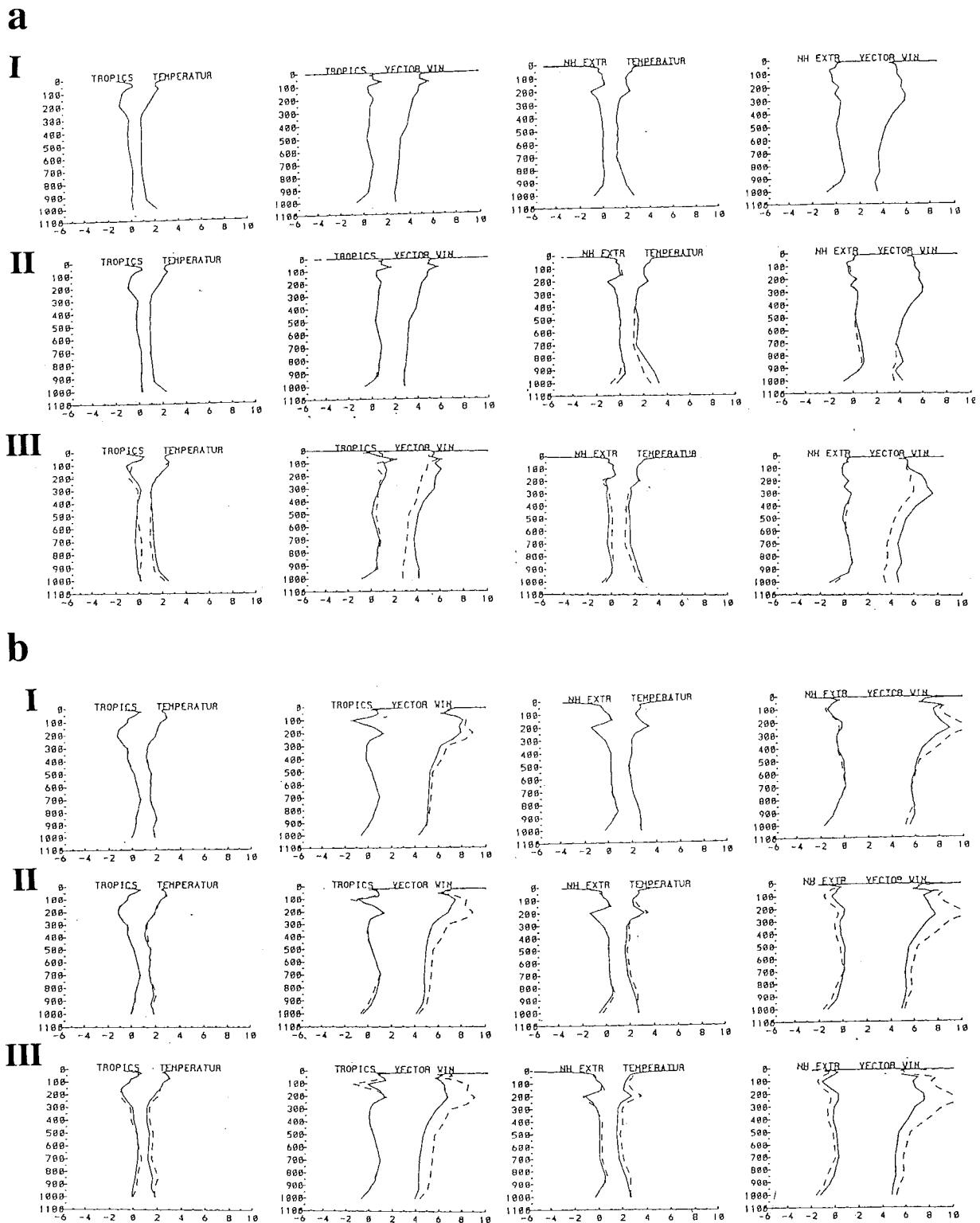


FIG. 9. The root-mean-square error of the sensitivity and forecast field fit the rawinsonde (observations) data. In each plot, the left two curves represent the bias, and the right two curves represent the rms error. Dashed line for the control field and solid line for the sensitivity field. The vertical axis denotes the pressure: (I) Adjoint one iteration (II), adjoint five iterations, and (III) quasi-inverse TLM. (a) For sensitivity initial condition, and (b) for one-day sensitivity forecast.

ditions tend to fit the data worse than the control analysis, which by definition tries to optimally fit the data and the first guess.

After 24 h the perturbed forecasts are better than the control forecasts in both the Tropics and the extratropics, except for the one-iteration adjoint, which fits the winds worse than the control in the low levels of the extratropics. The improvement after five adjoint iterations is larger than after one iteration, but the TLM improvement is comparable or larger. This relationship is maintained after 3 and 5 days (not shown). The improvement in the bias is also better for the quasi-inverse linear sensitivity forecast than for the adjoint forecast.

Checks of the fit of the initial conditions and forecasts against other data (aircraft reports, cloud track winds, satellite temperature soundings, and surface reports) gave similar results: the linear estimate of the initial error resulted in a fit to the data worse than the control analysis (which is designed to fit the data well), but the corresponding forecasts resulted in better fit to the data than either the control or the adjoint sensitivity forecasts.

c. Sensitivity corrections and forecast error

We now consider the evolution of the changes in the initial condition introduced by the sensitivity patterns and how they improve the nonlinear forecast with respect to the control. This is done by measuring how parallel the sensitivity correlation is to the control error. We define as $X = A(\text{analysis}) - C(\text{control forecast})$ the perceived control forecast error, and $Y = S(\text{sensitivity forecast}) - C$ is the sensitivity forecast correction. The angle between them is (e.g., Gill et al. 1981)

$$\alpha = \cos^{-1} \frac{(x, y)}{\|x\| \|y\|}, \tag{11}$$

where (\cdot) denotes an inner product and $\|\cdot\|$ is an Euclidian L_2 norm, in this case the total energy. If the sensitivity perturbations were able to perfectly correct the forecast, the sensitivity forecast correction would remain parallel to the control forecast error.

Figure 10 shows the variation of the angles with the forecast day. Note that at $t = 0$ the perceived control forecast error is zero, so that the corrections cannot be compared. At 24 h the quasi-inverse sensitivity correction is much more parallel to the control error than the adjoint one- and five-iteration sensitivity corrections. The advantage for the quasi-inverse method remains clear for the first 2 days, but then the angle increases quickly and, by day 5, it is close to the angle between the adjoint sensitivity forecast error and control error. The advantage of the quasi-inverse method may be explained by the fact that its perturbation includes all components of the error, both growing and decaying, as does the real forecast error, and therefore the correction and the error are more parallel. The adjoint sensitivity perturbations, as shown in Fig. 10, on

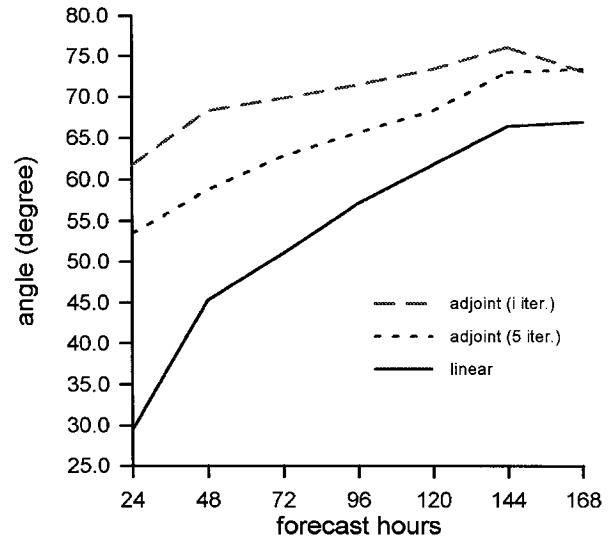


FIG. 10. The angle between forecast error and correction by sensitivity forecast. For one iteration adjoint (long-dashed line), five-iteration adjoint (short-dashed line), and quasi-inverse linear estimation (solid).

the other hand, contain only corrections in the fast growing errors and are therefore less parallel to the total error. After the first day or two, the growing errors dominate the total forecast error, and the advantages of the quasi-inverse linear procedure become less dominant. Since the quasi-inverse linear sensitivity recovers both growing and decaying errors in the initial conditions, and the adjoint sensitivity recovers only the fastest growing errors, it is not surprising that the amplitude of linear initial estimate of the initial error is significantly larger than the adjoint sensitivity perturbation (Figs. 8 and 9).

d. Sensitivity patterns, bred vectors, and singular vectors

To further interpret the different estimations of initial errors, let us look at formula (8) again. We will show that the perturbations obtained by the quasi-inverse linear method are strongly related to the bred vectors, whereas the perturbations obtained by the one-iteration adjoint are in turn related to the singular vectors [as shown first by Rabier et al. (1996)]. The quasi-inverse method attempts to find the initial perturbation δX_0 that inverts the following equation:

$$L\delta X_0 = \delta X_t, \tag{12}$$

where the rhs of (12) is the perceived error at time t (24 h in our experiments):

$$\delta X_t = X_t^a - M(X_0). \tag{13}$$

Let us assume for the moment that diabatic and dissipative effects can be neglected, which is a reasonable

assumption for short integrations. The adiabatic inviscid dynamics of the model are a Hamiltonian system (see appendix) and have properties that include conservation of volume in phase space and conservation of total energy. Other related properties are the fact that eigenvalues of the matrix \mathbf{L} appear in sets of 4 and that if λ is an eigenvalue, then $1/\lambda$ is also an eigenvalue. As an important consequence, stretching by the dynamics in a given phase-space direction is always compensated by a contraction in another direction, preserving the volume of phase-space elements.

This implies that the most unstable and stable directions in the phase space of the model dynamics exist always together: there are no large instabilities without large decays.

Therefore, an inverse adiabatic TLM model, and its numerical realization the quasi-inverse TLM, must reproduce not only the unstable initial perturbation, but also the decaying initial perturbation that feeds its energy to the growing structure. A straightforward consequence is that *the initial amplitude of the quasi-inverse TLM perturbations must necessarily be larger than if only growing perturbations were included*. Note, however, that *this larger amplitude does not shorten the time interval for which a (linear) TLM integration is valid*, since the decaying initial components dominate the larger amplitude. These results are an intrinsic property of the dynamics of the model, independent of any choice of coordinates or norm.

The quasi-inverse perturbations are related to the “bred” or leading local Lyapunov vectors used as initial perturbations in the ensemble forecasting at NCEP since 1992 (Toth and Kalnay 1993, 1996; Kalnay and Toth 1996). The breeding method used to generate perturbations simulates the development of growing errors in the analysis cycle, where errors grow during a 6-h forecast, and then they are “squashed down” by the analysis, which combines the short-range forecast with observations.³ The bred perturbations are obtained as the difference between two nonlinear forecasts; this difference is carried forward upon the evolving atmospheric analysis field and scaled down at regular intervals. Breeding is a nonlinear generalization of the method used to generate leading Lyapunov vectors, where only a linear model is used for propagating the perturbations. As long as their amplitude remains small, and the phys-

ical processes are not dominant, the bred vectors generated by the difference between two nonlinear forecasts can be considered as the result of a forward integration of the TLM as in (12). Therefore, to the extent that forecast errors at time t are dominated by growing errors (leading local Lyapunov vectors), the inverse method will result in the Lyapunov vectors at time $t = 0$. Of course, in addition to these vectors, there are decaying vectors as well in the analysis error. These decaying errors, when integrated backward with the TLM or its quasi inverse, will result in larger amplitudes at the initial time than at the final time.

In order to clarify the relationship between the quasi-inverse and the adjoint perturbations, let us scale the model variables so that the total energy norm becomes the Euclidian norm; that is, $\|\mathbf{X}\|^2 = \mathbf{X}^* \mathbf{X}$. In what follows we use the same notation for the transformed variables as that which was used for the variables before the transformation. If we expand the final perturbation field on the basis formed by the singular vectors (eigenvectors of $\mathbf{L}^* \mathbf{L}$ after the model integration to the final time), an inverse model integration is formally

$$\delta \mathbf{X}_0 = \mathbf{L}^{-1} \delta \mathbf{X}_t = \sum_{i=1}^n \frac{\mathbf{v}_i^*(t) \delta \mathbf{X}_t}{\sigma_i} \mathbf{v}_i(0), \quad (14)$$

while an adjoint integration is

$$\delta \mathbf{X}_0 = \mathbf{L}^* \delta \mathbf{X}_t = \sum_{i=1}^n [\sigma_i \mathbf{v}_i^*(t) \delta \mathbf{X}_t] \mathbf{v}_i(0), \quad (15)$$

where $\mathbf{v}_i(0)$, $\mathbf{v}_i(t)$, $i = 1, \dots, n$ denote the singular vectors at the initial and final times, respectively, and σ_i are the corresponding singular values.

Comparing (14) and (15), we see that *the adjoint integration overemphasizes the growing components and underestimates the role of the decaying components of the true inverse by a factor of σ_i^2* . This occurs for any choice of norm, although the structure of the fastest growing singular vectors itself is strongly dependent on the choice of norm. For this reason, as pointed out by Rabier et al. (1996), the gradient of the cost function (error energy) computed by one iteration of the adjoint method is dominated by the leading singular vectors, and the structure of the adjoint sensitivity pattern is also very similar to that of (energy norm) leading singular vectors, which are used at ECMWF as initial perturbations in their ensemble forecasting system (Molteni et al. 1996).

In general the minimization of a quadratic function requires multiple iterations, unless the problem is perfectly conditioned (all the eigenvalues of the minimization algorithm are equal). One adjoint minimization provides the gradient of the cost function, that is, the shape of a small perturbation that results in a maximum change of the cost function. But the full adjoint minimization problem needs multiple iterations depending on the conditioning number (ratio of the largest and

³ We have done additional experimentation by taking as “analysis error” the difference between an operational and a parallel (experimental) analysis cycle. We found that the backward integration of the “24-h analysis error” had significant similarity with the (known) “0-h analysis error”, that is, the difference between the operational and parallel analyses at $t = 0$ h. This supports the basic assumption of the breeding method, which is that, to a large extent, the analysis errors are carried forward by the model first guess, thus “breeding” Lyapunov vectors into the analysis error. It does not support the assumption of white noise errors in the analysis, which is used in the development of singular vectors.

smallest eigenvalue). If all the eigenvalues σ_i^2 of L^*L were equal, then the adjoint minimization (15) would find the same solution as the inverse approach. But in fact, there are eigenvalues greater and smaller than one, corresponding to different growing and decaying singular vectors.

If the model was exactly reversible, then finding the initial field with zero 24-h forecast error would be a deterministic problem, and the solution would be unique. In this case, if the iterated adjoint minimization converged to the lowest possible minimum (i.e., the field that results in a 24-h forecast with zero error energy), it would have to reach necessarily the same result as the exact inverse procedure, that is, find the true initial error.

In reality, of course, the model, even if linearized, is not exactly reversible, as discussed earlier in the paper. For this reason we are forced to use an approximate inverse or “quasi-inverse” model, in which the dynamics are maintained exactly as in the true inverse, but diffusive terms are handled as in the adjoint method. This approach gives good results, as shown in previous sections, because during short integrations the role of dissipation and diffusion is quite small, except near the surface.

In summary, the adjoint procedure (with a single iteration) can be considered optimal, in the sense that by determining the gradient of the error energy with respect to initial perturbations, it finds the smallest perturbation that results in a maximum decrease of the error energy. The quasi-inverse integration, on the other hand, finds a close approximation to the *exact* initial perturbation that corrects the perceived 1-day forecast error. If successful, the iterated adjoint scheme would converge to the same exact solution.

6. Possible improvement of future forecast skill using initial error estimates

Our results showed that the linear sensitivity patterns can improve substantially the original forecast, even beyond the period for which the error was computed (section 3). But a crucial question for the operational practice is whether we can use this procedure to improve the skill of future forecasts. For possible operational applications, we should compare the sensitivity forecast with the regular forecast from latest initial conditions. This means, in our experiments, that we should compare the 5-day sensitivity forecast scores with the 4-day forecast from the original analysis field, since the 1-day analysis data needed for the computation of the sensitivity perturbation introduced a 1-day delay. Figure 11 compares the anomaly correlation scores for 500-mb geopotential heights and shows that although the 5-day sensitivity forecast is much better than the original 5-day control forecast, it is still worse than the corresponding 4-day control forecast in most cases. The same results were also obtained by using

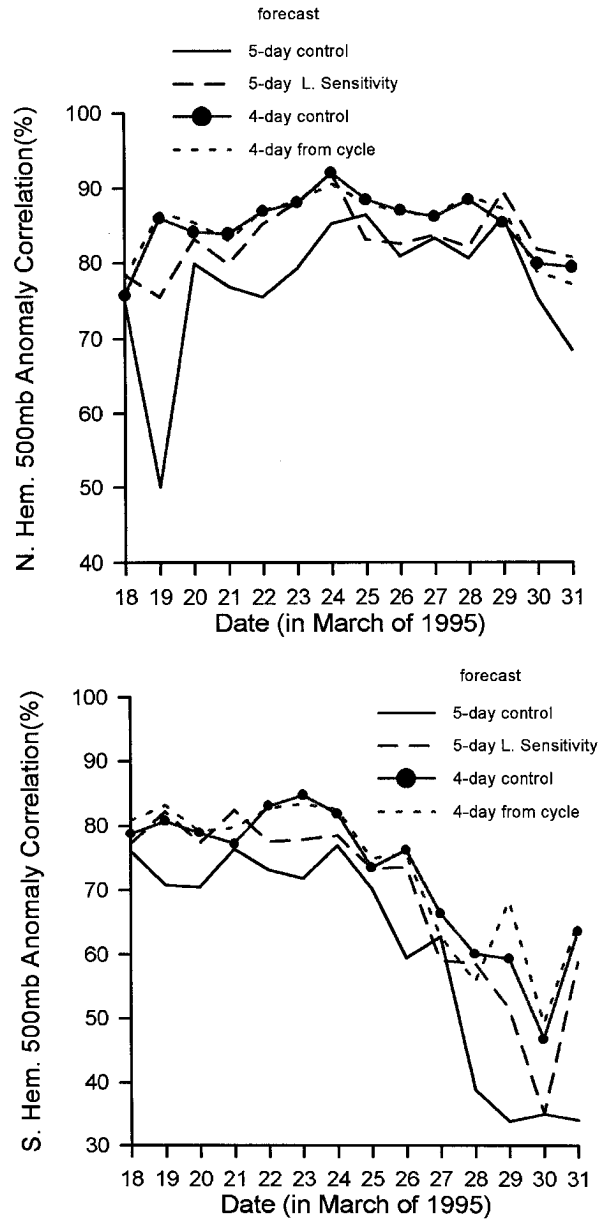


FIG. 11. The anomaly correlation scores for 500-mb geopotential heights, for 5-day sensitivity forecast, corresponding 4-day control (operational) forecast, 5-day control forecast, and for the 4-day forecast from the new cycle (iterated). Dates (March 1995) on the horizontal axis denote the starting dates of forecasts.

the adjoint sensitivity perturbation (Pu et al. 1996, 1977; Rabier et al. 1996). This result suggests that the latest control forecast, based on the NCEP analysis–forecast cycle, makes better use of the data available every 6 h than the sensitivity forecasts, which are 1 day longer and only use the information present in the latest analysis. However, the fact that the sensitivity 5-day forecasts are better than the original 5-day forecasts also indicates that they start from a better initial

analysis than the control 5-day forecast. In order to improve the future forecast skill, Pu et al. (1997) suggested a technique that takes advantage of both the better starting point for the 5-day forecast provided by the sensitivity analyses and of the use of the data in the latest day by the analysis–forecast cycle. This technique referred to as “iterated cycle” can be described as follows. (a) At 0000 UTC today, calculate sensitivity perturbation for initial condition at 1 day ago ($t = -24$ h) from the present (today) 1-day forecast error. (b) Adjust the 1-day-old ($t = -24$ h) initial condition by using the sensitivity perturbation. (c) Using this adjusted new (and presumably better) initial condition at $t = -24$ h as a starting point, repeat the NCEP three-dimensional analysis system SSI cycle every 6 h, until the present time (0000 UTC today) is reached, so that a new analysis field is obtained for today. (d) Perform the medium-range weather forecast from this new obtained analysis field.⁴ Following this procedure, Pu et al. showed that the iterated cycle with adjoint sensitivity perturbations was an improvement of the future forecast skill: the medium-range forecast from the iterated cycle was better than the original corresponding forecast. In a similar way, we now perform an iterated cycle by using the quasi-inverse TLM initial error estimate. As can be seen from Fig. 11, and the new 4-day forecasts from this iterated cycle are better than the corresponding control forecasting in several cases, especially in the Southern Hemisphere. Figure 12 shows the 14-case average of forecast anomaly correlation scores, which verified against the control forecast for 1–5-day forecast 500-mb geopotential heights. The iterated cycle leads to a small improvement in the medium-range weather forecasts, especially in the Southern Hemisphere, indicating that it might be possible to use the quasi-inverse TLM initial error estimate to improve future forecast skill. However, we have not addressed here the problem that the background error covariance should be modified, since we have inserted new structures (the error corrections) into the background. The results of the iterated cycle would presumably improve if this was properly taken into account.

7. Summary and discussion

We have presented a quasi-inverse linear method to study the sensitivity forecast errors to initial conditions for the NCEP global spectral model. The inverse is approximated by running the tangent linear model (TLM) of the nonlinear forecast model with a negative time step, but reversing the sign of friction and diffusion

⁴ The “iterated cycle” (Pu et al. 1996, 1997) is very similar to the “poor man’s 4D VAR” developed independently by Huang et al. (1996) at HIRLAM. At NCEP, however, it is denoted “poor woman’s.”

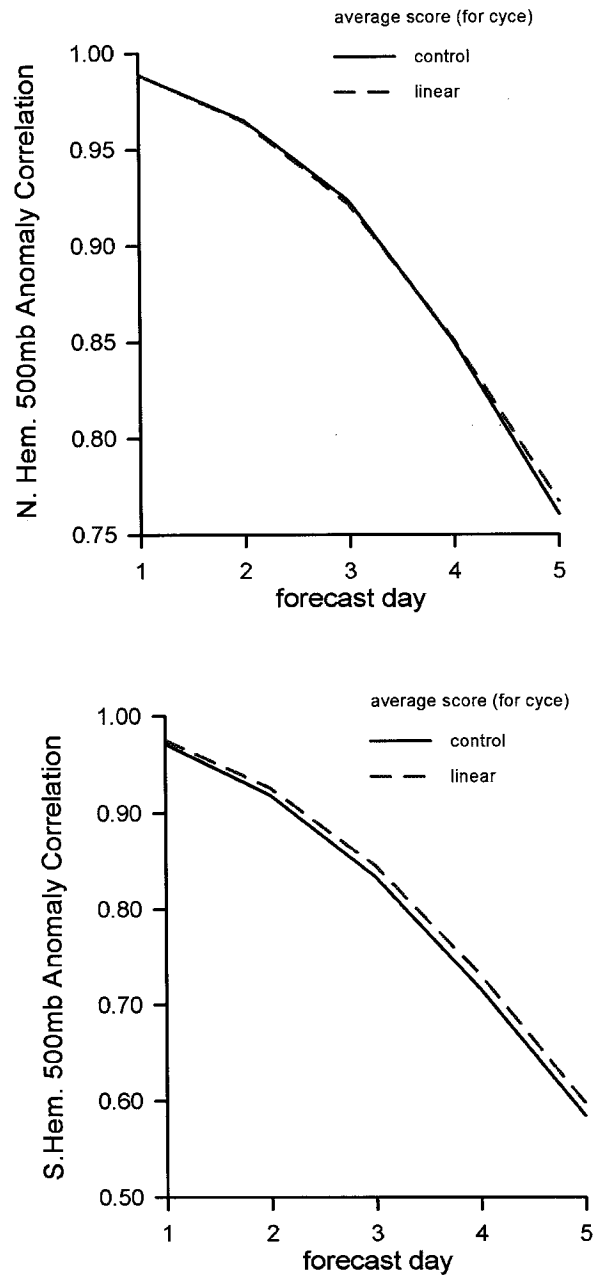


FIG. 12. Comparison of the average anomaly correlation scores of 1–5-day forecast for 500-mb geopotential height between the iterated cycle (dashed) and control forecast (solid). Starting dates of the forecasts ranged from 18 March 1995 to 31 March 1995.

terms (in the same way as in an adjoint integration). This avoids the computational instability that would be associated with these terms if they were run backward. As done using the adjoint model integrations, we started the quasi-inverse TLM at the time of the verified forecast error and integrated backward to the corresponding initial time. However, instead of minimizing an error cost function through successive iterations, as done with

the adjoint sensitivity method, the quasi-inverse linear perturbation is obtained by a single, deterministic integration. This has the advantage that it is faster and does not depend on the choice of the norm used in the definition of the error cost function.

A numerical experiment performed using a known perturbation from the NCEP ensemble shows that this quasi-inverse linear estimation is able to trace back the differences between two perturbed nonlinear 1-day forecasts and recover with good accuracy the known difference between the two forecasts at the initial time. The results show that both the linear estimation and the quasi-inverse linear estimation are quite close to the nonlinear evolution of the perturbation in the nonlinear forecast model, suggesting that we should be able to apply the method to study the sensitivity of forecast errors to initial conditions.

We then perform experiments tracing back actual forecast errors. We calculate the perturbation field at the initial time (linear initial error estimate) by using perceived 1-day forecast errors as initial conditions for a backward integration using the quasi-inverse TLM. As could be expected from the previous experiments, when the estimated error is subtracted from the original analysis, the new initial conditions lead to an almost perfect 1-day forecast. The forecasts beyond day 1 are also considerably improved, indicating that the initial conditions have indeed been improved.

We then compare the quasi-inverse linear sensitivity method with the adjoint sensitivity method (Rabier et al. 1996; Pu et al. 1997) for medium-range weather forecasting. We find that although both methods are able to trace back the forecast error to sensitivity perturbations that improve the initial conditions, the forecast improvement obtained by the quasi-inverse linear method is considerably better than that obtained with a single adjoint iteration and similar to the one obtained using five iterations of the adjoint method. This is true even though each adjoint iteration (except the first one) requires at least twice the computer resources of the inverse TLM estimation. As indicated above, the quasi-inverse TLM estimation method does not depend on the definition of a norm, it does not require the estimation of an optimal step size, and it provides an optimal correction throughout the globe.

We point out that (as shown by Rabier et al. 1996) the adjoint forecast sensitivities are closely related to singular vectors. In fact, the adjoint sensitivities show several characteristics also observed in the singular vector behavior (Szunyogh et al. 1997). In the initial perturbations, the wind perturbations are rather small compared to the temperature perturbations, and they are maxima at relatively low levels. After 1 day, however, the maximum winds of the adjoint perturbations grow and are observed closer to the tropopause levels.

The quasi-inverse linear sensitivities are also related to ensemble perturbations, the bred (Lyapunov) vectors used for ensemble forecasting at NCEP (Toth and Kal-

nay 1993). If the final error is a Lyapunov vector, the inverse method will also yield an exact Lyapunov vector (except for the effect of changed sign in the dissipative terms). However, since the analysis errors also contain decaying errors, these will be magnified during the backward integration. This is the reason why the quasi-inverse perturbations are much larger than those obtained with the adjoint method. We point out that the adjoint procedure (with a single iteration) can be considered optimal, in the sense that by determining the gradient of the error energy with respect to initial perturbations, it finds the smallest perturbation that results in a maximum decrease of the error energy. The quasi-inverse integration, on the other hand, finds a close approximation to the *exact* initial perturbation that corrects the perceived 1-day forecast error. *If successful, the iterated adjoint scheme would converge to the exact same solution.*

Finally, the possibility of the use of the initial error estimate to improve future forecast skill is discussed, and preliminary experiments encourage us to further test this rather inexpensive method for possible operational use. Although the results are somewhat positive, this method would have to address the problem that the data is effectively used twice, and the inverse method introduces perturbations in essentially all degrees of freedom in the model, whereas the iterated analysis using adjoint perturbation method (Pu et al. 1997) only introduced changes into a few degrees of freedom (the leading singular vectors).

Acknowledgments. We would like to thank Drs. Zoltan Toth and O. Talagrand for helpful suggestions and Drs. Milija Zupanski and J. Purser for their suggestions in the preparation of the manuscript. The thoughtful and constructive comments by two anonymous reviewers are also gratefully acknowledged. The first author is supported by the UCAR/NCEP Visiting Scientist Program.

APPENDIX

Growing and Decaying Perturbations in Hamiltonian Systems

The system of the inviscid adiabatic primitive equations has a Hamiltonian structure (e.g., Shepherd 1990). We can therefore assume that the local (in phase space) dynamics of the model is Hamiltonian. On the timescale used in this paper (24 h) this approximation is quite realistic for synoptic-scale motions, although it does not hold in a strict sense due to (i) the presence of diabatic and dissipative processes, (ii) the lack of Coriolis force terms related to $l=2\Omega \cos\phi$, and (iii) the application of non-structure-preserving spatial and temporal discretization schemes. Some of the considerations discussed below are valid only for this idealized system, but we will demonstrate that they describe well the local (in phase space) qualitative behavior of the GCMs.

Hamiltonian systems have an amazing variety of symmetries, among them that the eigenvalues of the matrix \mathbf{L} appear in quadruples [see in Olver (1993), Theorem 6.22, and Arnold (1989), 226], and that if λ is an eigenvalue, then $1/\lambda$ is also an eigenvalue. As an important consequence, stretching in a given phase space direction is always compensated by a contraction in another direction, preserving the volume of phase space elements. The phase space is incompressible (Liouville's theorem). Therefore, the most unstable and stable directions in the phase space are twins; there are no large instabilities without large decays. If we only take into account growing directions [as done in the adjoint sensitivity, or in the pseudoinverse approach, it means that we are creating an expanding phase space, that is, a forced (but not dissipative) physical system. The intrinsic geometry of the phase space, and hence the important symmetry of the eigenvalues, are generated by the dynamics, so that they are independent of any free choice of local coordinates and norm. The time evolution of bred and quasi-inverse TLM perturbations attempts to simulate this norm-independent stretching–contracting property of the dynamical system.

The quantification of stretching (contraction), however, requires the use of local Euclidean coordinates and a corresponding norm. Recall that if $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on the space of real n -dimensional vectors \mathcal{R}^n , then there exist positive constants, c_1 and c_2 , such that

$$c_1\|\mathbf{X}\|_a \leq \|\mathbf{X}\|_b \leq c_2\|\mathbf{X}\|_a, \quad (\text{A1})$$

where $\mathbf{X} \in \mathcal{R}^n$. It means that the two norms are equivalent on \mathcal{R}^n only when the vector is infinitesimally small or infinitely large, but one may get substantially different results with the use of different norms measuring finite-time instabilities. Indeed, since

$$\frac{c_1\|\delta\mathbf{X}_t\|_a}{c_2\|\delta\mathbf{X}_0\|_a} \leq \frac{\|\delta\mathbf{X}_t\|_b}{\|\delta\mathbf{X}_0\|_b} \leq \frac{c_2\|\delta\mathbf{X}_t\|_a}{c_1\|\delta\mathbf{X}_0\|_a}, \quad (\text{A2})$$

a perturbation can be growing and decaying at the same time when it is measured by two different norms.

A quantity whose conservation is closely related to the preservation of phase volume may appear as a reasonable candidate to be a proper norm. The implementation of this strategy, however, may lead to unresolvable difficulties for a particular system. Suppose that the energy of the system is defined by the quadratic form $\mathbf{X}^*\mathbf{X}$, then the quantity $(\mathbf{X} + \delta\mathbf{X})^*(\mathbf{X} + \delta\mathbf{X}) - \mathbf{X}^*\mathbf{X}$ is conserved but not quadratic, while $(\delta\mathbf{X})^*(\delta\mathbf{X})$ is quadratic but not conserved. In other words, the first candidate is not a useful norm, while the second one is not conserved. Perhaps the best choice for a norm is the sum of the kinetic and the available potential energy of perturbations, the pseudoenergy, which is quadratic, and also conserved if the basic state is steady (Shepherd 1990, 1993). An example of a norm of this type is the total energy, which is an unbiased norm but still not conserved for a time-dependent background flow.

These arguments suggest that the one-iteration adjoint method may generate unrealistic initial patterns because the physical quantity used as a norm is not conserved in the perturbed nonlinear system. However, there is a less obvious but more important problem. Toth et al. (1996) computed the most unstable and stable singular values and vectors for the T10 L18 version of the NCEP Medium Range Forecast Model with respect to the total energy norm [the estimated full spectrum is shown in Toth and Kalnay (1996)]. They found that most of the first and the last few hundred singular values have the symmetry of being in a reciprocal relation to each other. However, *the most unstable singular vectors were unrealistically far from thermal wind balance*. This suggests that the associated initial state is a state of the system that is possible but highly unlikely.

This is a typical problem with high-dimensional Hamiltonian systems. To show an extreme case of the danger of obtaining highly unstable but very improbable initial states through the use of an arbitrary norm, we borrow an example that Arnold (1989) used to exhibit a paradoxical conclusion from Poincaré's recurrence theorem: "If you open a partition separating a chamber containing gas and a chamber with vacuum, then after a while the gas molecules will again collect in the first chamber." The resolution of this paradox is that the gas molecules can indeed move back to the first chamber, but the probability of this event occurring is less than $1/(\text{duration of the solar system's existence})$. Let us open the door between the two chambers and use the number of molecules in the left chamber as the dynamical variable of the system. The absolute value of the perturbation of this number can also be used as the norm of the system. Intuitively, the most unstable initial perturbation that such a norm would generate occurs when most of the molecules are moved to the right chamber, since the system wants to recreate the thermal equilibrium immediately. This initial state might also happen spontaneously, but the probability of such an event is again exceedingly low.

This discussion suggests that an optimal norm should take into account the probability of the different initial states. Houtekamer (1995) and J. Barkmeijer (1996, personal communication) involved information from the analysis error covariance matrix to achieve this goal in practice. R. Pasmanter (1996, personal communication) suggested a norm based on considerations from statistical physics. These norms produce more plausible singular vectors than the energy norm, but an optimal norm still does not seem to exist. On the other hand, the adjoint integration has the formal advantage that the result always exists and is unique, while the inverse calculation cannot be performed if any of the singular values is zero or very small in a finite precision computational environment. This may happen when there exist initial perturbation patterns that disappear from the model by the end of the integration or, in other words,

in the presence of strong dissipation. The solution that we have adopted in this paper is to use an inverse linear model that handles the dissipation and the diffusion in a same way as an adjoint model does. This approach gives realistic results if the role of dissipation and diffusion is restricted to the control of numerical and gravity wave noise. In this case there are no extremely small singular values and when the adjoint scheme is iterated, it should eventually converge (like the quasi-inverse method) to the exact solution of (12) and, if the linearity assumption is valid, to the solution of the nonlinear error equation (4).

One may argue that the atmosphere is a forced/dissipative system rather than Hamiltonian, and therefore these arguments are not relevant to a GCM. This is true in a global but not in a local sense, since the Reynolds number in the free atmosphere is somewhere between the simple and double precision roundoff error and smaller than the neglected terms of the Coriolis force.

REFERENCES

- Andersson, E., P. Courtier, C. Gaffard, J. Haseler, F. Rabier, P. Uden, and D. Vasiljevic, 1996: 3D-Var—The new operational analyses scheme. *ECMWF Newsllett.*, **71**, 2–5.
- Arnold, V. I., 1989: *Mathematical Methods of Classical Mechanics*. 2d ed. Springer, 508 pp.
- Buizza, R., 1994: Sensitivity of optimal unstable structures. *Quart. J. Roy. Meteor. Soc.*, **120**, 429–451.
- , J. Tribbia, F. Molteni, and T. Palmer, 1993: Computation of optimal unstable structure for a numerical weather prediction model. *Tellus*, **45A**, 388–407.
- Cohn, S. E., N. S. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867.
- Courtier, P., J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4-D VAR using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.
- Derber, J. C., 1987: Variational four-dimensional analysis using quasi-geostrophic constraints. *Mon. Wea. Rev.*, **115**, 998–1008.
- , 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2446.
- , D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Wea. Forecasting*, **6**, 538–547.
- Development Division, NMC/NWS/NOAA, 1988: Documentation of the NMC global model. 244 pp. [Available from Environmental Modeling Center, NCEP/NWS/NOAA, 5200 Auth Road, Washington, DC 20233-9910.]
- Errico, R. M., and T. Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU–NCAR mesoscale model. *Mon. Wea. Rev.*, **120**, 1644–1660.
- , —, and K. Raeder, 1993: Examination of the accuracy of a tangent linear model. *Tellus*, **45A**, 462–477.
- Gill, P. E., W. Murray, and M. H. Wright, 1981: *Practical Optimization*. Academic Press, 401 pp.
- Hong, S.-Y., and H.-L. Pan, 1996: Nonlocal boundary layer diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339.
- Houtekamer, P. L., 1995: The construction of optimal perturbations. *Mon. Wea. Rev.*, **123**, 2888–2898.
- Huang, X.-Y., N. Gustafsson, and E. Källén, 1996: A poorman's 4DVAR: The use of the adjoint model in an intermittent data assimilation system. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 132–134.
- Kalnay, E., and Z. Toth, 1994: Removing growing errors in the analysis. Preprints, *10th Conf. on Numerical Weather Prediction*, Potland, OR, Amer. Meteor. Soc., 212–215.
- , and —, 1996: The breeding method. *Proc. of the ECMWF Seminar on Predictability*, Reading, United Kingdom, ECMWF, 69–82. [Available from ECMWF, Reading RG2-9AX, United Kingdom.]
- , and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kanamitsu, M., and Coauthors, 1991: Recent changes implemented in the global forecast system at NMC. *Wea. Forecasting*, **6**, 425–435.
- Lacarra, J. F., and O. Talagrand, 1988: Short range evolution of small perturbation in a barotropic model. *Tellus*, **40A**, 81–95.
- LeDimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 91–110.
- Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269–298.
- , R. Buizza, T. Palmer, and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- 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.
- Olver, P. J., 1993: *Applications of Lie Groups to Differential Equations*. 2d ed. New York, 513 pp.
- Pan, H.-L., and W.-S. Wu, 1995: Implementing a mass flux convection parameterization package for the NMC medium-range forecast model. NCEP Office Note, 409, 42 pp. [Available from NOAA/NWS/NCEP, Environmental Modeling Center, WWB, Room 207, Washington, DC 20233.]
- Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Pu, Z. X., 1996: Application of adjoint and quasi-inverse linear model of the NCEP operational global spectral model to sensitivity analysis and 4-D VAR data assimilation. Ph.D. dissertation, Lanzhou University, China, 130 pp. [Available from EMC/NCEP, Washington, DC 20233.]
- , E. Kalnay, J. C. Derber, and J. Sela, 1996: Using past forecast error to improve the future forecast skill. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 142–143.
- , —, —, and —, 1997: Using forecast sensitivity patterns to improve future forecast skill. *Quart. J. Roy. Meteor. Soc.*, **123**, 1035–1054.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121–150.
- Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281–1305.
- Sela, J. G., 1980: Spectral modeling at the National Meteorological Center. *Mon. Wea. Rev.*, **108**, 1279–1292.
- , and Coauthors, 1988: Documentation of the research version of the NMC medium range forecasting model. 504 pp. [Available from NOAA/NWS/NCEP, Environmental Modeling Center, WWB, Room 207, Washington, DC 20233.]
- Shepherd, T. G., 1990: Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Advances in Geophysics*, Vol. 32, Academic Press, 287–338.
- , 1993: A unified theory of available potential energy. *Atmos.–Ocean*, **31**, 1–26.
- Simmons, A. J., 1995: High-performance computing requirements for medium-range weather forecasting. *ECMWF Newsllett.*, **69**, 8–13.

- , R. Mureau, and T. Petroligis, 1995: Error growth and estimates of predictability from the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771.
- Szunyogh, I., E. Kalnay, and Z. Toth, 1997: A comparison of Lyapunov vectors and optimal vectors in a low resolution GCM. *Tellus*, **49A**, 200–227.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1996: Ensemble forecasting at NCEP. *Proc. of the ECMWF Seminar on Predictability*, Reading, United Kingdom, ECMWF, 39–60. [Available from ECMWF, Reading RG2-9AX, United Kingdom.]
- , I. Szunyogh, and E. Kalnay, 1996: Singular, Lyapunov and bred vectors in ensemble forecasting. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 53–55.
- Vukicevic, T., 1991: Nonlinear and linear evolution of initial forecast error. *Mon. Wea. Rev.*, **119**, 1602–1611.
- Zou, X., I. M. Navon, and F. X. LeDimet, 1992: An optimal nudging data assimilation scheme using parameter estimation. *Quart. J. Roy. Meteor. Soc.*, **118**, 1163–1186.
- Zupanski, D., and F. Mesinger, 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.
- , 1995: An iterative approximation to the sensitivity in calculus of variations. *Mon. Wea. Rev.*, **123**, 3590–3604.
- , and D. Zupanski, 1996: A quasi-operational application of a regional four-dimensional variational data assimilation. Preprints, *11th Conf. on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 94–95.