

A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems

DUSANKA ZUPANSKI

NCEP/UCAR Visiting Research Program, Washington, D.C.

(Manuscript received 23 July 1996, in final form 31 December 1996)

ABSTRACT

A technique to apply the forecast model as a general weak constraint in a complex variational algorithm, such as NCEP's regional 4DVAR data assimilation system, is presented. The proposed definition of the model error has a flexible time resolution for the random error term. It has a potential for operational application, because the coarse time resolution of the random error term and a diagonal in time random error covariance matrix, as used in this study, require less computational space.

The results presented in this study strongly indicate the need for a weak constraint (as opposed to a strong constraint formulation) in order to get the full benefit of a 4DVAR method. The inclusion of the model error term, even only the systematic error part, gives a main contribution to the capability of the 4DVAR method to outperform the optimal interpolation method.

1. Introduction

It is well known that forecast models are not perfect. The model error should be taken into account if a sophisticated data assimilation method is to be developed. In Kalman filter theory the solution to the problem of including the model error is well established and forms an integral part of the Kalman filter itself. However, due to limited computer power and insufficient observations, only crude approximations to this complex problem have been considered so far. For example, there are studies dealing with one-dimensional models (Ghil et al. 1981; Cohn et al. 1981; Dee et al. 1985) or two-dimensional shallow-water models (Cohn and Parrish 1991; Boutier 1993; Todling and Ghil 1994; Cohn et al. 1994). There are also some other simplified versions of the Kalman filter, applicable to more complex models, such as the extended Kalman filter (e.g., Ghil 1989; Gautier et al. 1993) or simplified dynamics filter (Dee 1991, 1995). The possibility of employing an ensemble-prediction approach in the Kalman filter technique in order to make it more suitable for realistic (nonlinear) forecast models has been examined recently, but using an idealized quasigeostrophic model (e.g., Evensen 1994; Evensen and van Leeuwen 1996). Even though the aforementioned papers provide a very useful theoretical background for a sophisticated data assimilation system, the common problem that has not been ad-

ressed yet is how to make this theory applicable to the state-of-the-art forecast models.

An approach to take the model deficiency into account when designing a sophisticated data assimilation system was considered recently in the framework of ensemble prediction by Houtekamer et al. (1996). In this "Monte Carlo" study, the ensemble members are obtained by perturbing both the observations and the forecast model in order to generate appropriate forecast error statistics. This is a valid approach that has the advantage of being applicable to nonlinear forecast models. The computational cost, however, being determined by the number of the necessary ensemble members, might be a limiting factor.

Another promising method, the four-dimensional variational (4DVAR) data assimilation method, while becoming increasingly applicable to realistic primitive equation models (Navon et al. 1992; Thepaut et al. 1993) and the most complex diabatic and mesoscale forecast models (e.g., Zou et al. 1993; Zou et al. 1995; D. Zupanski 1993; Tsuyuki 1996), has not paid enough attention so far to the treatment of model error. In order to account for the model error in 4DVAR methods, one should apply the forecast model as a weak constraint, rather than as a strong constraint, as was pointed out long ago by Sasaki (1970). This approach is, however, difficult to apply in realistic 4DVAR data assimilation systems because of the requirements for tremendous computer resources. Thus, a straightforward application of the theory would require storage of the model error term every time step of the integration period, which, for a complex model, would require an error vector of size $N = 10^8$ – 10^9 and a model error covariance matrix

Corresponding author address: Dusanka Zupanski, NCEP/WWB2, Room 207, 5200 Auth Rd., Camp Springs, MD 20746.
E-mail: dzupanski@sg179.wwb.noaa.gov

of the size $N \times N$. Obviously, these tremendous requirements are far from being fulfilled in realistic data assimilation experiments and remain a big obstacle for both the Kalman filter and 4DVAR method. However, as was argued by Dee (1995), even in the case of having available much more powerful computers, the amount of information that is available (observations, prior knowledge about the model error) is clearly insufficient to describe the complex behavior of the model error. Consequently, only crude approximations have to be considered. It is our hope, however, that a better knowledge about the model error can be obtained in the future and the situation may improve.

The model deficiency was first accounted for in a 4DVAR method through the definition of a systematic error term by Derber (1989). Later, this idea was further examined using a shallow-water model (Wergen 1992), in a barotropic hurricane track model (DeMaria and Jones 1993), and in a complex 4DVAR data assimilation system (M. Zupanski 1993a). The results of these studies show a considerable positive effect on the forecast, even with this approximate definition of the model error. More recently, a weak constraint was defined through a more general error term, including both the systematic and the random error parts, and applied to data assimilation for a barotropic (Bennett et al. 1993) and a complex primitive equation model (Bennett et al. 1996). These results, even with simple error covariance models, and without physical parameterizations included in the adjoint model, show an encouraging benefit of applying the forecast model as a weak constraint in a 4DVAR data assimilation method. A study by Menard and Daley (1996), where the equivalence between the fixed-interval Kalman smoother and the weak constraint 4DVAR optimization method (in their formulation Pontryagin optimization) was employed, also indicated the importance of the weak constraint assumption in the 4DVAR method.

In this paper we propose a general weak constraint (including both systematic and random error parts) applicable to the most complex 4DVAR data assimilation systems and in operational practice. The plan of the paper is as follows. In section 2 we present the theoretical background. In section 3 some problems that need special attention, such as minimization and preconditioning, are addressed. The model and data used in this study are described in section 4. The experimental design is given in section 5, the results in section 6, and finally, in section 7, the summary and conclusions are presented.

2. Theoretical background

a. The functional

The 4DVAR method can be defined as a process of minimizing the following functional (e.g., Jazwinski 1970; Lorenc 1986):

$$J = \frac{1}{2} \sum_{n=1}^N (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}})^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \frac{1}{2} \sum_{k=1}^M \sum_{m=1}^M \Phi_m^T \mathbf{Q}_{m,k}^{-1} \Phi_k + \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \text{gravity wave penalty}, \quad (2.1)$$

with the forecast model G and a postprocessing operator, H , imposed as weak constraints, defined by

$$\mathbf{x}_m = G(\mathbf{x}_{m-1}) + \Phi_m \quad (2.2)$$

$$\mathbf{y}_n = H_n(\mathbf{x}_m) + \epsilon_n. \quad (2.3)$$

In (2.1)–(2.3) the superscript T stands for a transpose, obs denotes observations, b is a background value, the index n defines observational times, and indexes m and k are model time steps. Matrices \mathbf{R} , \mathbf{B} , and \mathbf{Q} are observational, background, and model error covariances, respectively. The model error correction term Φ_m accounts for the model error growth from time t_{m-1} to time t_m . At this point we do not make any assumption about the model error. For simplicity, we assume that the observational error ϵ_n is stationary and white in space and time, with the mathematical expectation equal to $\langle \epsilon_n \epsilon_k^T \rangle = \mathbf{R} \delta_{n,k}$ (more about justification for this assumption is given in section 5). In the most general case, the observational noise ϵ_n can be considered as a control variable of the minimization problem (2.1)–(2.3). In our experiments we neglect ϵ_n in (2.3), that is, apply H as a strong constraint. The control variable (\mathbf{z}) includes the initial conditions \mathbf{x}_0 and model error Φ_m ; that is,

$$\mathbf{z} = (\mathbf{x}_0, \Phi_m). \quad (2.4)$$

For limited-area models we can also consider the model error term as a two-component vector: one is the model error inside the integration domain, and the other is the boundary condition's error. Under these conditions, the definition of the 4DVAR problem given by (2.1)–(2.4) becomes identical to the general 4DVAR problem defined in Bennett et al. (1993) and Bennett et al. (1996).

The sufficient conditions for the existence of a local minimum of (2.1) are 1) the total variation of J is $\delta J = 0$, for the arbitrary $\delta \mathbf{x}_0$ and $\delta \Phi_m$, and 2) the Hessian of J is positive definite. As an approximation, we use only a first variation of J , which can be expressed using gradients of J with respect to \mathbf{x}_0 , denoted $\nabla_{\mathbf{x}_0} J$, and Φ_m , denoted $\nabla_{\Phi_m} J$, as

$$\delta J = \delta \mathbf{x}_0^T \nabla_{\mathbf{x}_0} J + \delta \Phi_m^T \nabla_{\Phi_m} J. \quad (2.5)$$

After taking a first variation of (2.1)–(2.4) and an adjoint of a linear operator in a Hilbert space (detailed steps given in appendix A), we have

$$\begin{aligned} \delta J = & \delta \mathbf{x}_0^T \sum_{n=1}^N \mathbf{L}_{1,n} \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta \mathbf{x}_0^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \delta \Phi_M^T \sum_{n=1}^N \mathbf{H}_n^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta \Phi_M^T \sum_{j=1}^M \mathbf{Q}_{M,j}^{-1} \Phi_j + \dots \\ & + \delta \Phi_{k-1}^T \sum_{n=1}^N \mathbf{L}_{k,n} \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta \Phi_{k-1}^T \sum_{j=1}^M \mathbf{Q}_{k-1,j}^{-1} \Phi_j + \dots + \delta \Phi_1^T \sum_{n=1}^N \mathbf{L}_{2,n} \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta \Phi_1^T \sum_{j=1}^M \mathbf{Q}_{1,j}^{-1} \Phi_j, \end{aligned} \quad (2.6)$$

where the linear operator $\mathbf{L}_{k,n}$ is defined as

$$\mathbf{L}_{k,n} = \mathbf{G}_k^T \dots \mathbf{G}_{M-1}^T \mathbf{G}_M^T \mathbf{H}_n^T. \quad (2.7)$$

In the above equations \mathbf{G} and \mathbf{H} are linearizations of G and H , and \mathbf{G}^T and \mathbf{H}^T are the corresponding adjoint (transpose) operators. We assume that Φ_0 and $\delta \Phi_0$ are equal to zero at the initial time t_0 (there is no model error at the initial time).

The first two terms in (2.6) represent a variation of J with respect to the initial conditions $\delta \mathbf{x}_0$, where the gradient $\nabla_{\mathbf{x}_0} J$ is calculated by applying the adjoint model in the way identical to the case of a strong constraint. The remaining terms define a variation of J with respect to the model error $\delta \Phi$, obtained also by integrating the adjoint model. Note that for each model time step m the adjoint model produces a new value of the gradient $\nabla_{\Phi_m} J$.

The gradient with respect to the model error includes also the model error covariance terms (the “ \mathbf{Q} terms”). Because of the large size of the matrix \mathbf{Q} , the weak constraint approach becomes much more demanding concerning the computational space, as compared to the strong constraint approach. One possible way to overcome this obstacle is to define the model bias as a control variable of the 4DVAR problem (Derber 1989) while keeping the initial conditions constant. In that case the size of the problem becomes greatly reduced: the control variable is of the same size as in the case of applying a strong constraint and adjusting the initial conditions only. If the number of degrees of freedom in the data is large enough [e.g., if gridded data is assimilated, as in the paper by Derber (1989)], then the model error covariance term may not be necessary, which reduces the size of the problem even more. Another approach, used in M. Zupanski (1993a), is to adjust both the initial conditions and the Derber’s model bias term but exclude the model error covariance term from the definition of the functional. The problem formulated in this way is computationally feasible on the presently available computers. Practical applications in a complex, state-of-the-art 4DVAR data assimilation system (M. Zupanski 1993a; Zupanski and Mesinger 1995; Zupanski and Zupanski 1995) indicate that the approach has the capability of outperforming the optimal interpolation (OI) method. The problem with this approach is that it is not general enough (does not include the random part of the model error), and, on the other hand, the 4DVAR prob-

lem is not posed in a mathematically correct way, being greatly underdetermined without the constraint of the model error covariance term. The third approach, used in Bennett et al. (1993) and Bennett et al. (1996), is more general (includes both systematic and random error parts) and is mathematically well posed. One of the important features of Bennett’s approach is the definition of the minimization in the observational space, rather than in the model space, which makes the 4DVAR algorithm more efficient concerning the computational space requirements, when the size of the observational vector is small as compared to the size of the control variable in the model space. However, in order to apply the approach in a state-of-the-art 4DVAR data assimilation system, using all available observations, some further simplifications might be needed. In seeking a mathematically correct general weak constraint solution, applicable in the most complex 4DVAR data assimilation systems, we proceed with the following derivation.

From (2.6) and (2.7) the relation between the model error gradients at two different times, say $m-1$ and $k-1$, with $m-1 > k-1$, is

$$\begin{aligned} \nabla_{\Phi_{k-1}} J = & \mathbf{G}_k^T \dots \mathbf{G}_{m-1}^T \left(\nabla_{\Phi_{m-1}} J - \sum_{j=1}^M \mathbf{Q}_{m-1,j}^{-1} \Phi_j \right) \\ & + \sum_{n=1}^N \mathbf{G}_k^T \dots \mathbf{G}_{m-1}^T \mathbf{H}_n^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) \\ & + \sum_{j=1}^M \mathbf{Q}_{k-1,j}^{-1} \Phi_j. \end{aligned} \quad (2.8)$$

By examining (2.8), we immediately conclude that, for the case of neglecting the effect of the \mathbf{Q} term ($\mathbf{Q}^{-1} = 0$) and in the absence of additional observations between the times $k-1$ and $m-1$, the two gradients are linearly dependent, which means that there would be no benefit in using fine time resolution (e.g., each model time step) gradient information during the time interval $(k-1, m-1)$. Similar consideration for the space dimension would lead to the same conclusion. In a more realistic case we have $\mathbf{Q}^{-1} > 0$ but still big data gaps (e.g., data are collected every 3 h, the observations over the oceans are sparse). Then the relation between the two gradients would depend exclusively on our knowledge of \mathbf{Q} in the time intervals or areas with data

gaps. Since our knowledge of the model error covariance is very limited, it seems reasonable, as a first step, to define a minimization problem less dependent on model error covariance matrix. Based on this, we define an approximation for the model error described in the following section.

b. The model error

We define the model error Φ_m as a first-order Markov variable following Daley (1992), but we assume that the random part of the error (\mathbf{r}) is defined on a coarser timescale; that is,

$$\Phi_m = \frac{\mu \Phi_{m-1}}{\mu + (1 - \mu^2)^{1/2}} + \frac{(1 - \mu^2)^{1/2} \mathbf{r}_N}{\mu + (1 - \mu^2)^{1/2}}, \quad (2.9)$$

where index m denotes, as before, the model time steps; the index N defines some coarser time steps (of the order of several hours); and μ is a constant having a value between 0 and 1. As in the original paper (Daley 1992), the parameter μ controls the relative weights given to the systematic, time correlated [first term in (2.9)] and the random [second term in (2.9)] error parts. Unlike in the original paper, we normalize the coefficients by the constant $\mu + (1 - \mu^2)^{1/2}$ to make the comparison between different experiments easier. The model error given by (2.9) has these important features:

- 1) it is a serially (and spatially) correlated error, where the value at time m depends only on the value from a previous time step, $m - 1$;
- 2) the model error defined by (2.9) automatically includes error at the boundaries of the model integration domain in the case of a limited area model;
- 3) given the initial value of the model error (Φ_0), the control variable of the minimization includes only a random error part (\mathbf{r}_N), which considerably reduces the size of the problem;
- 4) the model bias at any time is a function of the initial model bias, defined as

$$\langle \Phi_m \rangle = \alpha \langle \Phi_{m-1} \rangle = \dots = \alpha^m \langle \Phi_0 \rangle;$$

$$\alpha = \frac{\mu}{\mu + (1 - \mu^2)^{1/2}}$$

(in our experiments we assume $\Phi_0 = \mathbf{0}$, as a reasonable choice, since its effect diminishes exponentially with time, but it is also possible to optimize it along with \mathbf{r}_N); and

- 5) the model error covariance matrix is fully nondiagonal in space and time but dependent only on the diagonal in time random error covariance matrix (the derivation given is in the appendix B).

Given the knowledge of the diagonal in time random error covariance matrix \mathbf{W}_N , one can conveniently find the minimum of

$$J = \frac{1}{2} \sum_{n=1}^N (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}})^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \frac{1}{2} \sum_{N=1}^{N_{\text{max}}} \mathbf{r}_N^T \mathbf{W}_N \mathbf{r}_N + \frac{1}{2} (\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \text{gravity wave penalty} = \text{MIN}, \quad (2.10)$$

where N_{max} is the total number of random error terms. The problem defined by (2.10) with the constraints (2.2), (2.3), and (2.9) is equivalent to the problem defined by (2.1) with the same constraints. The model error dependence only on the random error terms \mathbf{r}_N and the initial model error Φ_0 (zero in our experiments) allows us to replace the large size problem (2.1) by a much smaller size problem (2.10) (proof can be found in Jazwinski 1970, chapter 3.9). Regarding this simplification, there is an important difference between the 4DVAR and Kalman filtering methods. In 4DVAR method one can easily exploit the advantage of matrix \mathbf{W} being diagonal in time, as explained above. In Kalman filtering this cannot be done in a straightforward way due to the need for explicit treatment of the model error covariance. Thus, unless a special effort is taken to eliminate the systematic model error, all the terms in appendix B need to be carried out through the time. In return, the Kalman filter offers the forecast error covariance as an end product. In the 4DVAR method, however, the forecast error covariance matrix is only implicitly used and, therefore, is not available.

To further simplify the problem, we assume that the random error covariance matrix is stationary and denote it by \mathbf{W} . However, it is still a nontrivial problem to define it, given very little knowledge about the true matrix. One of the simplest ways to solve this problem is to use the background error covariance matrix \mathbf{B} , multiplied by a constant, as a first step. Then, as a next step, the random error covariance matrix can be defined using the optimal values of \mathbf{r}_N obtained from actual data assimilation experiments. One possible approach is to calculate \mathbf{W}^{-1} by applying an iterative Sherman–Morrison formula using perturbations (or in our case \mathbf{r}_N vectors) from a number of synoptic cases, as was done when calculating \mathbf{B}^{-1} in Zupanski and Zupanski (1995). Another possibility is to define appropriate correlation functions and their parameters (e.g., Gaspari and Cohn 1996) given the same data (\mathbf{r}_N vectors) from actual data assimilation experiments.

3. Minimization and preconditioning

A serious practical difficulty associated with a weak constraint is that the minimization problem can become increasingly ill conditioned when the model error terms are considered as control variables. One possible solution is to define the minimization part of the data assimilation problem in the observational space (e.g., Bennett et al. 1993; Bennett et al. 1996). This approach can simplify the preconditioning problem in the case

where the size of the observational vector is considerably smaller than the size of the model error vector. Our experience, however, is that the preconditioning can be handled well and the minimization can converge very fast, even with the control variables kept in the model space. One reason might be that we have greatly reduced the number of degrees of freedom, owing to the coarse time resolution of the random error term (\mathbf{r}_N). The size of the control variable in our experiments is 0.6×10^6 (initial conditions) + $N_{\max} \times 0.6 \times 10^6$ (model error and boundary conditions) $\leq 3 \times 10^6$, N_{\max} being the maximum number of the random error terms ($N_{\max} \leq 4$, for a 12-h period). This is a substantial reduction compared to a general case, where the size would be equal to 0.6×10^6 (initial conditions) + $M \times 0.6 \times 10^6$ (model error and boundary conditions) = 1.3×10^8 , M being the maximum number of model time steps ($M = 217$, for a 12-h period), assuming the same resolution model (80 km/17 layers) is used in both cases. The difference in the number of degrees of freedom between the most general and our approach is even more dramatic when measured by the dimensionality of the error covariance space. It is $\dim(\mathbf{B}) + M \times M \times \dim(\mathbf{W})$, for the most general model error, and $\dim(\mathbf{B}) + \dim(\mathbf{W})$, for the error defined by (2.9), with the assumption of stationarity in model error covariance matrix, in both cases.

The minimization method used in this study is the quasi-Newton memoryless algorithm (Shanno 1978, 1985) implemented as in M. Zupanski (1993b), Zupanski and Mesinger (1995), and M. Zupanski (1996). To define a preconditioning matrix, we follow the simple principal, proposed in M. Zupanski (1993b, 1996): calculate the ratio of the expected decrease of the functional and the corresponding gradient norm to obtain the elements of the diagonal preconditioning matrix. We start from the following general definition, given in M. Zupanski [(1996), Eqs. (3.4), (4.8), (4.9), and (4.10)]:

$$\mathbf{P}_c = \left[\text{const} \frac{[(F_{\text{ana}}^{\text{total}})_c]^{1/2}}{[\mathbf{g}^T \mathbf{R}^{-1} \mathbf{g}]^{1/2}} \right]^{-1} \mathbf{R}_c + \mathbf{A}_c^{-1}, \quad (3.1)$$

where \mathbf{P} is the inverse of the preconditioning (diagonal) matrix; \mathbf{R} is an estimate of the analysis error variance multiplied by a constant; $F_{\text{ana}}^{\text{total}}$ is the total functional defined over the model (analysis) space, without the terms included in calculation of \mathbf{A}^{-1} ; and \mathbf{g} is the gradient of the functional. The matrix \mathbf{A}^{-1} makes contributions to the preconditioning matrix from the terms of the functional calculated exactly (e.g., it includes \mathbf{B}^{-1} for the background term of the functional, or \mathbf{W}^{-1} for the model error penalty term). The index c defines a subspace that, in this case, corresponds to a model variable (u, v, T, p_s, q) at a specific model level (l). Thus, for a specific model variable, say $u(l)$, the elements of the preconditioning matrix are calculated by dividing the functional value $(F_{\text{ana}}^{\text{total}})_{u(l)}$, obtained using only u -wind observations at level l , by the corresponding gra-

dient norm $(\mathbf{g}^T \mathbf{R}^{-1} \mathbf{g})_{u(l)}$. As shown in the original paper, the preconditioning defined by (3.1) is applicable to both the model bias term, as defined by Derber (1989), and to the initial conditions.

In the case of a general model error term, as defined in this study, a straightforward substitution in (3.1) of the model error gradient norm did not work well. One obvious problem was the big disparity in magnitudes of the two gradients, when measured using the same definition of the norm. Thus, the norm of the model error gradient was typically two orders of magnitude greater than the initial conditions gradient norm, which after applying (3.1) would produce a descent direction (calculated in the first iteration of the minimization as $\mathbf{d} = \mathbf{P}^{-1} \mathbf{g}$) almost parallel to the model error gradient. To eliminate this problem, we employ a more general definition of the preconditioning matrix, applicable to a broader range of magnitudes of the specific components of the gradients. We define different norms for the initial condition and the model error gradients, using different weighting matrices \mathbf{R}_{x_0} and \mathbf{R}_Φ , and assume the following relation between the two weights:

$$\mathbf{R}_\Phi^{-1} = \alpha^2 \mathbf{R}_{x_0}^{-1}, \quad (3.2)$$

where α is an empirical constant to be determined. We then require that the two gradients are of the same magnitude; that is,

$$(\mathbf{g}_{x_0}^T \mathbf{R}_{x_0}^{-1} \mathbf{g}_{x_0})^{1/2} = (\mathbf{g}_\Phi^T \mathbf{R}_\Phi^{-1} \mathbf{g}_\Phi)^{1/2}, \quad (3.3)$$

where \mathbf{g}_{x_0} is the *total* gradient with respect to the initial conditions and \mathbf{g}_Φ is the *total* gradient with respect to the model error. From (3.2) and (3.3) we calculate α as

$$\alpha = \frac{(\mathbf{g}_{x_0}^T \mathbf{R}_{x_0}^{-1} \mathbf{g}_{x_0})^{1/2}}{(\mathbf{g}_\Phi^T \mathbf{R}_{x_0}^{-1} \mathbf{g}_\Phi)^{1/2}}. \quad (3.4)$$

And finally, from (3.1), (3.2), and (3.4) we obtain the following new (inverse) preconditioning for the gradient with respect to the model error \mathbf{g}_{Φ_m} :

$$\mathbf{P}_{\Phi_m, c} = \left[\alpha \text{const} \frac{[(F_{\text{ana}}^{\text{total}})_c]^{1/2}}{[\mathbf{g}_{\Phi_m}^T \mathbf{R}_{x_0}^{-1} \mathbf{g}_{\Phi_m}]^{1/2}} \right]^{-1} \mathbf{R}_{x_0}^c + \mathbf{W}_c^{-1}, \quad (3.5)$$

while the inverse preconditioning for the gradient with respect to the initial conditions \mathbf{g}_{x_0} remains as in the original paper; that is,

$$\mathbf{P}_{x_0, c} = \left[\text{const} \frac{[(F_{\text{ana}}^{\text{total}})_c]^{1/2}}{[\mathbf{g}_{x_0}^T \mathbf{R}_{x_0}^{-1} \mathbf{g}_{x_0}]^{1/2}} \right]^{-1} \mathbf{R}_{x_0}^c + \mathbf{B}_c^{-1}. \quad (3.6)$$

In (3.5) and (3.6) we included the random error covariance matrix \mathbf{W} in the preconditioning for the model error and the background error covariance \mathbf{B} in the expression for the initial conditions. Even though the two error covariance matrices are nondiagonal in space, in our application of the preconditioning we keep only diagonal components, for simplicity. This approximation, however, did not have a critical effect on the efficacy of the preconditioning. As before, the precon-

ditioning matrix has components for each model variable and at each model level (the subspace denoted by c). In addition, there is a time component (subspace denoted by m) for the model error preconditioning matrix.

4. Model and data

The regional National Centers for Environmental Prediction (NCEP) Eta Forecast Model (Mesinger et al. 1988; Janjic 1990, 1994; Janjic et al. 1995; Black 1994) with steplike mountains and the η vertical coordinate is used in this study. The model's free atmosphere turbulence is simulated by the Mellor–Yamada level 2.5 (Mellor and Yamada 1982) scheme, implemented as in Janjic (1990). The turbulent exchange between the model lowest layer and the earth's surface is parameterized using Mellor–Yamada level 2 scheme with the simplified second-order closure model defined in Loboocki (1993). A viscous sublayer is defined over water (Janjic 1994, and references therein). The moist processes are described by the smooth (Zupanski and Mesinger 1995) Betts–Miller–Janjic (Betts 1986; Betts and Miller 1986; Janjic 1994; Betts and Miller 1992) cumulus convection scheme and large-scale condensation and evaporation (Janjic 1990). In the smooth cumulus convection scheme we eliminated only those on–off switches with the most serious negative effect on the minimization, as explained in D. Zupanski (1993) and Zupanski and Mesinger (1995). Many other discontinuous switches are kept as in the original model. The effect of these discontinuities was negligible in our experiments. The remaining physical processes include radiation (Lacis and Hansen 1974; Schwarzkopf and Fels 1991), surface processes, and second-order horizontal diffusion (Janjic 1990, and references therein). The integration domain of the model covers the North American continent and a part of each of the adjacent oceans. The horizontal resolution is approximately 80 km, and there are 17 layers in the vertical with the top at 50 hPa. The time step is 200 s for the fastest processes, 400 s for the advection and large-scale precipitation, and 800 s for the remaining physical processes other than radiation (convection, turbulence, and surface processes). The radiation is calculated at larger time intervals, but its contributions are added at every time step.

The adjoint model includes all the processes of the forecast model, except radiation. The integration domain, space resolution, and time resolution, are the same as in the forecast model. The basic state used to define adjoint operators is updated every fourth time step.

In this study we use all the observations (radiosonde, surface, aircraft, satellite profiles, etc.) as used in the operational Eta Data Assimilation System (EDAS, Rogers et al. 1995). The observations of T , u , v , h_s (surface height derived from the surface pressure observations) and q are used in the experiments. The synoptic situation chosen is the 8–10 May 1995 case with severe weather

developing over the central United States and moving to the east. Figure 1 shows the NCEP's subjectively analyzed sea level pressure map valid at 1200 UTC 10 May 1995.

5. Experimental design

a. Data assimilation experiments

1) BASIC DEFINITIONS

In the experiments presented we minimize the functional defined as in (2.10), using a digital filter (Lynch and Huang 1992) gravity wave penalty term as in Zupanski and Zupanski (1995), and with the weak constraint defined by (2.2), (2.3), and (2.9). The data assimilation period is 12 h, and the data insertion frequency is 3 h. With the specific model error definition, the control variable (\mathbf{z}) of the data assimilation problem includes the initial conditions and the coarse timescale random error term; that is,

$$\mathbf{z} = (\mathbf{x}_0, \mathbf{r}_N). \quad (5.1)$$

The control variable (5.1) includes the same components for both the initial conditions and the model error. They are temperature (T), wind (u , v), surface pressure (p_s), and specific humidity (q). The model error term is defined over the entire model integration domain, with the boundary conditions automatically included in the definition of the control variable (5.1). The time resolution of the boundary conditions is the same as the random error time resolutions. Since the new boundary conditions are inserted into the model every time step, they are obtained from the coarse timescale boundary conditions, by applying a linear interpolation.

2) ERROR COVARIANCES

The observation error (co)variance matrix \mathbf{R} is diagonal in space in time. Its diagonal elements (variances) depend on the observed variable, type of the observation, and the vertical level. In Table 1 we present the observation errors (defined as a square roots of the variances) for the variables used as observations in our experiments: temperature (T), wind (u , v), specific humidity (q), and surface height (h_s). The errors for temperature, wind, and surface height are derived from the values used in the EDAS analysis system. We altered the original EDAS variances to make a similar contribution of each observed field (T , u , v , h_s , q) to the total functional. To calculate the observational errors for specific humidity, we assume a 12% error in the relative humidity field and employ the temperature errors for a specific type of observations and a specific level. With this choice of the model error covariance matrix, the 4DVAR experiments presented later in the text are not consistently comparable with the OI experiments, since there is a difference in the definition of \mathbf{R} . This was necessary, however, since the 4DVAR method, as a

WEDNESDAY, MAY10, 1995

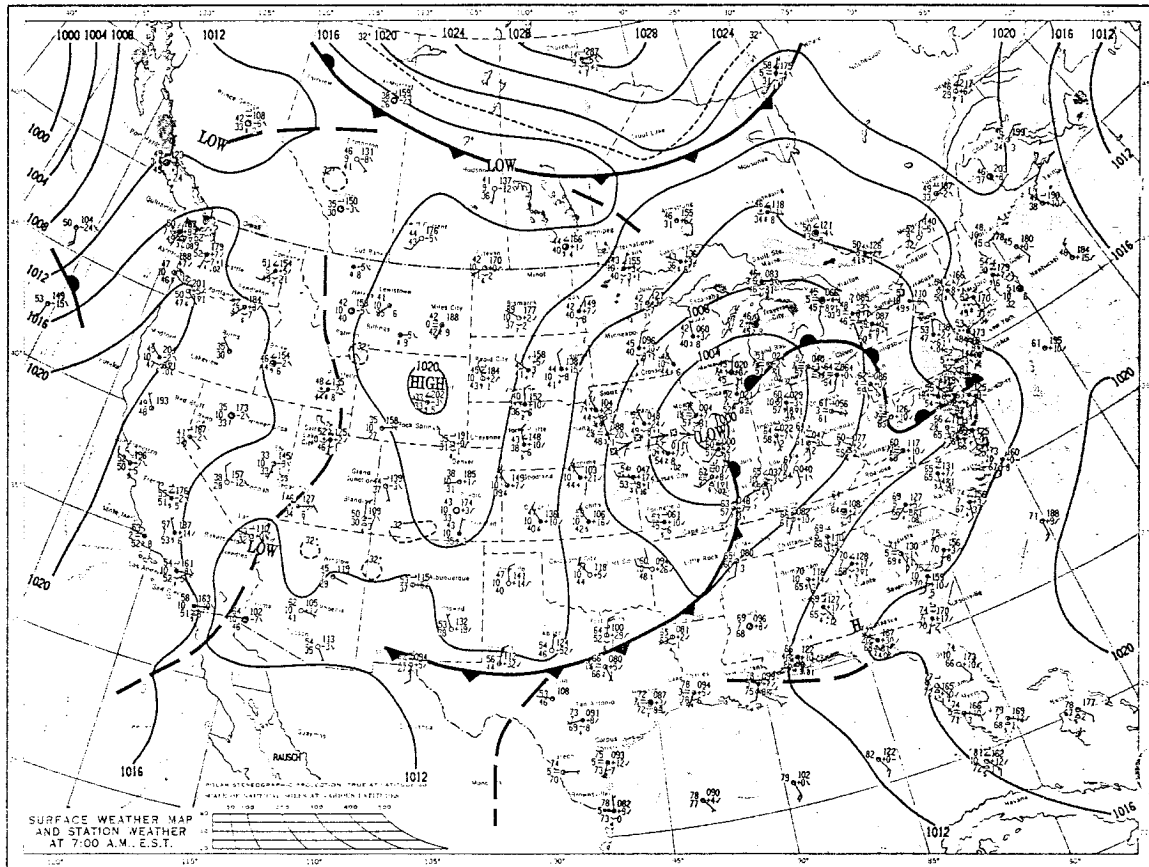


FIG. 1. NCEP's surface weather map valid at 1200 UTC 10 May 1995.

global solver, is more sensitive to the relative weights given to particular observed fields than the OI method. By experimenting with a range of values of \mathbf{R} , we found that the data assimilation results are not critically sensitive to particular choices of \mathbf{R} , once a reasonable range is chosen. Therefore we expect that the different values of \mathbf{R} used in 4DVAR and OI experiments should not have a critical impact on the experimental results.

It should be noted that the observation error covariance matrix used in our experiments is a very crude approximation to the real covariance matrix. First of all, the errors for the TIROS (Television Infrared Observation Satellite) Operational Vertical Sounder (TOVS) soundings seem to be too small. Also, some important correlations were neglected. Thus, the correlation in the vertical for the radiosonde observations, as well as horizontal and serial correlations for satellite observations, should be taken into account. As discussed in Daley (1992), there is a possibility that the error of some measurement instruments is correlated with the signal, which would make the observation error serially correlated. On the other hand, since the observational network is not fixed (in time), the assumption of stationarity

of \mathbf{R} might not be justified. These assumptions need to be reexamined, which is in our plans for future work.

The random error covariance matrix \mathbf{W} is stationary and diagonal in time. It is assumed to be equal to $\text{const} \times \mathbf{B}$. To calculate \mathbf{B}^{-1} , we apply the same procedure as in Zupanski and Zupanski (1995). The model error covariance matrix \mathbf{Q} is only implicitly defined. As explained in appendix B, it is fully nondiagonal in space and time.

3) THE EXPERIMENTS

Since the model errors are time correlated, they are more predictable than the pure white noise. Thus, the length of the correlation timescale, defined by the random error time interval in our experiments, would have an impact on the predictability of the model error. For longer correlation time periods, the effect of the systematic error becomes more dominant, which results in the more predictable model error, and vice versa. To examine the effects of different correlation timescales, as well as the effect of neglecting the model error, we define these data assimilation experiments:

TABLE 1. Observation errors.

Pressure level (hPa)	1000	700	500	300	100	50
Temperature (K)						
Radiosonde	1.20	1.40	1.47	1.72	2.53	2.53
Clear TOVS sounding		1.00	1.25	1.80	1.89	1.89
Cloudy TOVS sounding		1.00	1.25	1.80	1.89	1.89
Surf. marine	1.80					
Surf. land	1.80	1.80				
Specific humidity (g kg ⁻¹)						
Radiosonde	1.82	1.15	0.55	0.87		
Surf. marine	3.32					
Surf. land	3.06	3.30				
Wind (m s ⁻¹)						
Radiosonde	1.44	1.92	3.04	4.32	3.60	2.80
AIREP (flight level)	4.80	4.80	4.80	4.80	4.80	4.80
ACARS (flight level)	3.60	3.60	3.60	3.60	3.60	3.60
ASDAR (flight level)	3.92	3.92	3.92	3.92	3.92	3.92
Profiler winds	1.84	3.04	4.40	8.00	6.40	4.80
Pibal winds		3.04	4.40			
Low cloud drift (NESDIS)		3.36				
High cloud drift (NESDIS)		6.00	6.00	6.00	6.00	6.00
Low cloud drift (ESA)		5.76				
High cloud drift (ESA)		6.72	6.72	6.72	6.72	6.72
Surf. marine	2.80					
Surface height (m)						
Surf. marine	21.00					
Surf. land	16.20	24.45				

- 1) ERROR 12, with a 12-h correlation timescale (only one random error term is included);
- 2) ERROR 03, with a 3-h correlation timescale (four random error terms during a 12-h data assimilation period); and
- 3) NO ERROR, where no model error term is included (a strong constraint assumption).

In all the experiments we assimilate the same data over the same data assimilation period.

b. Forecast experiments

We perform the forecast experiments using the initial conditions defined at the end of 12-h data assimilation period. As in the data assimilation experiments, we also define two 4DVAR experiments with a weak constraint (ERROR 12 and ERROR 03) and one 4DVAR experiment with a strong constraint (NO ERROR). As a control experiment when examining the effects of the 4DVAR data assimilation method, we use the optimal interpolation experiment (OI EXP), where the optimal interpolation analyses are created as in operational EDAS system (except for different space resolution). In this experiment we define the same data assimilation period and assimilate the same data as in the 4DVAR experiments. The only difference between the 4DVAR and OI experiments is in the definition of **R**, as explained in section 5a.

The same model, without the model error term, is used in all the experiments. The verification of the fore-

casts is performed by calculating the rms differences between the forecast and all available observations every 12 h over a 48-h forecast period. The rms differences are calculated as mean values on 20 pressure levels starting from 1000 hPa with the pressure increments of -50 hPa.

6. Results

a. Data assimilation experiments

As was pointed out in section 3, one serious problem that needs special attention when dealing with a weak constraint in the 4DVAR framework is the preconditioning of the Hessian matrix. A common experience is that the minimization of the functional converges very slowly (the number of iterations measured in hundreds, converging not necessarily to the true minimum) due to ill conditioning of the 4DVAR problem, unless special care is taken to precondition the Hessian matrix. This problem becomes even worse in the case of a 4DVAR with a weak constraint. Therefore, it is important to pay attention to the convergence of the minimization in the data assimilation experiments with the model error. In Fig. 2 we present the value of the functional J (excluding the model error covariance “**W** term”) as a function of the iteration number obtained in the experiments NO ERROR, ERROR 12, and ERROR 03. The reason for excluding the **W** term is to make the comparison between the three experiments possible. In the case of a strong constraint (NO ERROR) this term is always equal

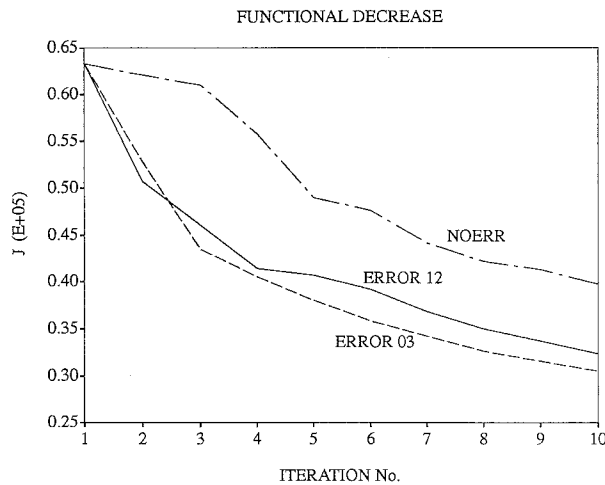


FIG. 2. The functional J (excluding the model error covariance term) plotted as a function of the iteration number for experiments NO ERROR (dashed-dotted line), ERROR 12 (solid line), and ERROR 03 (dashed line).

to zero and the total functional value cannot be compared with the other two experiments. As the figure indicates, starting from the same value, all functionals decrease relatively fast. The functional decreases most in the case of a general weak constraint (ERROR 03). The difference between experiments ERROR 12 and ERROR 03 is small, and both experiments show considerably larger reductions ($\approx 20\%$) of the functional as compared to the experiment NO ERROR. A considerable difference between all three experiments can be seen in particular parts of the functional. As an example, Fig. 3 shows the observational part [first term in (2.10)] of the functional calculated after 10 iterations of the minimization and plotted as instant functions of time (before the summation over time was taken). By examining the values of the functional, we immediately notice that the model error contributes more to the better fit to data at the times closer to the end than to the beginning of data assimilation period. It becomes even more pronounced with the finer time resolution of the model error, as can be seen by comparing the results of the experiments ERROR 12 and ERROR 03. This is in agreement with the Kalman filter theory, which says that the effect of the model error is in degrading the value of information in the distant past (e.g., Jazwinski 1970, chapter 8.10). On the other hand, overweighting the most recent data in the strong constraint formulation would have an effect similar to the model error in the weak constraint formulation (e.g., Courtier and Talagrand 1990). The better fit to the most recent data, by itself, however, is not sufficient to improve the analysis and forecast. For example, overweighting the most recent data in the nudging technique would result in a perfect fit to these data, but since the other constraints of the general minimization problem defined by (2.10) are not appropriately taken into account, the forecast

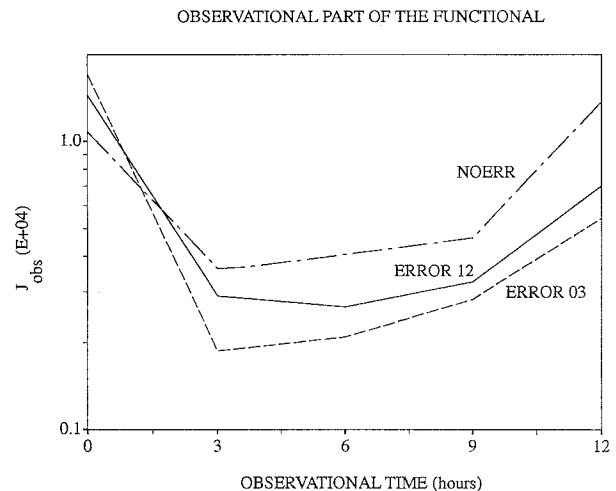


FIG. 3. The observational part of the functional calculated after 10 iterations of the minimization and plotted as an instant function of time. Note the better fit to the data at the end of the data assimilation period in the weak constraint experiments (ERROR 12 and ERROR 03) as compared to the strong constraint experiment (NO ERROR).

after such data assimilation would not probably be the optimal one. Only in the case when the improved fit to the most recent data is accompanied with a perfect balance between the observational, model error, background, and gravity wave penalty term, defined by (2.10), can one (theoretically) expect the solution of the problem to be close to optimal. (Practically, as in our experiments, due to poor knowledge about the error covariances, the solution of the problem is only sub-optimal.) It is also reasonable to expect that the better fit to the data at the end of the data assimilation period can improve the first-guess analysis for the next 4DVAR cycle and result in reducing the number of necessary iterations of the minimization. Results of cycled 4DVAR, obtained in Zupanski and Zupanski (1996), indicate that an improved first-guess analysis can contribute to considerably reducing the number of iterations. More applications related to using a 4DVAR method in a cycled mode can be found in Cohn et al. (1994).

We also present for experiment ERROR 03 the *total* functional decrease along with the background and \mathbf{W} term and gravity penalty term as functions of the iteration number in Fig. 4. As the figure shows, the convergence of the total functional is also relatively fast. The gravity penalty term decreases with the iteration number, indicating a good control of the gravity wave noise. The background and the \mathbf{W} term slowly increase with the iteration number. This is something one would expect, because both terms are initially equal to zero (the initial analysis is equal to the background field, and the model error is equal to zero). When the relative weights to the particular terms of the functional J are appropriately chosen, the increase of J_{back} and J_{w} terms should slow down, as the solution approaches the min-

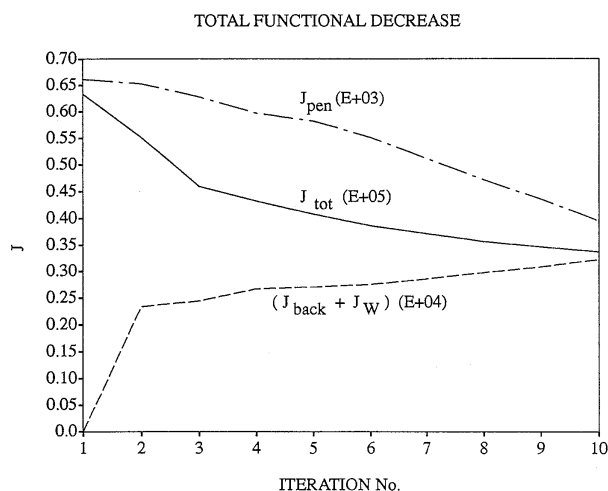


FIG. 4. The total functional decrease (solid line) along with the background and “ W term” (dashed line) and gravity penalty term (dashed-dotted line) plotted as functions of the iteration number for the experiment ERROR 03. The presented terms have different orders of magnitude, as indicated in parentheses.

imum. The results presented in Fig. 4, therefore, indicate that some more weight would need to be given to both background and model error covariance terms in order to achieve saturation. This however, would degrade the data assimilation results, probably due to a crude definition of the error covariances. It can be noted that the magnitudes of the particular terms of the functional are different (the gravity penalty term is approximately two orders of magnitude smaller, and the sum of the background and W term is an order of magnitude smaller than the total value of the functional). This is a consequence of the relative weighting given to the particular terms, and the values are obtained by performing the test experiments in several different synoptic situations.

It is important to note that in all the experiments presented only 10 iterations of minimization were performed. We did not run more iterations simply for reasons of economy. Our experience with continuing the minimization process (figure not included) is that the functional decreases further by several percent and probably 20–30 iterations would be necessary to reach the true minimum [by the criterion as in M. Zupanski (1996)]. The effect of further decreasing the functional is also reflected in a further (but small) improvement in the subsequent forecast.

One of the differences between the presented 4DVAR method and the Kalman filter method is an explicit calculation of the optimal model error term in the 4DVAR method. In the following figures we present the model error fields as yet another important component of the data assimilation experiments. In Figs. 5a–d the surface pressure component of the 3-h random error term, obtained after 10 iterations of the minimization process, is presented (experiment ERROR 03). We also present the 12-h random error term (experiment ERROR 12) in

Fig. 6. By comparing the 3- and 12-h random error terms, we can see that the 3-h random error allows for more variability in time and space. For example, the locations of the maximum values change from one time level to another in the 3-h random error case, while the 12-h random error case is dominated by the two maxima (off the Yucatan Peninsula, and in the central United States) over the whole data assimilation period.

Figures 7, 8, and 9 show the “optimal” perturbations to the surface pressure initial conditions obtained after 10 iterations of the minimization from experiments ERROR 03, ERROR 12, and NO ERROR, respectively. From the results presented in Figs. 7–9 we can learn more about another important aspect of the weak constraint that may have a relevance to the sensitivity studies. By examining the figures, we notice that the initial conditions perturbations are substantially different in the three experiments. The differences are even larger than it may seem at first sight, because in Fig. 9 (NO ERROR) we doubled the contouring interval. Obviously, in experiment NO ERROR, the calculated optimal perturbation to the initial conditions takes into account not only the initial conditions error, but the model error as well. The consequence is that, when attempting to find the origin of the forecast error, using the model as a strong constraint, as is done in some sensitivity studies, misleading conclusions may result.

By comparing the model error terms (Figs. 5 and 6) with the initial conditions perturbations (Figs. 7–9), we can immediately see that the magnitude of the model error term is rather small. It is between one and two orders of magnitude smaller than the average initial condition perturbations. The effect of the model error, however, is better seen by comparing the two forecasts: with and without the model error term. This integral effect of the model error on the forecast is quite substantial and comparable to the effect of the initial conditions error, as can be seen later in the text.

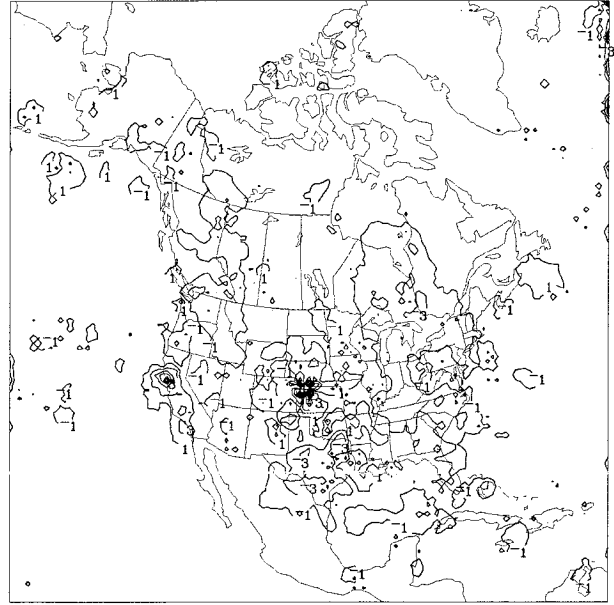
Even though the presented data assimilation results indicate the importance of the weak constraint and are in agreement with the theory that requires a weak constraint formulation in order to get a full benefit of a 4DVAR method, some caution should be taken when making conclusions about the effects of the model error. First, we use a very crude approximation to the model error covariance, by deriving it from the background error covariance. This might have, to some extent, limited the benefits of the weak constraint assumption. It was argued by Dee (1995), using an idealized model and data, that a misrepresentation of the model error covariance matrix can considerably deteriorate the data assimilation results. Second, it is important to note that the model error definition given by (2.9) is an approximation to the Markov process variable, because of the coarser timescale used for the random error term. In fact, we use an average random error term during a time interval (3 h or more in our experiments), rather than an instantaneous one. At this point it is not clear how

4D VAR 12-H ASSIM, ITER=10
SRFCE PRESS MODEL ERROR VALID 03Z 8 MAY 95

3-H FCST
80KM/17LYRS

4D VAR 12-H ASSIM, ITER=10
SRFCE PRESS MODEL ERROR VALID 06Z 8 MAY 95

6-H FCST
80KM/17LYRS



4D VAR 12-H ASSIM, ITER=10
SRFCE PRESS MODEL ERROR VALID 09Z 8 MAY 95

9-H FCST
80KM/17LYRS

4D VAR 12-H ASSIM, ITER=10
SRFCE PRESS MODEL ERROR VALID 12Z 8 MAY 95

12-H FCST
80KM/17LYRS

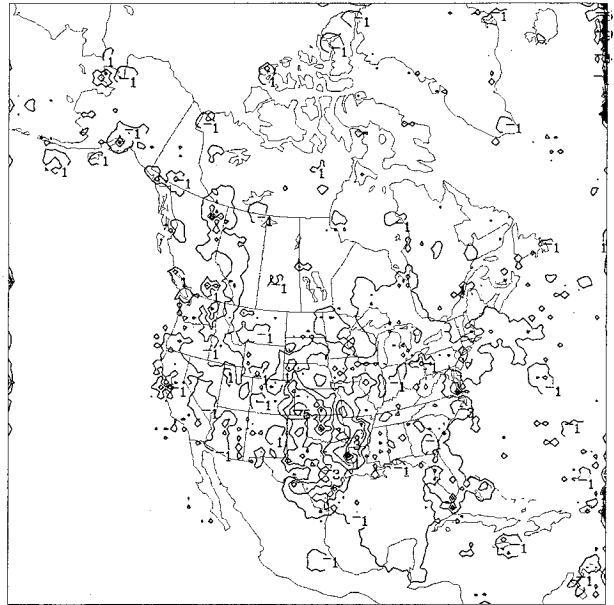
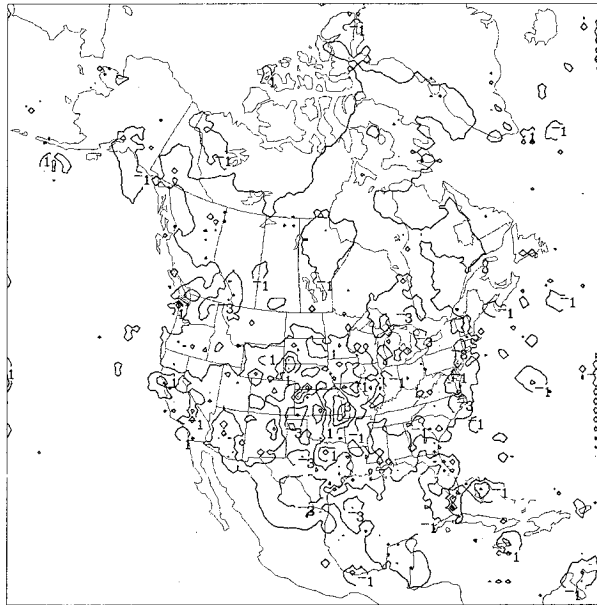


FIG. 5. Surface pressure component of the 3-h random error term obtained after 10 iterations of the minimization process (experiment ERROR 03). (a)–(d). Four time components during 12-h data assimilation periods. Contouring interval is 2 Pa.

much the results are sensitive to these assumptions. Even though the answer to this question cannot be obtained before the experiments with the less crude assumptions are conducted, the forecast experiments presented in the following section will, at least, contribute to a better understanding of the model error effects.

b. Forecast experiments

In the figures below we present the forecast results (0–48-h forecasts initiated at the end of the data assimilation period at 1200 UTC 8 May 1995). We compare the results of different data assimilation approaches us-

4D VAR 12-H ASSIM, ITER=10
SRFCE PRESS MODEL ERROR VALID 12Z 8 MAY 95

12-H FCST
80KM/17LYRS

4D VAR 12-H ASSIM, WITH MODEL ERROR, ITER=10
IC PERTURBATION VALID 00Z 8 MAY 95

SURFACE PRESSURE
80KM/17LYRS

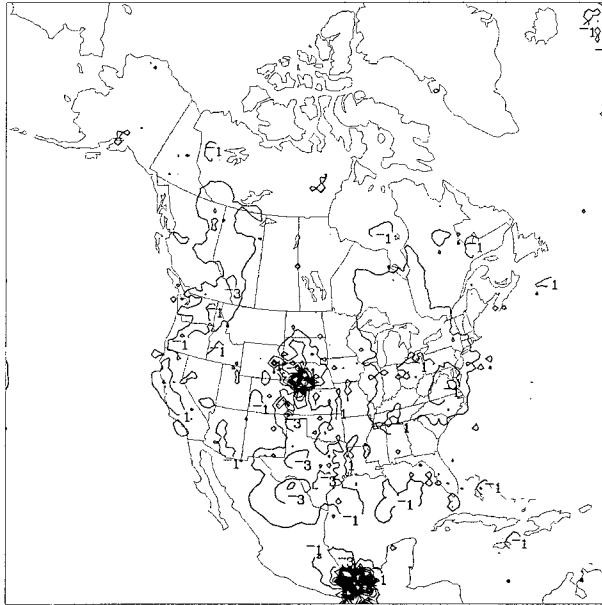


FIG. 6. Surface pressure component of the 12-h model error term obtained after 10 iterations of the minimization (ERROR 12).

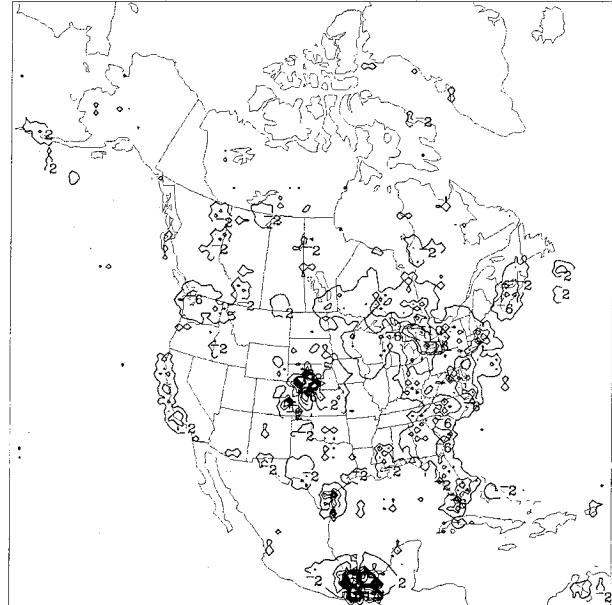


FIG. 8. As in Fig. 7 but for experiment ERROR 12.

4D VAR 12-H ASSIM, WITH MODEL ERROR, ITER=10
IC PERTURBATION VALID 00Z 8 MAY 95

SURFACE PRESSURE
80KM/17LYRS

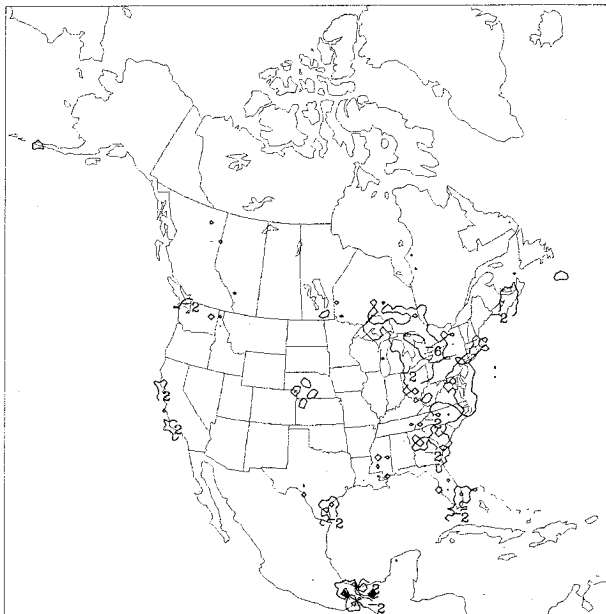


FIG. 7. The optimal perturbation to the surface pressure initial conditions obtained after 10 iterations of the minimization process in the experiment ERROR 03. The optimal perturbation is calculated by taking the difference between the initial conditions in the first and in the last iteration. Contouring interval is 4×10 Pa.

4D VAR 12-H ASSIM, NO MODEL ERROR, ITER=10
IC PERTURBATION VALID 00Z 8 MAY 95

SURFACE PRESSURE
80KM/17LYRS

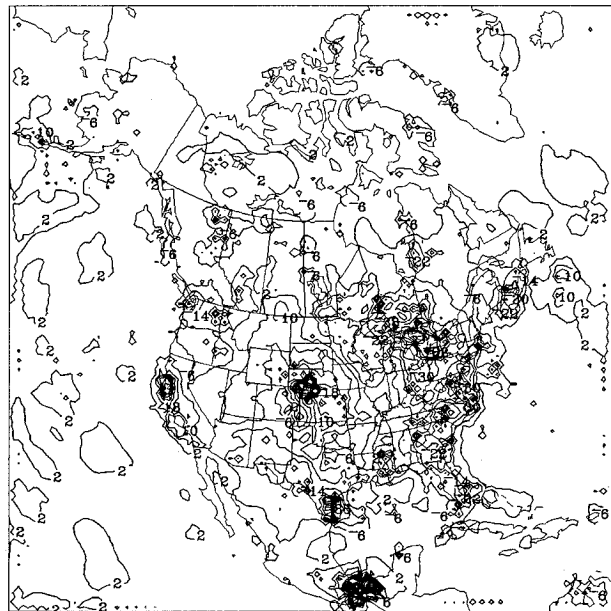


FIG. 9. As in Fig. 7 but for experiment NO ERROR. Note that the contouring interval is doubled here (8×10 Pa).

ing scatter diagrams. Figures 10 and 11 show scatter diagrams, experiment ERROR 12 versus ERROR 03, for temperature and wind rms errors, respectively. By examining the figures, we notice that there is a benefit

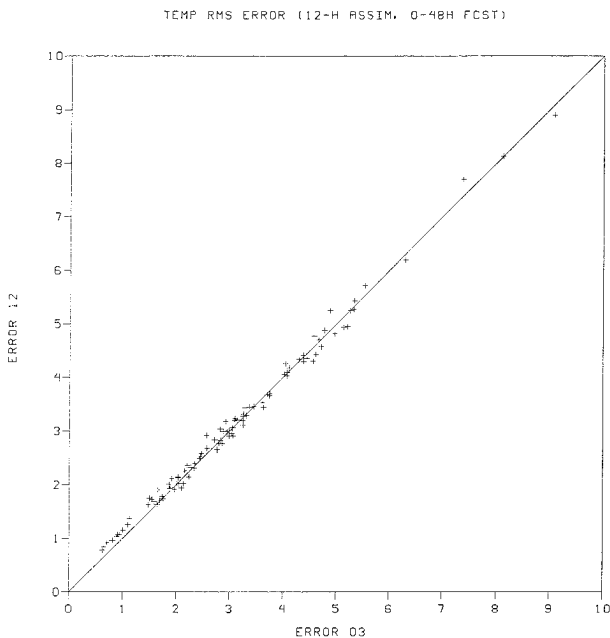


FIG. 10. Scatter diagram of the rms temperature error (K) for experiment ERROR 12 vs ERROR 03. The results of the forecast (0–48 h) originated at the end of data assimilation period, verified against observations every 12 h, at 20 pressure levels (as explained in section 5) are shown.

of using a finer-resolution random error term in the 4DVAR method. The majority of the points show smaller rms errors in experiment ERROR 03 than in experiment ERROR 12. The difference between the two experiments is rather small in the temperature and it is more pronounced in the wind field. It is not clear why this is so. For clarification, we ran the 4DVAR experiments in several different synoptic situations. The results showed that, generally after the 4DVAR data assimilation, there is a slight trend in the direction of having a better forecast for the wind field as compared to the other fields. This is not due to a better performance of the minimization for this variable, because the functional decreases similarly for all components of the control variable. It is more likely that, at present, our 4DVAR system is somewhat better tuned to wind than to mass field. For example, we know that in the mass field, especially in the lowest eta layers, there is a rather crude definition of the interpolation operator H . Also, there are some indications that the observational error (co)variances for the surface heights are less accurate than for the other fields, which can also hurt the mass field. This problem will be further examined in the future.

In order to see the impact of the weak constraint we compare experiments ERROR 03 and NO ERROR. In Figs. 12 and 13 we present the scatter diagrams, NO ERROR versus ERROR 03, for temperature and wind rms errors, respectively. As we see, a great majority of

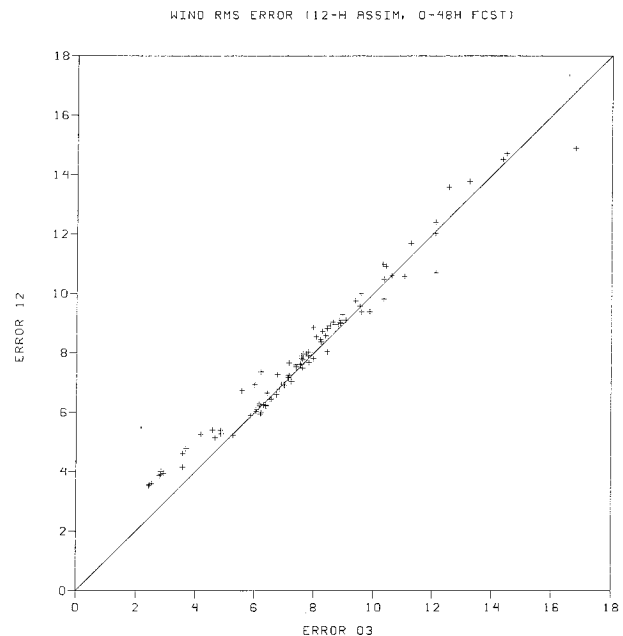


FIG. 11. As in Fig. 10 but for the wind component (m s^{-1}).

the points lie on one side of the diagram, indicating a substantial improvement in the forecast when using the 4DVAR method with a weak constraint to define the initial conditions. It is even more pronounced with the longer data assimilation periods (e.g., 4DVAR data assimilation over a 24-h period; figures not presented). A consistent improvement is also found in all other fields,

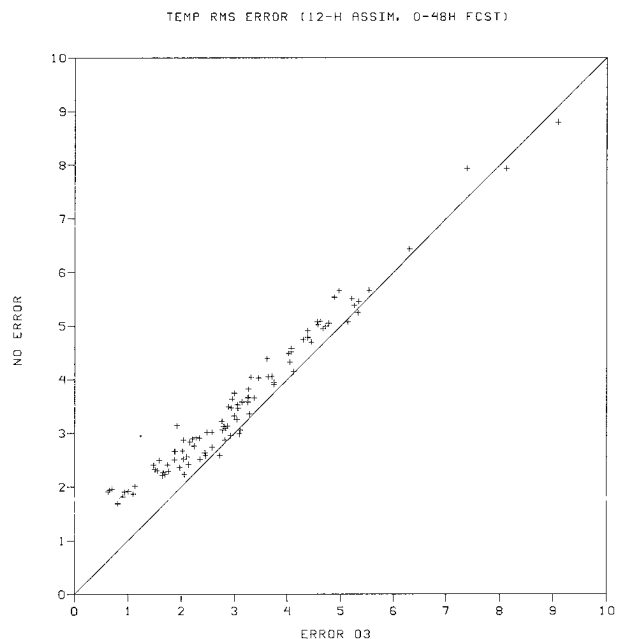


FIG. 12. As in Fig. 10 but for experiment NO ERROR vs ERROR 03.

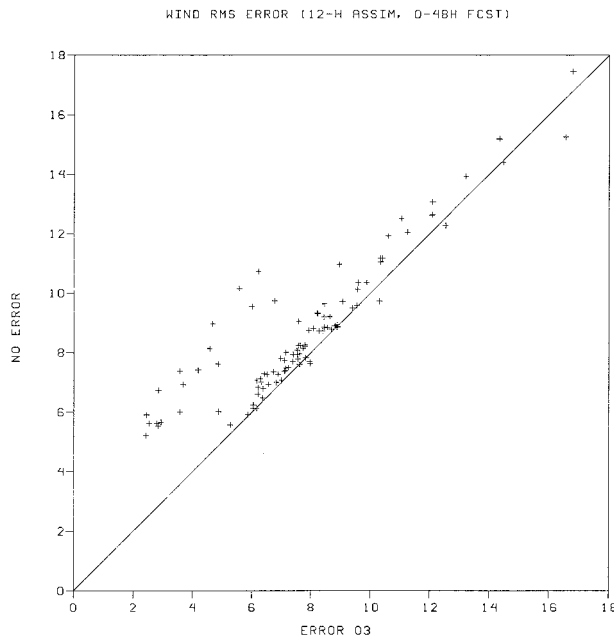


FIG. 13. As in Fig. 11 but for experiment NO ERROR vs ERROR 03.

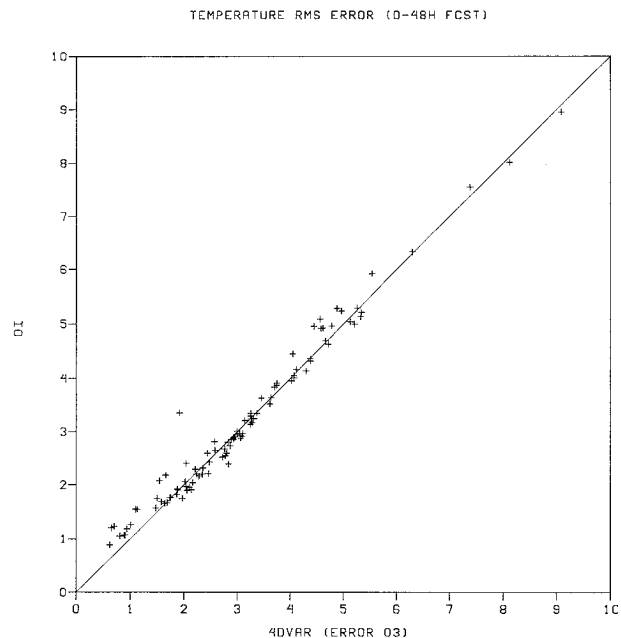


FIG. 14. As in Fig. 10 but for experiment OI vs ERROR 03.

such as surface pressure, humidity, precipitation, etc. As mentioned before, the assumptions made in the present 4DVAR system might have limited, to some extent, the capability of the 4DVAR system. Therefore, any further improvement, such as using a more realistic model error covariance, should lead to even better performance of the 4DVAR method with a weak constraint. By comparing Figs. 12 and 13 with the previous figures (Figs. 10 and 11), we conclude that it is indeed important to apply the forecast model as a weak constraint, and even by accounting for the systematic error only (ERROR 12 in our case) one can significantly improve the results. Evidence that a systematic error term can substantially improve the 4DVAR results is also shown in M. Zupanski (1993).

The importance of a weak constraint can also be verified when comparing the results of the 4DVAR experiments with the OI experiments. In Figs. 14 and 15 the scatter diagrams for experiments OI versus ERROR 03 are presented. By comparing these figures with the rms errors in the previous scatter diagrams, we can immediately see that the weak constraint *was* the factor that made it possible for the 4DVAR to slightly outperform the OI method. These results indicate that, for 12-h and longer data assimilation periods, the model error is, at least, equally, or even more important component of a 4DVAR system, compared to the initial conditions.

By examining specific forecast fields, we also found a considerable difference between different data assimilation approaches. To illustrate this, we show the plots of the 48-h forecast of the sea level pressure fields for experiments OI, ERROR 03, and NO ERROR, respec-

tively, in Figs. 16, 17, and 18. The verification is given in Fig. 1, where the NCEP surface weather map valid at the same time (1200 UTC 10 MAY 1995) is presented. By comparing Figs. 16 and 17 with the verification, we see that the quality of the two forecasts is comparable. However, one important difference can be seen: the center of the cyclone over the central United States is too deep and the size of the low pressure system is too large

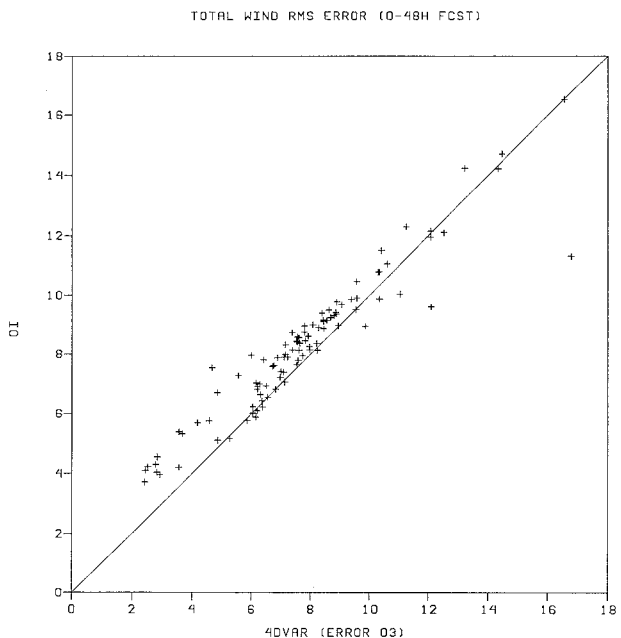


FIG. 15. As in Fig. 11 but for experiment OI vs ERROR 03.

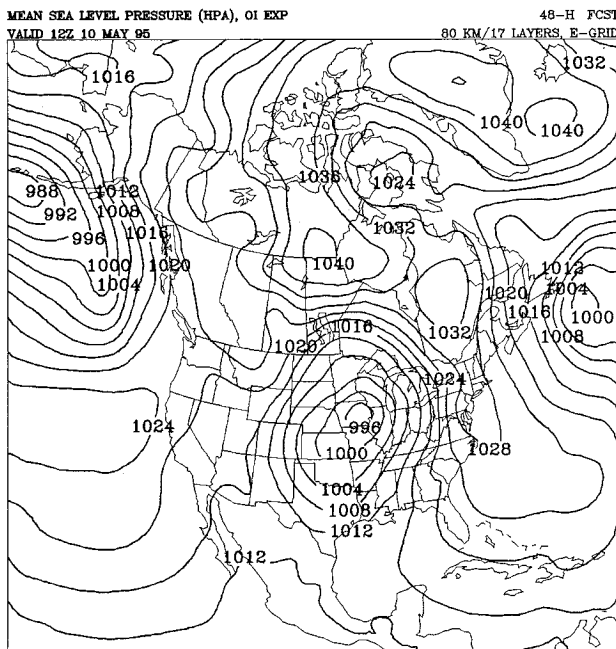


FIG. 16. The 48-h forecast of the sea level pressure field obtained in experiment OI. Contouring interval is 4 hPa.

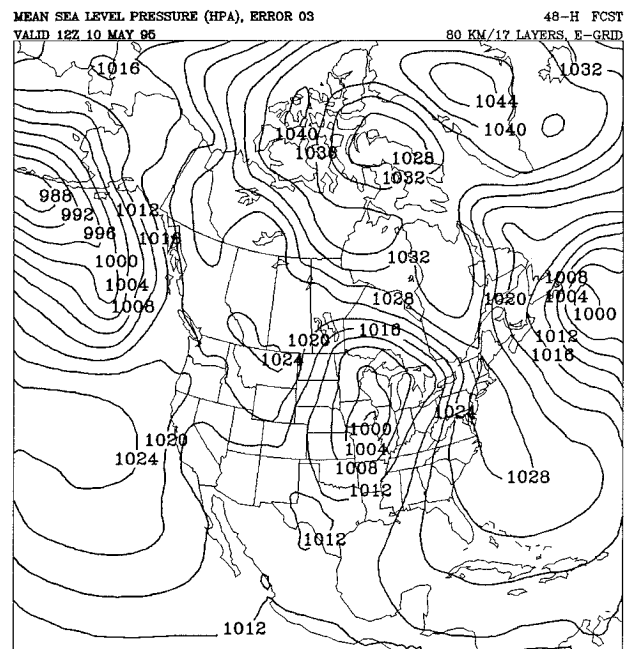


FIG. 17. As in Fig. 16 but for experiment ERROR 03.

in the OI experiment, as compared to the verification map and the weak constraint experiment (ERROR 03). The result of the third experiment, NO ERROR (strong constraint), is inferior to the other two experiments because of the positioning of the center of the cyclone too far in the southwest direction, and the cyclone being too shallow. Obviously, in this case, as in the previous results, the weak constraint was a factor that critically improved the 4DVAR forecast.

c. Computer power requirements

It is also important to assess the feasibility of the proposed weak constraint 4DVAR method from the aspect of the computational cost. As it is well known, the 4DVAR method is very expensive, concerning both computational space and time, even using only a strong constraint formulation. Because of this, only approximations to a general 4DVAR method should be considered as possible solutions for operational work, at this time. As compared to the strong constraint formulation, the weak constraint 4DVAR presented here adds a negligible amount of the computational time (needed to add the model error term to the model equations) for each time step. Concerning the computational space, the increase in the size of the control variable is considerable: from 0.6×10^6 , in the strong constraint, to 3×10^6 , in the weak constraint formulation, and this difference to the power of 2, to store the error covariances, as explained in section 3. These requirements, however, can be reduced to a great extent by using ad-

ditional disk space, instead of central processor space, by processing one vector at a time.

7. Summary and conclusions

We have presented a technique for applying the forecast model as a general weak constraint in a complex

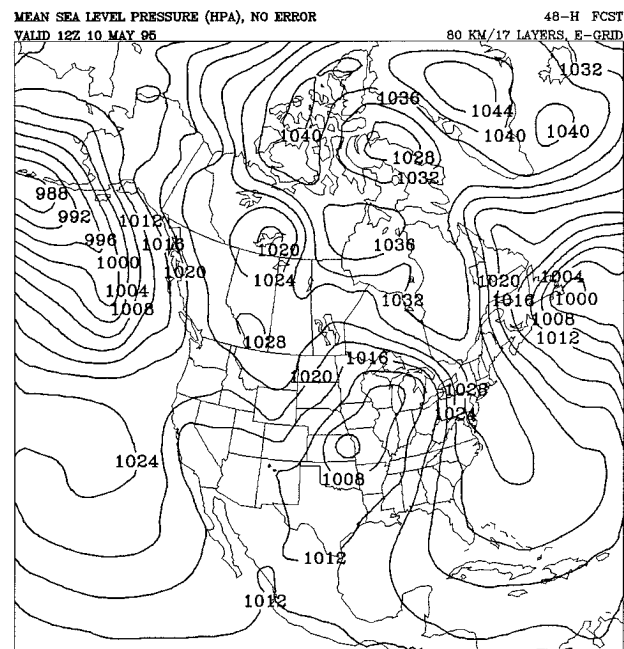


FIG. 18. As in Fig. 16 but for experiment NO ERROR.

variational algorithm, such as NCEP's regional 4DVAR data assimilation system. The proposed definition of the model error has a flexible time resolution for the random error term and has the potential for operational application, because the coarse time resolution and a diagonal in time random error covariance matrix \mathbf{W} require less computational space. Even though we did not consider a coarser space resolution for the random error term, the derivation is straightforward and can be used in order to further reduce the high demands for computational space for very fine resolution models.

The results presented in this study strongly indicate the need for a weak constraint (as opposed to a strong constraint formulation) in order to get a full benefit of a 4DVAR method. The inclusion of the model error term, even with the largest correlation timescale (12 h), gives a main contribution to the capability of the 4DVAR method to outperform the OI method. Reducing the correlation timescale to 3 h further improves the results, but less substantially, as our results indicate. This is not surprising, because the random perturbations tend, to some extent, to cancel each other out. Also, there is a possibility that the method might not get the full benefit from the complex model error due to the poor definition of the random error covariance matrix \mathbf{W} at this stage. In the next stage, the actual optimal values of the model error term can be used to calculate statistically, or to define a mathematical model of, the error covariance matrix and to improve our knowledge about it. In that case, a finer resolution of the random error term, which is more sensitive to the definition of the model error covariance matrix, can be used, if the computational space is available, to further improve the results.

Our results also indicate that the optimal initial conditions perturbation can be very different depending on the choice of the constraint (strong or weak) assumed for the forecast model. This is very important for sensitivity studies because, when determining the origin of the forecast error, the strong constraint assumption might be misleading.

In this study we have presented the results from a single synoptic case. Our experience from several different synoptic situations is very similar and indicates that the 4DVAR method with a weak constraint works well and that not more than 10 iterations of the minimization may be enough to considerably improve the forecast compared with OI results. Also, when running the 4DVAR data assimilation in a cycled mode (the results from the previous data assimilation cycle are used in the current cycle, as it is normally done in operational data assimilation systems), the number of iterations can be further reduced, because the initial state is closer to the minimum in that case [e.g., experiments in Zupanski and Zupanski (1996) showed that 3–5 iterations in a cycled mode would give the results similar to 10 iterations of a noncycled data assimilation]. This

is very encouraging for NCEP's plans for operational implementation in the near future.

Further improvements are also possible and should be expected. As mentioned above, the 4DVAR method can improve itself by using the actual data assimilation results to estimate background and model error statistics. (This statistical knowledge about the model error can be beneficial not only in 4DVAR, but in Kalman filter studies as well.) Other possible improvements of the 4DVAR method can be obtained by using observations such as precipitation, precipitable water, and clouds in a way consistent with the model dynamics. The ability to use observations not directly related to the model variables, such as satellite radiances and radar reflectivities, represents yet another advantage of the 4DVAR (and 3DVAR) system, as compared to OI data assimilation system. As explained in section 5a, the observation error covariance matrix should be correlated in time and in space, for some types of observations. These improvements will be considered in our future work.

Acknowledgments. I would like to thank M. Zupanski for many valuable discussions during this study and for providing the algorithm to calculate the background error correlation. The constructive comments by J. Purser and D. Parrish are very much appreciated. I am very thankful to R. McPherson, E. Kalnay, and J. Purser for thoughtful reviews of the manuscript. Special thanks go to J. Purser for correcting the English errors. My gratitude is extended to E. Rogers for help in preparing the observations and the EDAS analyses. Very thoughtful remarks made by Dr. Carl Hagelberg and an anonymous reviewer are greatly appreciated. This research was financially supported by the UCAR/NCEP Visiting Research Program.

APPENDIX A

Derivation of Eq. (2.6)

To derive the expression for δJ , Eq. (2.6), we take the first variation of Eq. (2.1), that is,

$$\delta J = \sum_{n=1}^N \delta \mathbf{y}_n^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta \Phi_m^T \sum_{j=1}^M \mathbf{Q}_{m,j}^{-1} \Phi_j + \delta \mathbf{x}_0^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b), \quad (\text{A.1})$$

and do the same with the model equations (2.1) and (2.3). Assuming $\boldsymbol{\varepsilon}_n = 0$, we have

$$\delta \mathbf{x}_m = \mathbf{G}_m \delta \mathbf{x}_{m-1} + \delta \Phi_m \quad (\text{A.2})$$

$$\delta \mathbf{y}_n = \mathbf{H}_n \delta \mathbf{x}_m, \quad (\text{A.3})$$

where \mathbf{G} and \mathbf{H} are linearizations of G and H , respectively. Then, after combining (A.2) and (A.3) and repeating the process of taking the first variations, one can obtain the linearized (tangent linear) model equa-

tion, depending only on the initial perturbation $\delta\mathbf{x}_0$ and the model error perturbations $\delta\Phi_m$, defined as

$$\begin{aligned}\delta\mathbf{y}_n &= \mathbf{H}_n \mathbf{G}_m \mathbf{G}_{m-1} \mathbf{G}_{m-2} \cdots \mathbf{G}_1 \delta\mathbf{x}_0 + \mathbf{H}_n \delta\Phi_m \\ &\quad + \mathbf{H}_n \mathbf{G}_m \delta\Phi_{m-1} + \mathbf{H}_n \mathbf{G}_m \mathbf{G}_{m-1} \delta\Phi_{m-2} + \cdots \\ &\quad + \mathbf{H}_n \mathbf{G}_m \mathbf{G}_{m-1} \mathbf{G}_{m-2} \cdots \mathbf{G}_2 \delta\Phi_1.\end{aligned}\quad (\text{A.4})$$

Here we assume that the model error perturbation at initial time $\delta\Phi_0$ is equal to zero. This assumption, however, can be easily relaxed, by allowing for $\delta\Phi_0 \neq 0$, when deriving the tangent linear model (A.4), which would result in an additional term in (A.4).

Finally, after including (A.4) in (A.1) and taking an adjoint of a linear operator in a Hilbert space, we obtain

$$\begin{aligned}\delta J &= \delta\mathbf{x}_0^T \sum_{n=1}^N \mathbf{G}_1^T \cdots \mathbf{G}_{M-1}^T \mathbf{G}_M^T \mathbf{H}_n^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta\mathbf{x}_0^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) + \delta\Phi_M^T \sum_{n=1}^N \mathbf{H}_n^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) \\ &\quad + \delta\Phi_M^T \sum_{j=1}^M \mathbf{Q}_{M,j}^{-1} \Phi_j + \cdots + \delta\Phi_{k-1}^T \sum_{n=1}^N \mathbf{G}_k^T \cdots \mathbf{G}_{M-1}^T \mathbf{G}_M^T \mathbf{H}_n^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta\Phi_{k-1}^T \sum_{j=1}^M \mathbf{Q}_{k-1,j}^{-1} \Phi_j + \cdots \\ &\quad + \delta\Phi_1^T \sum_{n=1}^N \mathbf{G}_2^T \cdots \mathbf{G}_k^T \cdots \mathbf{G}_{M-1}^T \mathbf{G}_M^T \mathbf{H}_n^T \mathbf{R}^{-1} (\mathbf{y}_n - \mathbf{y}_n^{\text{obs}}) + \delta\Phi_1^T \sum_{j=1}^M \mathbf{Q}_{1,j}^{-1} \Phi_j.\end{aligned}\quad (\text{A.5})$$

APPENDIX B

Model Error Covariance Matrix

a. Model error definition

Let us now write Eq. (2.9) in the more convenient form

$$\begin{aligned}\Phi_m &= \alpha \Phi_{m-1} + \beta \mathbf{r}_N; \quad \alpha + \beta = 1 \\ N &= 1, \dots, N_{\text{max}} \\ m &= 1, \dots, M_{\text{max}},\end{aligned}\quad (\text{B.1})$$

where the index m defines model time steps in the time interval N , corresponding to the coarse timescale random error component \mathbf{r}_N . By applying the recursive formula (B.1) in different time intervals ($N = 1, \dots, N_{\text{max}}$ and assuming Φ_0 (the initial model error) is known, we obtain the equation for the model error in the form

$$\begin{aligned}\Phi_m(N) &= \alpha^m \Phi_{M_{\text{max}}}(N-1) \\ &\quad + (\alpha^{m-1} + \alpha^{m-2} + \cdots + \alpha^0) \beta \mathbf{r}_N, \\ &= \alpha^m \Phi_{M_{\text{max}}}(N-1) + (1 - \alpha^m) \mathbf{r}_N.\end{aligned}\quad (\text{B.2})$$

Note that, due to the coarser timescale of the random error term, the model error defined by (B.2) or (2.9) is not a first-order Markov process variable (first-order Markov variable at time m depends only on the random forcing at the same time and on the variable at the previous time $m-1$). However, by introducing the new variable $\mathbf{X}_N = \Phi_{M_{\text{max}}}(N)$, defined only on the coarse timescale, Eq. (B.2) becomes the expression for the classic first-order Markov process:

$$\mathbf{X}_N = \gamma \mathbf{X}_{N-1} + (1 - \gamma) \mathbf{r}_N; \quad N = 1, \dots, N_{\text{max}}, \quad (\text{B.3})$$

where

$$\gamma = \alpha^{M_{\text{max}}}. \quad (\text{B.4})$$

Therefore, the model error defined by (B.3) or (2.9) is a Markov process on the coarse timescale. On the fine timescale, within the random time interval N , it is a deterministic process, where the random error \mathbf{r}_N is applied as a mean deterministic forcing.

Let us now derive the relevant components of the model error covariance matrix. We first start with the covariances on the coarse timescale $\langle \mathbf{X}_N \mathbf{X}_K^T \rangle$.

b. Model error covariances on the coarse timescale

By multiplying (B.3), respectively, by \mathbf{X}_N^T , \mathbf{X}_K^T , \mathbf{r}_N^T , and \mathbf{r}_K^T and applying the mathematical expectation operator, after simple algebra we obtain the corresponding components of the model error covariance matrix as

$$\langle \mathbf{X}_N \mathbf{X}_N^T \rangle = \frac{(1 - \gamma)^2}{1 - \gamma^2} \mathbf{W} = \mathbf{Q}_{M_{\text{max}}, M_{\text{max}}, N, N} \quad (\text{B.5})$$

$$\langle \mathbf{X}_N \mathbf{X}_K^T \rangle = \gamma^{|N-K|} \frac{(1 - \gamma)^2}{1 - \gamma^2} \mathbf{W} = \mathbf{Q}_{M_{\text{max}}, M_{\text{max}}, N, K} \quad (\text{B.6})$$

$$\langle \mathbf{X}_N \mathbf{r}_N^T \rangle = (1 - \gamma) \mathbf{W} \quad (\text{B.7})$$

$$\langle \mathbf{X}_N \mathbf{r}_K^T \rangle = \gamma^{N-K} (1 - \gamma) \mathbf{W}; \quad N \geq K \quad (\text{B.8})$$

$$\langle \mathbf{X}_N \mathbf{r}_N^T \rangle = 0; \quad N < K. \quad (\text{B.9})$$

In the above equations a stationarity in $\langle \mathbf{r}_N \mathbf{r}_N^T \rangle$ and $\langle \mathbf{X}_N \mathbf{X}_N^T \rangle$ is assumed and $\mathbf{W} = \langle \mathbf{r}_N \mathbf{r}_N^T \rangle = \langle \mathbf{r}_K \mathbf{r}_K^T \rangle$ is defined. Also, for the Markov process variable, the correlations $\langle \mathbf{r}_N \mathbf{r}_K^T \rangle$ are equal to zero for $N \neq K$, and $\langle \mathbf{X}_N \mathbf{r}_K^T \rangle = 0$ for $N < K$. To denote the dependence of the model error covariance \mathbf{Q} on both fine (small letters) and coarse timescales (capital letters), we introduced the four indices m , k , N , and K .

As we can see from (B.6), the model error correlations

between different times decay exponentially with time. For $\alpha \approx 0.7$ and $M_{\max} \geq 50$, as in our experiments, these correlations are practically negligible.

c. Model error covariances on the fine timescale

On the fine timescale, the model error covariance matrix includes two different types of components: 1) $\langle \Phi_m(N) \Phi_k(N)^T \rangle$, where both model error components belong to the same random time interval N ; and 2) $\langle \Phi_m(N) \Phi_k(K)^T \rangle$, where the model error components belong to different random time intervals N and K .

1) DERIVATION OF $\langle \Phi_m(N) \Phi_k(N)^T \rangle$

As in the previous case, we obtain the various components of the model error covariance matrix by multiplying (B.2) by the appropriate transposed variable. As a result, we get

$$\langle \Phi_m(N) \Phi_k(N)^T \rangle = \alpha^{m+k} \langle \Phi_{M_{\max}}(N-1) \Phi_{M_{\max}}(N-1)^T \rangle + (1 - \alpha^m)(1 - \alpha^k) \langle \mathbf{r}_N \mathbf{r}_N^T \rangle. \quad (\text{B.10})$$

By recalling the definition (B.4) and applying Eq. (B.5), we obtain the model error covariance

$$\begin{aligned} \mathbf{Q}_{m,k,N,N} &= \langle \Phi_m(N) \Phi_k(N)^T \rangle \\ &= \alpha^{m+k} \frac{(1 - \alpha^{M_{\max}})^2}{1 - \alpha^{2M_{\max}}} \mathbf{W} \\ &\quad + (1 - \alpha^m)(1 - \alpha^k) \mathbf{W}, \end{aligned} \quad (\text{B.11})$$

which indicates a strong correlation throughout the whole time interval N . From (B.11) one can easily obtain the diagonal component

$$\mathbf{Q}_{m,m,N,N} = \alpha^{2m} \frac{(1 - \alpha^{M_{\max}})^2}{1 - \alpha^{2M_{\max}}} \mathbf{W} + (1 - \alpha^m)^2 \mathbf{W}, \quad (\text{B.12})$$

which for $m = M_{\max}$ reduces to (B.5).

2) DERIVATION OF $\langle \Phi_m(N) \Phi_k(K)^T \rangle$

Multiplying (B.2) by $\Phi_k(K)^T$ and assuming $K > N$, we have

$$\begin{aligned} \langle \Phi_m(N) \Phi_k(K)^T \rangle &= \alpha^{m+k} \langle \Phi_{M_{\max}}(N-1) \Phi_{M_{\max}}(K-1)^T \rangle \\ &\quad + (1 - \alpha^m) \alpha^k \langle \mathbf{r}_N \Phi_{M_{\max}}(K-1)^T \rangle. \end{aligned} \quad (\text{B.13})$$

After applying (B.8), we obtain

$$\begin{aligned} \mathbf{Q}_{m,k,N,K} &= \langle \Phi_m(N) \Phi_k(K)^T \rangle \\ &= \alpha^{m+k} \alpha^{(K-N)M_{\max}} \frac{(1 - \alpha^{M_{\max}})^2}{1 - \alpha^{2M_{\max}}} \mathbf{W} \\ &\quad + (1 - \alpha^m) \alpha^{(K-N)M_{\max}} (1 - \alpha^{M_{\max}}) \mathbf{W}. \end{aligned} \quad (\text{B.14})$$

As in the case of a coarse timescale, the correlations decay exponentially with time. For $m = k = M_{\max}$ (B.14)

reduces to (B.6). The model error covariance \mathbf{Q} is non-diagonal in time (and space), depending only on the diagonal in time random error covariance matrix \mathbf{W} . This allows us to simplify the minimization problem, as explained further in the text.

REFERENCES

- Bennett, A. F., L. M. Leslie, C. R. Hagelberg, and P. E. Powers, 1993: Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method. *Mon. Wea. Rev.*, **121**, 1714–1729.
- , B. S. Chua, and L. M. Leslie, 1996: Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys.*, **60**, 165–178.
- Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.*, **112**, 677–691.
- , and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and Arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.*, **112**, 693–709.
- , and —, 1992: The Betts–Miller scheme. *The Representation of Cumulus Convection in Numerical Models*, Meteor. Monogr., No. 46, Amer. Meteor. Soc., 107–121.
- Black, T. L., 1994: The new NMC mesoscale eta model: Description and forecast examples. *Wea. Forecasting*, **9**, 265–278.
- Bouttier, F., 1993: The dynamics of error covariances in a barotropic model. *Tellus*, **45A**, 408–423.
- Cohn, S. E., and D. F. Parrish, 1991: The behavior of forecast error covariances for a Kalman filter in two dimensions. *Mon. Wea. Rev.*, **119**, 1757–1785.
- , M. Ghil, and E. Isaakson, 1981: Optimal interpolation and the Kalman filter. *Proc. Fifth Conf on Numerical Weather Prediction*, Monterey, CA, Amer. Meteor. Soc., 36–42.
- , N. S. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867.
- Courtier, P., and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus*, **42A**, 531–549.
- Daley, R., 1992: The effect of serially correlated observations and model error on atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 164–177.
- Dee, D. P., 1991: Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.*, **117**, 365–384.
- , 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- , S. E. Cohn, A. Dalcher, and M. Ghil, 1985: An efficient algorithm for estimating noise covariances in disturbed systems. *IEEE Trans. Autom. Control*, **AC-30**, 1057–1065.
- DeMaria, M., and R. W. Jones, 1993: Optimization of a hurricane track forecast model with the adjoint model equations. *Mon. Wea. Rev.*, **121**, 1730–1745.
- Derber, J. C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2446.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10 143–10 162.
- , and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85–96.
- Gaspari, G., and S. E. Cohn, 1996: Construction of correlation functions in two and three dimensions. DAO Office Note 96-03, 38 pp. [Available from G. Gaspari, Code 910.3, NASA/Goddard Space Flight Center, Greenbelt, MD 20771; also on-line from <http://hera.gsfc.nasa.gov/subpages/office-notes.html>]
- Gauthier, P., P. Courtier, and P. Moll, 1993: Assimilation of simulated

- wind lidar with a Kalman filter. *Mon. Wea. Rev.*, **121**, 1803–1820.
- Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. *Dyn. Atmos. Oceans*, **13**, 171–218.
- , S. E. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction. *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil, and E. Kallen, Eds., Springer-Verlag, 139–224.
- Houtekamer, P. L., L. Lefaire, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- Janjic, Z. I., 1990: The step-mountain coordinate: Physical package. *Mon. Wea. Rev.*, **118**, 1429–1443.
- , 1994: The step-mountain eta coordinate model: Further development of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945.
- , F. Mesinger, and T. L. Black, 1995: The pressure-advection term and additive splitting in split-explicit models. *Quart. J. Roy. Meteor. Soc.*, **121**, 953–957.
- Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.
- Lacis, A. A., and J. E. Hansen, 1974: A parameterization of the absorption of solar radiation in the earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133.
- Lobocki, L., 1993: A procedure for the derivation of surface-layer bulk relationships from simplified second-order closure models. *J. Appl. Meteor.*, **32**, 126–138.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- Lynch, P., and X. Y. Huang, 1992: Initialization of the HIRLAM model using a digital filter. *Mon. Wea. Rev.*, **120**, 1019–1034.
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.*, **20**, 851–875.
- Menard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221–237.
- Mesinger, F., Z. I. Janjic, S. Nickovic, D. Gavrilo, and D. Deaven, 1988: The step mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Mon. Wea. Rev.*, **116**, 1493–1518.
- Navon, I. M., X. Zou, J. C. Derber, and J. G. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.*, **120**, 1433–1446.
- Rogers, E., D. G. Deaven, and G. J. DiMego, 1995: The regional analysis system for the operational “early” eta model: Original 80-km configuration and recent changes. *Wea. Forecasting*, **10**, 810–825.
- Sasaki, Y., 1970: Some basic formalisms on numerical variational analysis. *Mon. Wea. Rev.*, **98**, 875–883.
- Schwarzkopf, M. D., and S. B. Fels, 1991: The simplified exchange method revisited: An accurate, rapid method for computation of infrared cooling rates and fluxes. *J. Geophys. Res.*, **96**(D5), 9075–9096.
- Shanno, D. F., 1978: Conjugate gradient methods with inexact line search. *Math. Operations Res.*, **3**, 244–256.
- , 1985: Globally convergent conjugate gradient algorithms. *Math. Program.*, **33**, 61–67.
- Thepaut, J. N., D. Vasiljevic, P. Courtier, and J. Pailleux, 1993: Variational assimilation of conventional meteorological observations with a multilevel primitive equation model. *Quart. J. Roy. Meteor. Soc.*, **119**, 153–186.
- Todling, R., and M. Ghil, 1994: Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one-layer results. *Mon. Wea. Rev.*, **122**, 183–204.
- Tsuyuki, T., 1996: Variational data assimilation in the tropics using precipitation data. Part II: 3-D model. *Mon. Wea. Rev.*, **124**, 2545–2551.
- Wergen, W., 1992: The effect of model errors in variational assimilation. *Tellus*, **44A**, 297–313.
- Zou, X., I. M. Navon, and J. G. Sela, 1993: Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus*, **45A**, 370–387.
- , Y.-H. Kuo, and Y.-R. Guo, 1995: Assimilation of atmospheric radio refractivity using a nonhydrostatic adjoint model. *Mon. Wea. Rev.*, **123**, 2229–2249.
- Zupanski, D., 1993: The effects of discontinuities in the Betts–Miller cumulus convection scheme on four-dimensional variational data assimilation. *Tellus*, **45A**, 511–524.
- , and F. Mesinger, 1995: Four-dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.*, **123**, 1112–1127.
- Zupanski, M., 1993a: Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, **121**, 2396–2408.
- , 1993b: A preconditioning algorithm for large scale minimization problems. *Tellus*, **45A**, 578–592.
- , 1996: A preconditioning algorithm for four-dimensional variational data assimilation. *Mon. Wea. Rev.*, **124**, 2562–2573.
- , and D. Zupanski, 1995: Recent developments of NMC's regional four-dimensional variational data assimilation system. *Proc. Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, WMO, 376–372.
- , and —, 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.