Tellus (1993), 45A, 435–448
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

The influence of diffusion and associated errors on the adjoint data assimilation technique

By YONG LI[1,*] and KELVIN K. DROEGEMEIER, *School of Meteorology and Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma 73019, USA*

(Manuscript received 27 October 1992; in final form 1 February 1993)

ABSTRACT

We investigate the influence of diffusion, and errors associated with its representation, on the adjoint data assimilation technique. In order to determine how diffusion influences the retrieval of the initial state given a set of observations at later times, we perform a linear analysis of the one-dimensional diffusion equation and show that the retrieved initial state will be amplified (smoothed) if the diffusion in the prediction model is larger (smaller) than that present within the observations. This amplification (smoothing) not only increases dramatically as the length scale of the feature under consideration decreases, but also plays a rôle in suggesting an appropriate time period within which to assimilate observed data. These results are verified numerically for the simple case of a rising thermal in a neutrally-stratified environment using simulated pseudo-observations from a dry, three-dimensional Boussinesq model and its adjoint.

1. Introduction

One of the most challenging and long-standing problems in numerical weather prediction (NWP) is the construction of a model initial state that is compatible with the governing equations, yet preserves the relevant information content of the observations: the so-called initialization/assimilation problem (e.g., NRC, 1991). Many strategies have been devised to deal with this problem, ranging from complicated schemes grounded in fundamental principles of dynamics (e.g., nonlinear normal mode initialization; Daley, 1981) to more mechanistic approaches which, though simple in concept and formulation, have proven quite successful (e.g., nudging; Hoke and Anthes, 1976).

It is important to note that fundamental differences exist between data assimilation on the large and small scales. For example, in the context of the large-scale atmosphere, all fields necessary for initializing a prediction model, except for the vertical velocity, can be obtained with varying degrees of accuracy directly from remote and in situ observing systems. Using the well-known balances of geostrophy and hydrostatic equilibrium, the mass and wind fields can be mutually adjusted to render a set of balanced fields, though the prognostic model equations are typically used to obtain a state appropriate for use as initial conditions (e.g., nonlinear normal mode initialization). For smaller-scale phenomena (e.g., thunderstorms), only the time history of the radial wind component from a Doppler radar, along with a proximity sounding and radar reflectivity, are typically known; all other physical variables must be deduced (Lilly, 1990). Unfortunately, no known balances exist on the storm scale; rather, all relationships among the wind and mass fields, as well as information regarding those fields not directly observable, must be determined by the

* Current affiliation: Supercomputer Computations Research Institute Florida State University, Tallahassee, FL 32306, USA.
[1] Corresponding author: General Sciences Corporation, A Subsidiary of Science Application International Corporation, 6100 Chevy Chase Dr., Laurel, Maryland 20707, USA.

equations governing the flow. Thus, the initialization/assimilation problem on the storm scale inherently involves elements of data retrieval and prediction modeling.

One particularly attractive method of data assimilation is based upon variational principles that seek to minimize, in a least squares sense, the discrepancy between model forecasted states and a set of observations through successive adjustments to model control parameters, usually the initial conditions. Marchuk (1975), Penenko and Obratsov (1976), and Le Dimet (1982) were among the first to have implemented a particular variational method, known as the adjoint, to evaluate the sensitivity of a model forecast to the initial state. By running the forward-in-time prediction and backward-in-time adjoint models in a successive and iterative fashion, the misfit between the model solution and a set of observations can be objectively minimized, via adjustments to the initial conditions, until a converged state (i.e., optimal initial condition) is reached.

The adjoint assimilation technique has seen increasing use in atmospheric and oceanic applications as a dynamically-consistent way of blending model data and observations, e.g., Courtier (1985), Derber (1985), Lewis and Derber (1985), Le Dimet and Talagrand (1986), Thacker and Long (1988), Courtier and Talagrand (1990). Operational centers are now exploring the adjoint as a candidate data assimilation technique for global, hemispheric, and regional models, e.g., the National Meteorological Center (NMC) (Navon et al., 1992) and the European Centre for Medium Range Weather Forecasts (ECMWF) (Thépaut and Courtier, 1991). The adjoint method is not without its problems or limitations, however. For example, there exist fundamental questions concerning the uniqueness of the retrieved solution (e.g., Li, 1991; Gauthier, 1992), the representation of moist processes and other physics that are of a discontinuous (on/off) nature, and concern about the number of iterations required to achieve a converged state, particularly in an operational environment. With hope that these and other problems can be resolved, research in adjoint development and testing is proceeding at a rapid pace, principally because the adjoint technique represents a theoretically optimal approach to data assimilation.

Sun et al. (1991) demonstrated the potential of using the adjoint technique to retrieve the three-dimensional wind and temperature fields from a time series of model-generated pseudo-observations of a single horizontal velocity component similar to that obtainable from a single Doppler radar. Other related studies have also been conducted by Kapitza (1991) and Li (1990), and each suggests a potential for success while pointing to important insolved problems, such as the inclusion of moist processes, that must be addressed before the method is of practical use for deep convection.

Qiu and Xu (1992) recently developed a simplified adjoint method for retrieving wind data from single-Doppler observations. Because their technique utilizes only one equation that describes the Lagrangian conservation of reflectivity, it is, in contrast to other adjoint approaches, extremely efficient. The Qiu and Xu scheme has successfully been applied to turbulent boundary-layer as well as convective outflow (microburst) data, and is now being extended to other physical regimes (Q. Xu, personal communication).

Because the two models associated with the adjoint methodology (the forward-in-time prediction model and its associated backward-in-time adjoint) contain irreversible diffusive effects, both physical and numerical in most cases, information contained in the initial state may be damped, and in some cases even lost completely, as the physical system evolves. In such cases, it may be difficult or even impossible to infer the initial field from observations at later times. If the model diffusion is represented incorrectly, which is nearly always the case for sub-grid scale turbulence parameterizations, or if unphysical, numerically-induced diffusion is exceedingly large, the retrieved fields could be substantially distorted, especially for small-scale features.

In a study using uniformly-distributed data, Thacker (1989) found that the difference in retrieved states between the eastern and western halves of a modeled ocean basin could be explained by the effects of diffusion. In particular, he argued that the eastern boundary flow was subject to greater diffusion than that at the western boundary because the former was characterized by relatively short-wavelength Kelvin waves, as opposed to the Rossby waves for the latter.

In comparative experiments with and without explicit horizontal diffusion in the ECMWF general circulation model and its adjoint, Thépaut and Courtier (1991) found degradation in the quality of the retrieved fields when horizontal diffusion was included. Although they attributed this result to the loss of information associated with diffusive processes, additional analyses and experiments are needed to fully understand the role of diffusion in data retrieval, particularly with regard to inaccuracies in model-parameterized diffusion as well as spurious diffusion or inherent diffusion associated with numerical algorithms.

It is important to recognize that errors other than those induced by diffusion, some perhaps scale-selective in nature, can impact the quality of retrieved fields in variational data assimilation. For example, as pointed out by one reviewer, phase speed errors are a possible source of degradation in the retrieved state, and methods for dealing with it and similar problems are now being explored (e.g., Brewster, 1991; Wergen, 1992). Derber (1989) developed a somewhat novel approach (VCA, or variational continuous assimilation) for dealing with various sorts of errors that might be related to both the data and the model itself (e.g., systematic biases). Specifically, he added a correction term to the governing equations to serve as a control variable, and through adjustments to it in a continuous manner, obtained forecasts superior to those obtained using the more traditional adjoint approach.

As a step toward developing a more complete understanding of various types of errors, both physical and computational, in variational assimilation techniques, we investigate here the effects of second-order, constant-coefficient diffusion on the retrieval of a model initial state using the adjoint method. We begin with a linear analysis of the one-dimensional diffusion equation (Section 2) and show that the assimilated initial state will be amplified (smoothed) if the explicit diffusion in the prediction model is larger (smaller) than that of the observations. This amplification (smoothing) is shown to increase dramatically as the wavelength of disturbance decreases (Sections 2, 3). In Section 4, we verify numerically the theoretical results for the simple case of a rising thermal using a dry, three-dimensional Boussinesq model and its adjoint, and discuss the degree to which diffusion affects the efficiency of the assimilation process. Finally, we summarize and discuss our results in Section 5.

2. Linear analysis

In order to examine the influence of diffusion on the retrieval of a model initial state in a simple context, consider the one-dimensional linear diffusion equation:

$$\frac{\partial a(x, t)}{\partial t} = \nu \frac{\partial^2 a(x, t)}{\partial x^2}, \tag{1}$$

where $a(x, t)$ is a function of space (x) and time (t) and ν is a constant diffusion coefficient. The solution to (1) can be expressed by harmonic functions of the form:

$$a(x, t) = \sum_{k=0}^{\infty} a_k(x, t), \tag{2}$$

where

$$a_k(x, t) = A_k(t) \cos(kx). \tag{3}$$

Substituting (2) into (1) leads, with the use of (3), to the well-known result:

$$a_k(x, t) = a_k(x, 0)\, \mathrm{e}^{-k^2 \nu t}. \tag{4}$$

It is convenient to define an *intensity ratio*, $r_k(t)$, which measures the amplitude fraction present at time t relative to the initial state, as

$$r_k(t) = \frac{\|a_k(x, t)\|}{\|a_k(x, 0)\|} = \mathrm{e}^{-k^2 \nu t}, \tag{5}$$

where $\|*\|$ denotes the amplitude of a particular mode contained in the total field. As $t \to \infty$, every mode with $k \neq 0$ tends to vanish. If the intensity ratio of a particular mode for times $t > T$ is less than a small positive value ε, where ε is perhaps a function of observational and model errors or uncertainties, then it might be possible to retrieve the initial state only if information *prior* to time T is used. In other words, the exponentially-increasing effects of diffusion may become so large beyond a certain time as to render available data useless in the reconstruction of the initial state. To quantify this hypothesis, we define an *influence time scale*, T_c, such that when $t \geqslant T_c$, $r_k(t) \leqslant \varepsilon$. Thus, any observational information available beyond the influence time scale will not be useful for inferring the initial state.

If we set $r_k(t) = \varepsilon$, then eq. (5) can be written as:

$$T_c = \frac{L^2 \ln(1/\varepsilon)}{4\pi^2 \nu}, \tag{6}$$

where L is the wavelength defined by $L = 2\pi/k$. The physical interpretation of eq. (6) is straightforward: the influence timescale is inversely proportional to the diffusion coefficient, demonstrating that increased diffusion reduces the time interval over which observed data are useful in retrieving the initial state, as hypothesized. Further, because T_c varies as the square of the wavelength, small-scale features are much more difficult to retrieve than large-scale features because data for the former must be available more frequently, e.g., if observations are available every 15 min for a characteristic length scale L, virtually none of the information associated with L may be retrievable if, for example, $T_c = 5$ min.

Although the term involving ε in eq. (6) is somewhat difficult to quantify and is problematic, we can easily estimate how changes in ε will impact T_c. Consider the situation in which ε decreases by 4 orders of magnitude, say from 10^{-4} to 10^{-8}, representing, for example, an improvement in the accuracy of the model or the observations. In this case, $\ln(1/\varepsilon)$ doubles from 9.2 to 18.4, representing only a factor of two increases in the influence time scale.

Because most numerical models of atmospheric motion are not designed to represent features down to the molecular scale, the "large eddy simulation" (LES) approach is applied in which unresolvable processes are treated statistically through the use of closure models which parameterize subgrid-scale motions in terms of the resolved mean flow (e.g., Lilly, 1962; Smagorinsky, 1963; Deardorff, 1973, 1975; Mellor and Yamada, 1982).

Consider the one and one-half order turbulence closure scheme described by Deardorff (1975) and implemented in a three-dimensional convective storm model by Klemp and Wilhelmson (1978). In this parameterization, the eddy diffusivity of momentum is written

$$K_m = c_m E^{1/2} l,$$

where $E = \frac{1}{2}\bar{u}'^2$ is the subgrid-scale energy, $l = (\Delta x\, \Delta y\, \Delta z)^{1/3}$ is a length scale (Lilly, 1967), c_m is a coefficient with a nominal value of 0.2, and the overbar and prime represent a grid volume average and a deviation from it, respectively. Given a model resolution of $(\Delta x, \Delta y, \Delta z) = (1000, 1000, 500)$ m, as used by Klemp and Wilhelmson (1978), and $E = 1.0\ \mathrm{m^2\,s^{-2}}$, $K_m = 709.9\ \mathrm{m^2\,s^{-1}}$. Again using $\ln(1/\varepsilon) = 10$, the associated influence scale T_c is approximately 12 h for $L = 8$ km, a period comparable to the lifetime of squall lines and mesoscale convective systems. As a result, eddy diffusion may have a significant physical impact on small-scale features for these types of events, and thus its effects on data assimilation cannot be ignored.

Coupled with often large diffusivities associated with numerical solution techniques, as well as filtering imposed to maintain computational stability, one is faced in LES models with the use of physically-questionable parameterizations of subgrid-scale mixing processes. Consequently, it is important to investigate how inaccuracies associated with such parameterizations might influence the retrieval of the model initial state during data assimilation.

In the analysis that follows, we assume that subgrid-scale mixing can be represented by a constant-coefficient, second-order diffusion term. We seek to understand the behavior of the assimilaton process when the diffusivity of the model differs from that of the observations.

Let all variables generated by a forward prediction model be denoted "model", and all observed data denoted "obs". If $t = T$ in eq. (4), we can write for the observed data

$$a_k^{\mathrm{obs}}(x, T) = a_k^{\mathrm{obs}}(x, 0)\, \mathrm{e}^{-k^2 \nu^{\mathrm{obs}} T}, \tag{7a}$$

and, likewise for the model,

$$a_k^{\mathrm{model}}(x, T) = a_k^{\mathrm{model}}(x, 0)\, \mathrm{e}^{-k^2 \nu^{\mathrm{model}} T}. \tag{7b}$$

In this idealized situation, we can pose the retrieval problem as follows: given a set of observations at some particular time T, $a_k^{\mathrm{obs}}(x, T)$, find the associated model initial state, $a_k^{\mathrm{model}}(x, 0)$. For a perfect retrieval, the latter would be identical to $a_k^{\mathrm{obs}}(x, 0)$. Substituting $a_k^{\mathrm{obs}}(x, T)$ for $a_k^{\mathrm{model}}(x, T)$, we obtain

$$a_k^{\mathrm{model}}(x, 0) = a_k^{\mathrm{obs}}(x, 0)\, \mathrm{e}^{-k^2 (\nu^{\mathrm{obs}} - \nu^{\mathrm{model}})\, T}. \tag{8}$$

If the diffusion associated with the model is greater than that of the observations, i.e., $\nu^{\text{obs}} - \nu^{\text{model}} < 0$, a spurious amplification of the initial state will occur for $k \neq 0$. This amplification will be much stronger for small scales than for large scales, and thus the retrieved field would be expected to contain excessive small-scale noise.

Consider now the reverse situation, in which the model diffusion is smaller than that of the observations, i.e., $\nu^{\text{obs}} - \nu^{\text{model}} > 0$. In this case, eq. (8) shows that small scale features are damped in the assimilation process, resulting in a smoothed initial state in which the amplitudes of small-scale extrema are systematically reduced. These effects become more pronounced for larger T (see eq. 8), which is an obvious result given that the distortion of the initial field by diffusion depends exponentially upon time (eq. 4).

3. An analysis within the context of the adjoint method

We now extend the analysis of the previous section to the variational formulation of the adjoint method. In order to quantify the misfit between the model forecast and observations (in a least squares sense) over a period of time, we define a cost function J as follows:

$$J(a(x, t=0)) = \int_0^T \int_0^X \|a(x, t) - a^{\text{obs}}(x, t)\|^2 \, \mathrm{d}x \, \mathrm{d}t. \tag{9}$$

Decomposing the observational data into Fourier modes, observing the orthogonality property of the associated trigonometric functions, and considering that observational data are available at discrete time intervals T_i, it is more helpful to write the cost function as:

$$J(A_k(0), k = 0, 1, 2, \ldots, K) = \sum_{i=0}^{N} \sum_{k=0}^{K} (A_k(T_i) - A_k^{\text{obs}}(T_i))^2. \tag{10}$$

Without losing generality, let $T_0 = 0$ to yield

$$J(A_k(0), k = 0, 1, 2, \ldots, K) = \sum_{i=0}^{N} \sum_{k=0}^{K} (A_k(0)\, \mathrm{e}^{-k^2 \nu T_i} - A_k^{\text{obs}}(T_i))^2. \tag{11}$$

In order to arrive at the optimal initial state using the adjoint approach, J is minimized in an iterative fashion as follows. First, an initial guess is provided to the forward model and a forecast is generated, from which the cost function is computed given a time series of observations over the same period. Next, in order to improve the initial guess, the derivatives of J with respect to the initial conditions are computed from the backward-in-time adjoint model. Based on this information, the initial condition is adjusted and the process repeated until J is believed minimized.

Because we wish to examine the scale-selective nature of the diffusive process, it is prudent to express the derivatives of J with respect to the *amplitudes* of the initial modes, $A_k(0)$, as follows:

$$\frac{\partial J}{\partial A_k(0)} = \sum_{i=0}^{N} 2(A_k(0)\, \mathrm{e}^{-k^2 \nu T_i} - A_k^{\text{obs}}(T_i))\, \mathrm{e}^{-k^2 \nu T_i}. \tag{12}$$

The term $\partial J / \partial A_k(0)$ can be interpreted as the contribution by a particular mode to the gradient which, for the purposes of our discussion, may be written as:

$$\frac{\partial J}{\partial a(x, t=0)} = \sum_{k=0}^{K} \left[\frac{\partial J}{\partial A_k(0)} \Big/ \frac{\partial a(x, t=0)}{\partial A_k(0)} \right]. \tag{13}$$

Assume now that $\|\partial a(x, t=0) / \partial A_k(0)\|$ is of order unity for all k, which means that, at the initial time, all physically-relevant (i.e., model-resolved) scales in the total field are of equivalent amplitude. Suppose we insert data only at time T_i; then, from eq. (12), the amplitude of the kth mode contains the factor $\mathrm{e}^{-k^2 \nu T_i}$, which decreases dramatically as k increases, as demonstrated in the previous section. Even if data are inserted at multiple time levels, small-scale information would be expected to contribute very little to the gradient in eq. (13). In other words, the derivative of J with respect to the initial condition contains less information for small scales than for large scales even though both have the same amplitude at the initial time. Hence, one would expect the larger scales to be recovered *earlier* during the minimization process.

Consider now what happens when the model diffusion coefficient is dissimilar to that associated with the observations. We can express the observational data as

$$A_k^{\text{obs}}(T_i) = A_k^{\text{obs}}(0)\, \mathrm{e}^{-k^2 \nu^{\text{obs}} T_i}.$$

Ideally, when the minimization process converges, the gradient of the cost function vanishes. By setting the gradient to zero in eq. (12), we obtain

$$A_k(0) = \left[\frac{F(k, \nu^{\text{obs}}, \nu^{\text{model}}, N)}{F(k, \nu^{\text{model}}, \nu^{\text{model}}, N)} \right] A_k^{\text{obs}}(0), \tag{14}$$

where

$$F(k, \nu^1, \nu^2, N) = \sum_{i=0}^{N} \mathrm{e}^{-k^2(\nu^1 + \nu^2) T_i}.$$

The bracketed quantity in (14) is a multiplying factor that describes the distortion of the retrieved field resulting from the model's inaccurate representation of diffusion. If data are inserted at a single time level only, eq. (14) becomes the same as eq. (8) in the previous section.

When $\nu^{\text{model}} > \nu^{\text{obs}}$, the factor [ ] in (14) is larger than unity, and thus the mode being retrieved is amplified exponentially. When $\nu^{\text{model}} < \nu^{\text{obs}}$, small scale features are excessively damped (note that this effect increases as the assimilation period lengthens, see eq. 14), resulting in a retrieved initial field that may be smoother than desired.

For increasingly long assimilation periods, i.e., large T, the deviation of the multiplying factor in (14) from unity increases. This suggests that the influence of the model's inaccurate representation of diffusion also increases, which is consistent with the analysis given in the previous section.

4. Numerical experiments

4.1. *Models and experiment design*

Building upon the theoretical results of the preceding sections, we now verify our hypotheses using the dry 3-D Boussinesq prediction model and associated adjoint developed by Sun et al. (1991). The forward model equations can be written:

$$\frac{\partial v_i}{\partial t} = -\frac{\partial}{\partial x_j} v_i v_j - \frac{\partial \pi}{\partial x_i} + \delta_{i3}\theta + \nu \nabla^2 v_i,$$

$$\frac{\partial \theta}{\partial t} = -\frac{\partial}{\partial x_j} v_j \theta + \mathrm{Pr}^{-1} \nu \nabla^2 \theta,$$

$$\nabla^2 \pi = -\frac{\partial^2}{\partial x_i \partial x_j} v_i v_j + \frac{\partial \theta}{\partial x_3},$$

where θ, v_i, π, ν, Pr are the non-dimensional potential temperature, velocity components, pressure, diffusion coefficient and Prandtl number, respectively. Details regarding the numerical framework of both the forward and adjoint models may be found in Sun et al. (1991). Apart from the diffusion coefficient, all parameters are identical to those used by Sun et al.

The methodology of our experiments follows that adopted by Sun et al. (1991) and Kapitza (1991), in which a forward model is used to generate a set of pseudo-observations (hereafter, observations) that serve as Doppler radar data. From the set of model fields, we extract complete 3-D volumes of wind data every (nondimensional) timestep and use the horizontal components as observations in the adjoint retrieval technique. In reality, the most challenging and practical application of variational data assimilation involves retrieving all unobserved state variables given a horizontal (radial) wind component from a single Doppler radar. However, because our intent here is to emphasize the effects of diffusion on data assimilation, we chose a more basic scenario in which both horizontal wind components are known. Clearly, as demonstrated by others (e.g., Gal-Chen, 1978), this is sufficient for retrieving the vertical velocity (by continuity) as well as the thermodynamic and water substance (Ziegler, 1985) fields.

The model grid mesh used in all experiments contains $13 \times 13 \times 13$ points which, though quite coarse, is sufficient for our needs, particularly from a computational resources point of view since numerous integrations of the predictive and adjoint models are required to attain convergence in a single experiment. To avoid complications associated with gravity waves, we use a neutrally-stratified base state that is initially static and in hydrostatic balance. All lateral boundaries are periodic. Convective motion in the model is created by placing a warm thermal disturbance in

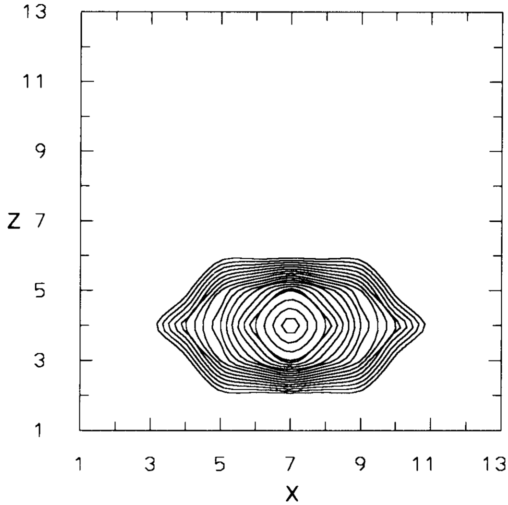

Fig. 1. $X-Z$ cross section of the initial temperature field for the control simulation. Contour interval is 0.2; maximum value is 4.0; and minimum value is 0.0.

the middle of the domain at the third grid level above the rigid, free-slip lower boundary. An $X-Z$ cross section through the center of the initial bubble at $t=0$ is shown in Fig. 1. This is the perfect initial temperature field against which retrieved initial states will be compared. As the model is integrated forward in time, the positive buoyancy of the thermal induces horizontal and vertical motions in response to associated pressure gradient forces, with the flow field characterized by one principal updraft and much weaker compensating downdrafts alongside.

The numerical experiments performed are listed in Table 1. Experiments 1 through 4 are designed to investigate the effects of diffusion when the diffusivity of the observations is equal to that used in the assimilation runs. The amplification of small-scale features in the retrieved fields, which occurs when the model diffusion exceeds that of the observations, is examined in experiments 5 through 7. Experiment 8 demonstrates the sensitivity of this amplification to the length of the assimilation window, while experiments 9 through 11 investigate the excessive smoothing of the retrieved fields when the model diffusion is smaller than that of the observations. Finally, experiment 12 extends the results of experiment 11 through a change to the length of the assimilation window.

In each of the experiments, 500 iterations (one iteration consists of two forward predictions and a backward adjoint integration) are performed in the minimization process. This rather large number is used to ensure a reduction in the cost function and gradient norm by at least 4 orders of magnitude, from $O(10^1)$ to $O(10^{-3})$, yielding a solution that is, for all practical purposes, converged. A zero first guess is used for the initial state in all experiments, and the numerical results justify the use of this condition and demonstrate that the cost function is well-behaved.

Table 1. *List of numerical experiments*

Experiment no.	Assimilation window (steps)	ν^{obs}	ν^{model}
experiment 1	12	0.003	0.003
experiment 2	12	0.006	0.006
experiment 3	12	0.009	0.009
experiment 4	12	0.012	0.012
experiment 5	12	0.003	0.006
experiment 6	12	0.003	0.009
experiment 7	12	0.003	0.012
experiment 8	18	0.003	0.006
experiment 9	12	0.006	0.003
experiment 10	12	0.009	0.003
experiment 11	12	0.012	0.003
experiment 12	18	0.006	0.003

4.2. *Numerical results*

4.2.1. *Diffusion of model and observations identical.* In the first suite of simulation experiments, we assume that the diffusivity of the observations is identical to that used in the assimilation runs. The diffusivity, ν, is specified to be 0.003, 0.006, 0.009, and 0.012 for experiments 1 through 4, respectively, of which the smallest value was chosen to maintain numerical stability while the largest produces fields that are smoothed but not excessively so. The cost function and gradient norm for these four experiments (plots not shown) all demonstrate similar convergence behavior. However, the larger diffusivity coefficients, the square error between the exact or control initial condition and that retrieved by the adjoint technique, shown in Fig. 2 for the temperature field, decreases at a significantly slower rate as the minimization proceeds. In other words, given a certain error tolerance for the retrieval, the number of iterations required to achieve the same level of accuracy increases with the diffusivity. As an

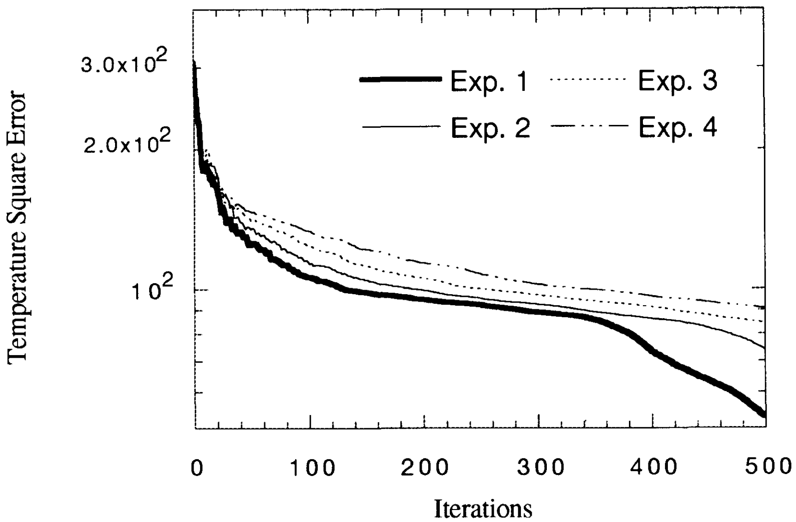

Fig. 2. Square Error of retrieved initial temperature field for experiment 1 ($\nu = 0.003$), experiment 2 ($\nu = 0.006$), experiment 3 ($\nu = 0.009$) and experiment 4 ($\nu = 0.012$).

example, 250 iterations with $\nu = 0.012$ are needed to produce a solution as accurate as 100 iterations with $\nu = 0.003$.

As suggested by the theoretical analyses in Sections 2 and 3, the differences described above can be attributed to the increased difficulty of retrieving small-scale features when larger diffusivities are used. To confirm this point with the numerical model, we plot the retrieved initial temperature field after 100 iterations for experiment 1, with $\nu = 0.003$, and for experiment 4, with $\nu = 0.012$, in Figs. 3a, b, respectively. The retrieved temperature field in experiment 4 is clearly smoother than that in experiment 1, with the amplitude of the thermal in the former case reduced by approximately 12%. These characteristics are still indentifiable even after 500 iterations (plots not shown).

It is also interesting to note that, for experiment 1, there exists a relatively steep slope in the square error of retrieved temperature after approximately 350 iterations (Fig. 2). Detailed plots (not shown here) indicate that this behavior results from the improved treatment of small-scale features which, though not significant to the overall quality of the retrieved thermal, are still indentifiable. None of the other three experiments exhibit this type of behavior in the square error.

4.2.2. Inaccuracy in the model representation of diffusion. As mentioned above, additional problems may be encountered if the diffusivity used in the numerical model does not match that of the observations. In experiments 5–8, all observations generated by the forward model are made using $\nu = 0.003$. However, the value of ν used in the adjoint model is larger, taking on values of 0.006, 0.009, and 0.012 for experiments 5–7 with an assimilation window of 12 time steps, respectively. Experiment 8 is the same as experiment 5, but with an assimilation window of 18 rather than 12 time steps. Again, the cost function and the gradient norm, which are not shown here, demonstrate convergence characteristics similar to those of experiments 1–4.

Fig. 4 shows the square errror of the retrieved temperature field as a function of iteration number for experiments 5–7. (Note that this quantity is not the cost function being minimized.) In each case, the errors decrease initially but then begin to increase as the iteration count grows. The larger the diffusion coefficient, ν, the greater the increase in the square error and the earlier this increase begins. A more thorough examination of the temperature field (not shown) suggests that the early decreases are due to the retrieval of the principal

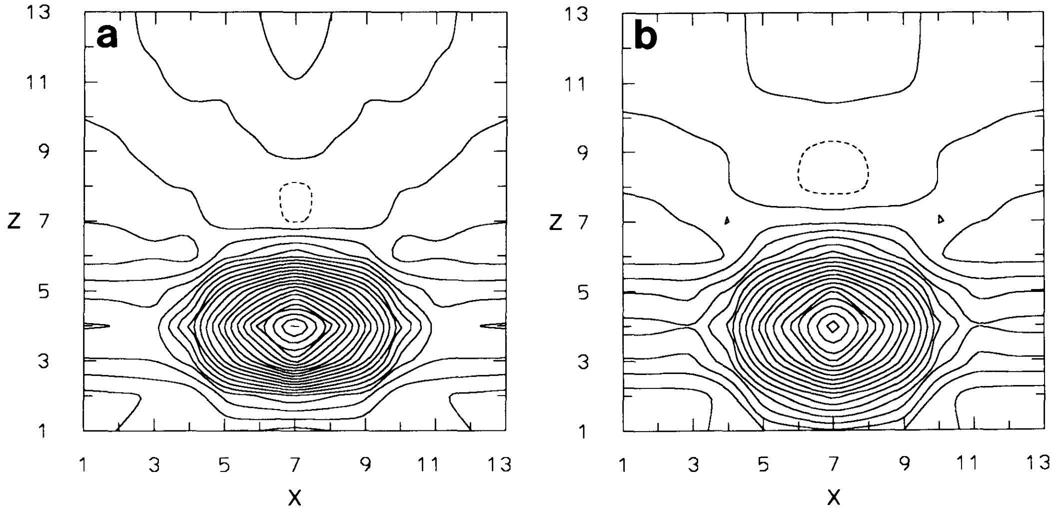

Fig. 3. Retrieved initial temperature field after 100 iterations. (a) Experiment 1 ($\nu = 0.003$; contour interval is 0.2; maximum value is 3.95; and minimum value is -0.2); (b) Experiment 4 ($\nu = 0.012$; contour interval is 0.2; maximum value is 3.6; and minimum value is -0.2).

structure of the thermal bubble, while the subsequent increase is due to the spurious amplification of small-scale features. To illustrate this point, we present in Fig. 5 the retrieved temperature field after 500 iterations for experiment 5. We indeed find a significant amplification of spurious small-scale noise, an effect that becomes more serious as the diffusion coefficient ν increases (plots not shown).

The theoretical analyses of the previous sections also suggest that a spurious amplification of small-scale features should accompany increases in the length of the assimilation window. This hypothesis is evaluated in experiment 8, which uses a

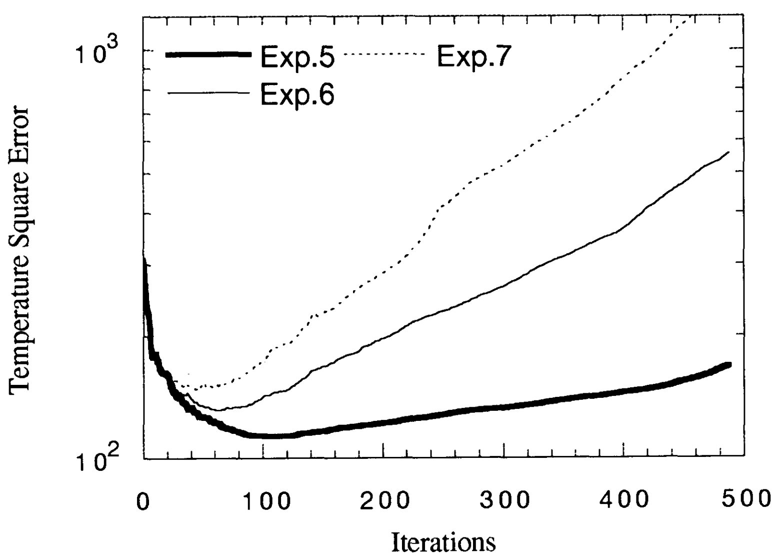

Fig. 4. Square error of retrieved initial temperature field for experiment 5 ($\nu^{\text{model}} = 0.006$), experiment 6 ($\nu^{\text{model}} = 0.009$) and experiment 7 ($\nu^{\text{model}} = 0.012$). The value of $\nu^{\text{obs}} = 0.003$ is used for all the three experiments.

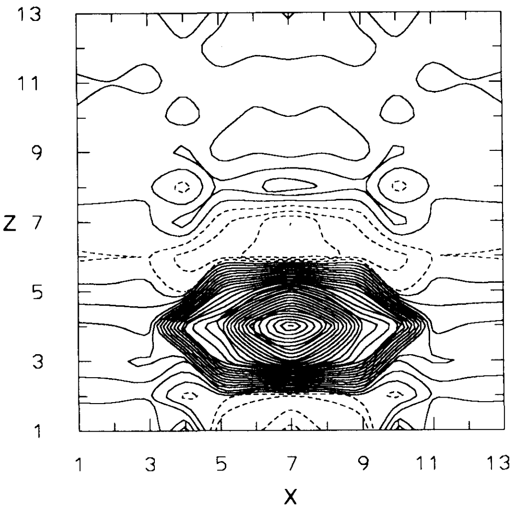

Fig. 5. Retrieved initial temperature field after 500 iterations for experiment 5 ($\nu^{obs} = 0.003$, $\nu^{model} = 0.006$). Contour interval is 0.2; maximum value is 5.8; and minimum value is -0.9.

lengthened assimilation window of 18 time steps, and visualized in Fig. 6 by comparing the associated square error of the temperature field with that produced by experiment 5 (assimilation window of 12 steps). During the first 80 or so iterations, the errors decrease at nearly the same rate for these two experiments. Thereafter, the error rises substantially faster for the longer assimilation window (experiment 8), a behavior that can be explained by examining the retrieved temperature field for experiment 8 (Fig. 7). The amplitudes of the spurious, small-scale features in Fig. 7 are increased (compare with Figs. 5 and 1), a feature which can be explained only by the differences in the length of the assimilation window.

Consider now the situation in which the model diffusion is smaller than that of the observations. Experiments 9–11 use a diffusivity of 0.003 for the forward/adjoint models, and values of 0.006, 0.009, and 0.012 for the observations. Fig. 8 shows the square error of the retrieved temperature fields. In each case, an initially rapid and then more subtle decrease in error, associated with the retrieval of the thermal bubble, is followed by a sudden increase which occurs at larger iteration counts for smaller differences in diffusivity between model and observations. This growth is easily traced to spurious damping induced by the diffusion process, as illustrated theoretically in the preceding section. This damping is documented more thoroughly, as an example, in Fig. 9, which shows

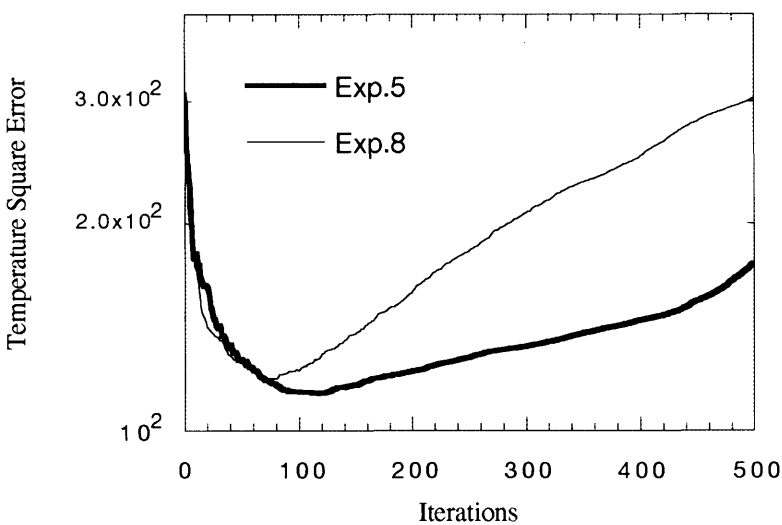

Fig. 6. Square error of retrieved initial temperature field for experiment 5 ($T = 12$ steps) and experiment 8 ($T = 18$ steps). In both these cases, $\nu^{obs} = 0.003$ and $\nu^{model} = 0.006$.

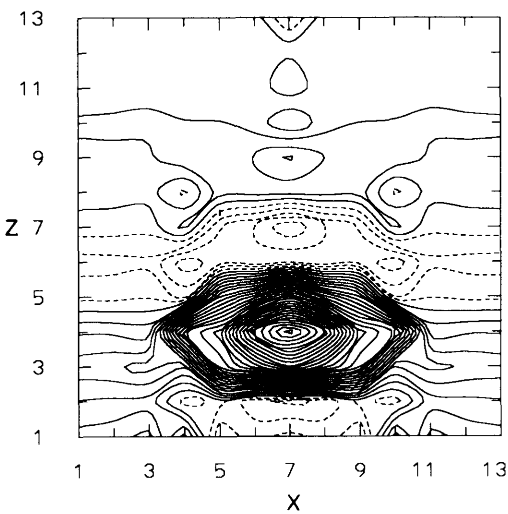

Fig. 7. Retrieved initial temperature field after 500 iterations for experiment 8 ($T = 18$ steps, $\nu^{obs} = 0.003$ and $\nu^{model} = 0.006$). Contour interval is 0.2; maximum value is 6.0; and minimum value is -1.2.

the retrieved temperature field for experiment 9 after 500 iterations. Note the excessive smoothing relative to the exact initial condition (compare with Figs. 3a, 1).

Finally, the results from experiment 12, which is identical to experiment 9 but uses an assimilation window lengthened to 18 timesteps, are shown in Fig. 10. Compared to the results shown in Fig. 9, the smoothing of the temperature gradient within the thermal, even though not significant, is nonetheless identifiable.

The above results suggest the existence of an optimal number of iterations for the minimization process. Interestingly, the use of additional iterations in the minimization appears to substantially worsen the quality of the retrieved instead of improving it when the diffusion is misrepresented by the models. Based on our results, it is not possible to determine the precise iteration count required for an acceptable retrieval, though it obviously depends upon the accuracy of the model's diffusion and the length of the assimilation

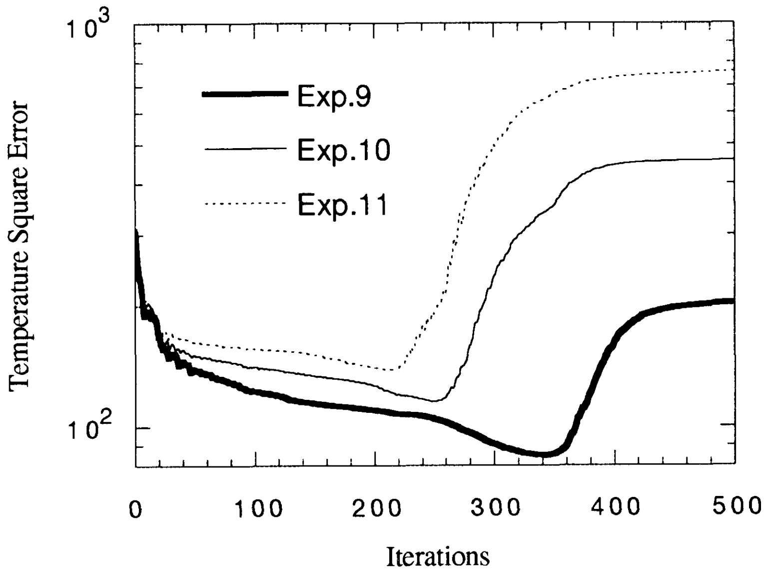

Fig. 8. Square error of retrieved initial temperature field for experiment 9 ($\nu^{obs} = 0.006$), experiment 10 ($\nu^{obs} = 0.009$) and experiment 11 ($\nu^{obs} = 0.012$). The value of $\nu^{model} = 0.003$ is used in all the three cases.

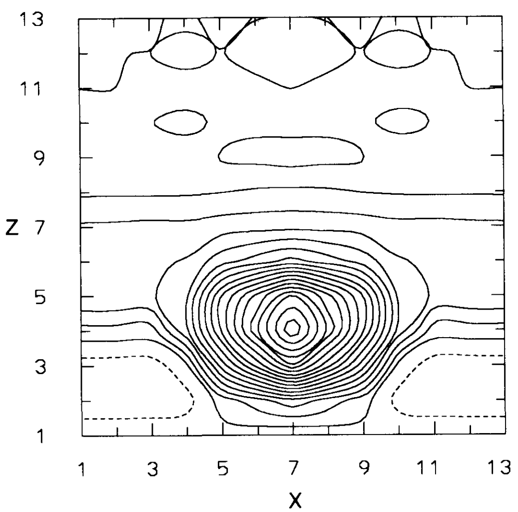

Fig. 9. Retrieved initial temperature field after 500 iterations for experiment 9 ($T = 12$ steps). The values of $\nu^{obs} = 0.006$ and $\nu^{model} = 0.003$. Contour interval is 0.2; maximum value is 3.6; and minimum value is -0.3.

window. Nevertheless, one should, with experience gained using a particular model, be able to make a rough estimate of the optimal iteration count, at least for a particular physical scenario.

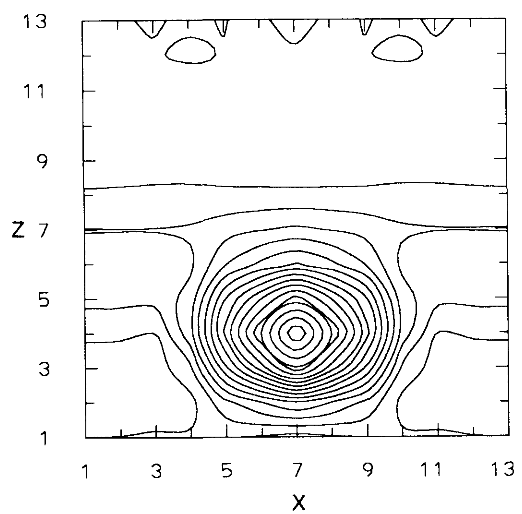

Fig. 10. Retrieved initial temperature field after 500 iterations for experiment 12 ($T = 18$ steps). The values of $\nu^{obs} = 0.006$ and $\nu^{model} = 0.003$. Contour interval is 0.2; maximum value is 3.4; and minimum value is -0.1.

5. Discussion and conclusions

In this paper, we examined the influence of second-order, constant-coefficient diffusion and its associated errors on the ability to retrieve the initial state of a model from subsequent observations using the adjoint method. Beginning with a linear analysis of the one-dimensional diffusion equation, we showed that the diffusive process damps information contained at the initial time. When the model diffusivity is larger (smaller) than that present in the observations, the retrieved initial field will be amplified (smoothed). These effects increase dramatically as the spatial scale of the structure of interest decreases, and as the length of the assimilation window increases.

To verify these results in a more complete manner, we used a 3-D Boussinesq model to simulate the evolution of a dry convective thermal, and then attempted to retrieve the known initial state from a subset of model data in a variety of circumstances. These simulation experiments supported the theoretical suppositions in all cases. More specifically, in situations where the model diffusion was prescribed to be less than that associated with the (model-generated) pseudo-observations, small-scale features were spuriously smoothed in the retrieved state. In the opposite situation, i.e., when the model diffusion was larger than that of the observations, small-scale features in the retrieved state were spuriously amplified, again leading to a misrepresentation of the initial state. In general, the spurious amplification (smoothing) of small-scale features increased with the length of the assimilation window (i.e., the time period over which data are assimilated). As a result, there appears to be an optimum assimilation timeframe within which the error between the correct initial state and that retrieved by the model is minimized. Of course, this timeframe will be intimately linked to the timescale of the phenomenon under consideration.

This study, which suggests that diffusion may play a key rôle in controlling the accuracy of the adjoint data assimilation methodology, is clearly not exhaustive. In particular, sensitivities associated with the frequency of data insertion and the quality of the inserted data need to be addressed, in both cases using more physically realistic scenarios and broader parameter spaces than those used here. In addition, the model-generated pseudo-observations must be replaced by real data, optimally where a single-Doppler wind component can be used in the assimilation procedure and the retrieved state verified against dual- or triple-Doppler data for the same event. Our study also points to the importance of developing improved parameterizations for sub-grid scale processes in convective- and meso-scale prediction and simulation models, and suggests that prudence should be exercised in the application of spatial filters designed to damp computational noise.

One strategy for improving the quality of the retrieval when the diffusivity is unknown or is unavailable is to include it as a control variable in the variational formulation. In this manner, it can be determined using the procedure described by Derber (1989) (see Section 1). For more realistic models, in which the diffusivity of heat, moisture, and momentum is a function of time and space, the size and complexity of the minimization problem will increase significantly, including the possibility of encountering multiple minima in the cost function and thus non-unique retrievals.

6. Acknowledgements

We wish to thank Drs. J. Sun and D. Lilly for providing the 3-D model for this study, and for their comments on the original manuscript. Discussions with Drs. J. Lewis, Y. Sasaki, I. M. Navon, W. C. Thacker, and A. Shapiro have proven very beneficial, and comments provided by anonymous reviewers improved the quality of the manuscript. The model experiments were performed on the Cray Y-MP and Cray-2 at the National Center for Supercomputing Applications, University of Illinois. This research was supported by the National Science Foundation under Grant ATM86-57013 to the second author and by Grant ATM88-09872 to CAPS, University of Oklahoma. The first author, while preparing this manuscript, was supported also by Dr. I. M. Navon at Florida State University through NASA Grant NAG 5-1660.

REFERENCES

Brewster, K. A. 1991. Phase-correcting data assimilation. Preprints, *9th Conference on Numerical weather prediction*, 14–18 October, Denver, Amer. Meteor. Soc., 610–613.

Courtier, P. 1985. Experiments in data assimilation using the adjoint model technique. *Workshop on High-resolution analysis*, ECMWF (UK) June 1985, 20 pp.

Courtier, P. and Talagrand, O. 1990. Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus 42A*, 531–549.

Daley, R. 1981. Normal mode initialization. *Rev. Geophys. Phys. 19*, 450–468.

Deardorff, J. W. 1973. The use of subgrid transport equations in a three dimensional model of atmospheric turbulence. *J. Fluids Eng. 95*, 429–438.

Deardorff, J. W. 1975. The development of boundary-layer turbulence models used in studying the severe storm environment. *Proc. SESAME Meeting*, Boulder, NOAA-ERL, 251–264.

Derber, J. C. 1985. *The variational 4-D assimilation of analyses using filtered models as constraints.* PhD thesis, University of Wisconsin-Madison, 142 pp.

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

Gal-Chen, T. 1978. A method for the initialization of the anelastic equations: Implications for matching models with observations. *Mon. Wea. Rev. 106*, 587–606.

Gauthier, P. 1992. Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model. *Tellus 44A*, 2–17.

Hoke, J. E. and Anthes, R. A. 1976. The initialization of numerical models by a dynamic initialization technique. *Mon. Wea. Rev. 104*, 1551–1556.

Kapitza, H. 1991. Numerical experiments with the adjoint of a nonhydrostatic mesoscale model. *Mon. Wea. Rev. 119*, 2993–3011.

Klemp, J. B. and Wilhelmson, R. B. 1978. The simulation of three-dimensional convective storm dynamics. *J. Atmos. Sci. 35*, 1070–1096.

Le Dimet, F. X. 1982. *A general formalism of variational analysis.* CIMMS Report, No. 22, School of Meteorology, University of Oklahoma, Norman, OK 73091, USA (available from the School of Meteorology).

Le Dimet, F. X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of Meteorological Observations: Theoretical Aspects. *Tellus 38A*, 97–110.

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

Li, Y. 1990. Assimilation research at Center for Analysis and Prediction of Storms: Application of an anelastic mesoscale model and its adjoint to data assimilation. *Proceeding of NCAR mesoscale data assimilation Colloquium.* June 1990, Boulder, Colorado (available from NCAR, P. O. Box 3000, Boulder, Colorado 80307, USA).

Li, Y. 1991. A note on the uniqueness problem of variational adjustment approach to four-dimensional data assimilation. *J. Meteor. Soc. Japan 69*, 581–585.

Lilly, D. K. 1962. On the numerical simulation of buoyant convection. *Tellus 14*, 148–172.

Lilly, D. K. 1967. The representation of small-scale turbulence in numerical simulation experiments. *Proc. IBM Sci. Comput. Symp. Environmental sciences*, White Plains, 195–210 (available from Center for Analysis and Prediction of Storms, University of Oklahoma, 100 East Boyd, Norman, Oklahoma 73019, USA).

Lilly, D. K. 1990. Numerical prediction of thunderstorms, has its time come? *Q. J. R. Meteorol. Soc. 116*, 779–798.

Marchuk, G. I. 1975. Formulation of the theory of perturbations for complicated models. *Appl. Math. and Optim. 2*, 1–45.

Mellor, G. L. and Yamada, T. 1982. Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys. Space Phys. 20*, 851–875.

National Research Council, 1991. *Four-dimensional model assimilation of data. A strategy for the earth system sciences.* National Academy Press, 78 pp.

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

Penenko, V. V. and Obratsov, N. N. 1976. A variational initialization method for the fields of the meteorological elements. English translations: *Soviet Meteorology and Hydrology 11*, 1–11.

Qiu, C.-J. and Xu, Q. 1992. A simple adjoint method of wind analysis for single-Doppler data. *J. Atmos. and Oceanic Tech. 9*, 588–598.

Smagorinsky, J. 1963. General circulation experiments with primitive equations (I). The basic experiment. *Mon. Wea. Rev. 91*, 99–164.

Sun, J., Flicker, D. W. and Lilly, D. K. 1991. Recovery of three-dimensional wind and temperature fields from simulated single-Doppler radar data. *J. Atmos. Sci. 48*, 876–890.

Thacker, W. C. and Long, R. B. 1988. Fitting dynamics to data. *J. Geophys. Res. 93*, 1227–1240.

Thacker, W. C. 1989. The rôle of the hessian matrix in fitting models to measurements. *J. Geophys. Res. 94*, 6177–6196.

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

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

Ziegler, C. L. 1985. Retrieval of thermal and microphysical variables in observed convective storms. Part I: Model development and preliminary testing. *J. Atmos. Sci. 42*, 1487–1509.