Tellus (1985), *37A*, 309–322

The use of adjoint equations to solve a variational adjustment problem with advective constraints

By JOHN M. LEWIS, *NOAA/NESDIS Development Laboratory, Systems Design and Applications Branch, 1225 West Dayton Street, Madison, Wisconsin 53706, USA* and JOHN C. DERBER *Space Science and Engineering Center, 1225 West Dayton Street, Madison, Wisconsin 53706, USA*

(Manuscript received August 20, 1984; in final form February 6, 1985)

ABSTRACT

A methodology is developed to guarantee time continuity in a sequence of analyses. Coupling is accomplished by requiring the least squares minimization of adjustment to the analyses subject to dynamic constraints. In this paper, the analyses are assumed to be governed by advective constraints such as those used in vorticity conservation models; however, the method can easily be applied to other constraints. The results correspond to the application of a strong constraint as introduced by Sasaki , but the procedure used to accomplish the minimization is an alternative to the traditional methods for solution of the Euler-Lagrange equation. The method allows easier inclusion of more time levels in the analysis sequence as well as accommodation of more complicated constraints.

The method is tested using both simulated and real data. The simulated data studies use a one-dimensional advection equation with progressively more complicated dynamics: constant advection velocity, spatially varying advection velocity, and nonlinear advection. The real data case study uses an advection of quasi-geostrophic potential vorticity constraint to examine the height adjustment process for three time periods on 6 March 1982. Separate studies are made for analyses derived from VAS and RAOB data. The results of this study indicate the method has excellent potential to reduce the random component of the analysis errors.

1. Introduction

Past data are routinely incorporated into the present analysis at operational prediction centers by using a background or guess field to initiate the analysis. The background field has traditionally been the 12 h forecast from a previous analysis. Thus, the past data are indirectly combined with the current data and the time coupling is pragmatically accomplished. With the advent of new atmospheric profiling systems such as the ground-based microwave profiler and the VISSR Atmospheric Sounder (VAS) (Hovermale, 1983), the coupling of successive analyses is complicated by the fact that observations are becoming quasi-continuous in time. Various methods for using this data have been proposed and referred to as four-dimensional data assimilation techniques. These methods have been thoroughly reviewed by Bengtsson (1975) with more recent work found in Bengtsson et al. (1981). The focus of this research is on developing a new methodology for dynamically coupling many analyses closely spaced in time.

Miyakoda and Talagrand (1971) attempted to use a sequence of simulated analyses for a period of several days in order to improve the initial conditions for a barotropic model. In their approach, the synthesized initial condition was a weighted average of the current analysis (initial condition) and predictions to the current time using the past analyses. They found that for a linear model with constant advecting velocity, the observational error can be curtailed by averaging the prediction results and analysis with equal weight. This is a special case of the method developed in this paper. However, for the nonlinear equation, a simple averaging did not improve the results. Instead, to account for the error growth with time, a weighted average was used by Miyakoda and Talagrand with the weights determined from the error covariances of the forecast model. Because of the statistical nature of the covariances, the optimal weights are difficult to find and generally vary from case to case.

If the emphasis is placed on temporal continuity in the sequence of data sets rather than just the initial conditions, the adjustments are best handled by minimizing the sum of squared discrepancies between the solution from a forecast model and the observations at the various times. Since the forecast model is required to fit the data at the various analysis times, it is unnecessary to ascribe weights according to the data's age. The primary advantage of this approach is the mutual adjustment of the analyses which effectively increases the data base at each time by using information at other analysis times through the forecast equations. This is the philosophy of traditional variational analysis as formulated by Sasaki (1958, 1970).

Although attractive in principle, the multiple time level variational problem is cumbersome for even simple advective constraints when traditional methods are used (Lewis and Panetta, 1983). The governing Euler-Lagrange (E-L) equations are typically a coupled set of high order differential equations that are difficult to solve under ideal circumstances, namely linear dynamics and analyses uniformly spaced in time. When more realistic circumstances prevail, rather large approximations are necessary to make the problem tractable (Bloom, 1983).

An attractive alternative to solving the E-L equations associated with the variational analysis problem has appeared in optimal control theory. LeDimet (1982) is developing a general formalism for meteorological variational analysis as an application of optimal control theory that builds on the ideas of Lions (1971). Rather than finding the solution through the E-L equations, the functional is minimized by finding its gradient with respect to one of the analysis states (e.g., the initial state) and then using methods such as steepest descent or conjugate gradient to iterate towards the optimal state. Finding the gradient of the functional involves the use of adjoint equations (e.g. Courant and Hilbert, 1962, p. 770) which are similar in complexity to the governing constraints and are used to integrate difference fields backwards in time. Unlike the more traditional variational approach, the methodology using the adjoint equations remains tractable even with complicated constraints.

The multi-time level variational adjustment problem is addressed here using these optimization strategies. Following the development of basic theory, some conceptual examples and numerical experiments will be presented using advection equations. Finally, the application of the method to a case study involving VAS data will be presented.

2. The method

In this section an interative method is described which minimizes the weighted squared difference between the original analyses at several times and the coincident solutions to the constraint for a given variable (q). This problem is solved using standard nonlinear minimization algorithms utilizing a gradient calculated from adjoint equations. The constraint is some forecast model providing estimates $q(t_p)$, where t_p refers to the analysis times. The final analyses are constrained to satisfy the model forecast ($q(t_p)$) from a set of initial conditions ($q(t_1)$). Mathematically, an equivalent problem is to find the initial conditions to the forecast model which produces a solution minimizing

$$J(q(t_1)) = \tfrac{1}{2} \sum_{p=1}^{P} \langle W(t_p)(q(t_p) - \tilde{q}(t_p)), \; q(t_p) - \tilde{q}(t_p) \rangle, \tag{1}$$

where $J(q(t_1))$ is the functional dependent on the initial conditions ($q(t_1)$), $\langle , \rangle$ is the inner product over space, $\tilde{q}(t_p)$ is a vector of length M containing the original analyzed gridpoint values of the variable q at time levels t_p, and $W(t_p)$ is a $M \times M$ diagonal weighting matrix, which can be used to discriminate confidence levels at the various analysis times. Here the functional is minimized with respect to $q(t_1)$ and accordingly the functional is expressed as $J(q(t_1))$. The functional could be minimized with respect to q at any time (e.g., $J(q(t_p))$). The result will be the same for any t_p when the constraint is linear. This can be seen intuitively by realizing that the model is considered perfect and once an optimal state is found at an arbitrary time, the associated optimal states are found by forecasting and/or hindcasting to the other times using the adjusted state as initial conditions. In the nonlinear case, more than one solution to the algorithm is possible with the solution determined by the initial guess. Thus, it is uncertain whether or not the algorithm will converge to the same solution when minimized with respect to different times. However, it can be shown that for the implicit time differencing used in this paper any solution found by minimizing with respect to one time level is also a solution with respect to another time level.

J is minimized by finding the gradient of J with respect to the initial conditions, $\text{Grad}\{J(q(t_1))\}$, followed by the use of a standard nonlinear minimization algorithm. In this study, the steepest descent algorithm was initially tested (e.g., Polak, 1971, pp. 28–40) and it worked well with uniform weighting. However, when nonuniform weighting was used, we found the conjugate gradient method (e.g., Polak 1971, pp. 53–54) to be more satisfactory. For simplicity only the application to steepest descent will be described. The generalization to conjugate gradient is straightforward.

The steepest descent algorithm is outlined as follows:

Step 1. Find the gradient of J, Grad J, with respect to some initial conditions $q^\nu(t_1)$. ν specifies the iteration.

Step 2. Find the optimal stepsize, ρ^ν, such that J is minimized along the line of possible initial conditions $q(t_1) = q^\nu(t_1) - \rho^\nu \text{ Grad } J^\nu$ ($\rho > 0$). For a linear model, ρ^ν can be calculated explicitly. However, for a nonlinear model the stepsize usually cannot be determined analytically. For the non-linear constraint in this paper a quadratic line search algorithm (e.g., Avriel (1976)) is used to estimate ρ^ν.

Step 3. Update the initial conditions $q^{\nu+1}(t_1) = q^\nu(t_1) - \rho^\nu \text{ Grad } J^\nu$.

Step 4. Check to see if some convergence criterion is met. If not, go to step 1. In this study, a convergence criterion requiring $|\text{Grad } J^\nu| < \varepsilon$ is used where ε is a small number. However, several different criteria could be successfully used and often the satisfaction of one implies satisfaction of the others. Note that if Grad $J = 0$, the functional is at a minimum for the linear constraint. If the constraint is nonlinear, the functional can be at any stationary point, i.e., a local minimum, local maximum, or saddle point. However, the functional value must be less than or equal to the original functional value.

In Subsection 2.1, the procedure for finding Grad J^ν is outlined. Subsection 2.2 describes a special case where the optimal initial field is a simple average of the analysis and hindcasts to the initial time. In Subsection 2.3, the relationship between this method and the one suggested by Thompson (1969) is described.

2.1. Determination of Grad J

Grad J^ν is determined using a method based upon a general formalism being developed by LeDimet (1982). In the development to follow, it will be noted that a linearization of the constraint is used to derive the expression for Grad J. Despite this linearization, the method produces the exact gradient since the discarded higher order terms are only important in evaluating higher order derivatives of the functional. If J' is defined as the first-order change in the functional resulting from a small perturbation $q'(t_1)$ about the initial condition $q(t_1)$, then J' is the directional derivate in the $q'(t_1)$ direction and is given by

$$J' = \langle \text{Grad } J, q'(t_1) \rangle. \tag{2}$$

J' can also be found using the definition of J in (1),

$$J' = \sum_{p=1}^{P} \langle W(t_p)(q(t_p) - \tilde{q}(t_p)), q'(t_p) \rangle \tag{3}$$

where $q'(t_p)$ is the perturbation in the forecast resulting from the perturbation to the initial conditions, $q'(t_1)$. By equating (2) and (3), Grad J can be found by expressing $q'(t_p)$ in terms of the initial conditions $q'(t_1)$.

To accomplish this, a linearized perturbation equation is derived. If the constraint is nonlinear, the basic state (around which the linearization is performed) changes with each iteration ν and is given by the solution to the complete nonlinear equations for that iteration. Using an implicit form of finite differencing, the linearized perturbation equation can be written as

$$q'(t) - q'(t - \Delta t) = A(t)(q'(t) + q'(t - \Delta t)), \tag{4}$$

where A is an $M \times M$ matrix which may be a function of $q(t)$ and $q(t - \Delta t)$, the solutions to the governing equation of constraint at iteration level ν.

Algebraic manipulations of (4) yield

$$q'(t) = [I - A(t)]^{-1}[I + A(t)]\, q'(t - \Delta t)$$
$$\equiv D(t) q'(t - \Delta t),$$
$$q'(t_p) = D(t_p) D(t_p - \Delta)$$
$$\times D(t_p - 2\Delta t) \ldots D(t_1 + \Delta t) q'(t_1)$$
$$= \left[\prod_{k=K_p}^{1} D(t_1 + k\Delta t)\right] q'(t_1), \qquad (5)$$

where K_p is the number of timesteps used to get from t_p to t_1.

With $q'(t_p)$ written in terms of $q'(t_1)$, (2) is equated to (3) and (5) substituted into the result giving

$$\langle \text{Grad}\, J, q'(t_1)\rangle$$
$$= \sum_{p=1}^{P} \left\langle \hat{q}_p(t_p), \left[\prod_{k=K_p}^{1} D(t_1 + k\Delta t)\right] q'(t_1)\right\rangle, \qquad (6)$$

where we have simplified the notation by introducing the adjoint variable $\hat{q}_p(t_p) \equiv W(t_p)(q(t_p) - \tilde{q}(t_p))$, a weighted difference between the forecast and analysis at $t = t_p$. Since the adjoint variable will be integrated back to the initial time t_1, the subscript p on $\hat{q}_p(t)$ is introduced to keep track of the time level where the adjoint variable was initialized. To isolate $q'(t_1)$ on the right hand side of the inner product, the identity for a real matrix A

$$\langle x, Ay\rangle = \langle A^{\mathrm{T}} x, y\rangle,$$

is applied repeatedly where A^{T} is the transpose of A, resulting in

$$\langle \text{Grad}\, J, q'(t_1)\rangle$$
$$= \sum_{p=1}^{P} \left\langle \left[\prod_{k=1}^{K_p} D^{\mathrm{T}}(t_1 + k\Delta t)\right] \hat{q}_p(t_p), q'(t_1)\right\rangle, \qquad (7)$$

thus implying that

$$\text{Grad}\, J = \sum_{p=1}^{P} \left[\prod_{k=1}^{K_p} D^{\mathrm{T}}(t_1 + k\Delta t)\right] \hat{q}_p(t_p). \qquad (8)$$

This form of Grad J is not very practical, with the $D^{\mathrm{T}}(t_1 + k\Delta t)$ matrices being very difficult to evaluate. Therefore, (8) is rewritten in a form similar to the basic constraint.

This is done by defining the adjoint variable at times preceding analysis time t_p as

$$\hat{q}_p(t - \Delta t) = D^{\mathrm{T}}(t) \hat{q}_p(t), \qquad (9)$$

and repeatedly substituting into (8) yielding

$$\text{Grad}\, J = \sum_{p=1}^{P} \hat{q}_p(t_1). \qquad (10)$$

Note that (9) expanded as in (4) gives

$$\hat{q}_p(t) - \hat{q}_p(t - \Delta t) = -A^{\mathrm{T}}(t) \cdot (\hat{q}_p(t) + \hat{q}_p(t - \Delta t)), \qquad (11)$$

which is referred to as the adjoint equation. Thus, Grad J^v can be found by integrating $\hat{q}_p(t_p) = W(t_p)(q(t_p) - \tilde{q}(t_p))$, the weighted differences between the model forecasts and the analyses, from $t = t_p$ to $t = t_1$ using the adjoint equation (11) for each t_p and summing the results. However, since (11) is linear, (10) can be evaluated with a single integration of the adjoint equation from the last analysis time to the initial time by inserting the weighted differences, $W(t_p)(q(t_p) - \tilde{q}(t_p))$, whenever one of the analysis times (t_p) is reached. Thus, Grad J is found with a single integration of the adjoint model.

2.2. *Unitary system*

A special case of this problem exists when $A(t) = -A^{\mathrm{T}}(t)$, thus implying that $D^{\mathrm{T}} = D^{-1}$. When $D^{\mathrm{T}} = D^{-1}$, D is referred to as a unitary matrix (Kreyszig, 1978, see pages 195–206). Then the adjoint integration of the weighted differences is equivalent to integrating these differences backwards in the model equations. The optimal stepsize can be shown to be $1/P$ and the method will converge in one iteration if all points are weighted equally ($W(t_p) = I$). Thus, at the minimum

$$q(t_1) = \frac{1}{P} \sum_{p=1}^{P} \left[\prod_{k=1}^{K_p} D^{-1}(t_1 + k\Delta t)\right] \tilde{q}(t_p). \qquad (12)$$

A unitary system with equal weights has optimal conditions given by the average of the initial analyses integrated to $t = t_1$. This result implies that the optimal forecast is the average of the forecasts. The linear vorticity equation with constant advection velocity studied by Miyakoda and Talagrand (1971) is an example of a unitary system.

2.3. Relationship to Thompson's scheme

When the integration of the linear advection equation from one analysis time to the other is done in one step, the results from this method can be shown to be equivalent to the Thompson (1969) approach as applied by Lewis (1982) or for multiple time levels as developed by Lewis and Panetta (1983). This intuitively follows since both methods are minimizing the same functional under the same constraint.

The minimization technique discussed in this paper has three major advantages over the Thompson method. First, the method allows multiple timesteps between analysis times thus eliminating the truncation problem associated with large separation between analysis times as discussed in Lewis (1982). Because of this, limitations on the maximum time separation between analysis times are eliminated. Also, the inclusion of many analysis times becomes almost trivial with this method, while as shown by Lewis and Panetta (1983) the inclusion of many times in the Thompson method is cumbersome. Finally, this method is easier to implement on more complicated model equation sets. A potential disadvantage of this method is the required amount of computer time. Since the method is iterative with each iteration requiring at least one forward model integration and one adjoint model integration, complex models could be very time consuming. However, as will be seen in the next section, the method appears to converge in only a few iterations.

3. Simulated data study: the one-dimensional advection equation

In order to assess the degree of error reduction that can be accomplished through temporal coupling of analyses, data sets are constructed from analytic solutions of the one-dimensional advection equation. This dynamical constraint is examined in progressively more complicated formulations starting from advection with constant velocity, followed by advection with spatially varying velocity, and finally nonlinear advection governed by Burgers' equation (Burgers, 1939; Platzman, 1964).

The initial state is taken to be

$$q = \sin x, \qquad 0 \leqslant x < 2\pi.$$

The governing constraints are:

$$\text{(a)} \quad \frac{\partial q}{\partial t} + c\frac{\partial q}{\partial x} = 0, \qquad c = \text{constant}, \tag{13a}$$

$$\text{(b)} \quad \frac{\partial q}{\partial t} + V(x)\frac{\partial q}{\partial x} = 0,$$

$$V(x) = \frac{6}{(2\pi)^2}\cdot x(2\pi - x), \tag{13b}$$

$$\text{(c)} \quad \frac{\partial q}{\partial t} + q\frac{\partial q}{\partial x} = 0, \tag{13c}$$

where the coefficient used in the definition of the parabolic profile, $V(x)$, in (13b) is chosen so that the average velocity over the interval is 1. The analytical solutions to these advection equations are:

$$q = \sin(x - ct),$$

$$q = \sin(2\pi x/[x + (2\pi - x)\exp([3/\pi]t)]),$$

$$q = \sin(x - qt).$$

The finite differences form of the linear advection model (13a, b) is given in (14):

$$\frac{q_i(t + \Delta t) - q_i(t)}{\Delta t} + V_i\cdot\left(\frac{\bar{q}_{i+1} - \bar{q}_{i-1}}{2\Delta x}\right) = 0, \tag{14}$$

where

$$\bar{q}_i = \frac{q_i(t + \Delta t) + q_i(t)}{2},$$

and V is a prescribed velocity field. The adjoint equations are given in (15)

$$\frac{\hat{q}_i(t + \Delta t) - \hat{q}_i(t)}{\Delta t} + \frac{(V_{i+1}\bar{\hat{q}}_{i+1} - V_{i-1}\bar{\hat{q}}_{i-1})}{2\Delta x} = 0, \tag{15}$$

where

$$\bar{\hat{q}}_i = \frac{\hat{q}_i(t + \Delta t) + \hat{q}_i(t)}{2}.$$

Note that if the advecting speed is constant as in (13a), the system is unitary.

The nonlinear advection equation (13c), is presented in (16) in finite difference form with the associated adjoint equation given in (17).

$$\frac{q_i(t+\Delta t) - q_i(t)}{\Delta t} + \bar{q}_i \cdot \left(\frac{\bar{q}_{i+1} - \bar{q}_{i-1}}{2\Delta x} \right) = 0, \tag{16}$$

$$\frac{\hat{q}_i(t+\Delta t) - \hat{q}_i(t)}{\Delta t} + \frac{\bar{q}_{i+1}\,\bar{\hat{q}}_{i+1} - \bar{q}_{i-1}\,\bar{\hat{q}}_{i-1}}{2\Delta x} - \bar{\hat{q}}_i \frac{\bar{q}_{i+1} - \bar{q}_{i-1}}{2\Delta x} = 0. \tag{17}$$

Both (16) and (17) are solved iteratively with periodic boundary conditions. Note that in all cases, the finite differencing of the adjoint equations are derived such that they are consistent with the finite differencing of the constraint. Derivation of the adjoint equations in analytic form followed by finite differencing will generally lead to inconsistent equations and an incorrect evaluation of Grad J preventing convergence to the minimum.

The functional for all experiments in this paper is given by (1). The weighting matrix $W(t_p)$ is the identity matrix for the simulated data study. However, for the real data case study, the weighting matrix is defined from the objective analysis routine, thus incorporating information on the expected spatial distribution of analysis quality.

Temporally correlated error has been treated by Thompson (1969) and Lewis (1982) and these studies have shown that the correlated error cannot be reduced to the same degree as random error. For perfectly correlated error in time (e.g., identical phase errors in the case of linear advection), no reduction in error is possible. The reduction in error after adjustment will be bounded by zero and the maximum limit that occurs in the case of an uncorrelated error in time. In this exercise, the analysis will be generated by adding a normally distributed uncorrelated error in space and time to the analytic solutions. The amplitude of the error is 10% of the amplitude of the basic solution.

In the case of linear advection with constant velocity the equation is unitary, thus only one iteration is necessary. An iteration is defined as a forward integration using the constraint followed by a backward integration to the initial time using the adjoint equation. As stated in the previous chapter, for a unitary system the solution to the minimization problem is the average of the forecast and hindcasts from each of the analysis times. For two time levels, the adjusted analysis at the second time level is the average of a forecast from the initial time and the analysis, while the adjusted analysis at the initial time is the average of the hindcast from the second time and the initial analysis. Therefore, the error in the adjusted analyses is the average of the forecast error and the analysis error. Note that the forecast error is composed of two components, the error resulting from the initial conditions and the modelling errors. For these simple test cases, the modelling error results only from the finite differencing. Therefore, as the time step in the model decreases, the model error decreases and the forecast error approaches the error resulting from errors in the initial condition. For the linear problem, this initial error is merely advected from one location to another location. Thus, when this error is averaged with the random error at the analysis time, the result should approach a 50% reduction in mean square error as the time step becomes small.

The reduction in mean square error for an ensemble of 300 experiments are displayed in Fig. 1 for various numbers of time steps used to integrate between the analyses. The reduction is plotted as a function of the time separation constant (= (propagation speed) × (time between analyses)/spatial increment) since the model truncation error is a function of this constant, the number of timesteps between analyses and the wavelength. Note that the time separation constant is equivalent to the Courant number in Lewis (1982) when one timestep is used between analyses. As expected, the reduction in mean square error approaches the theoretical limit of 50% as the number of time steps

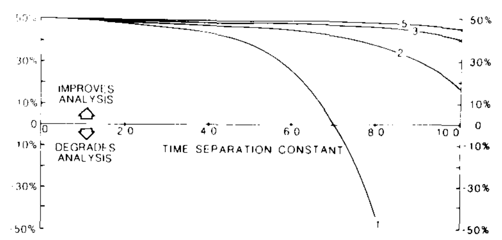


Fig. 1. Reduction in the mean square error as a function of the time separation constant (= propagation speed × time between analyses/spatial increment) for two time-level adjustment. The curve labeled "1" used one time step in the integration process whereas the curve labeled "2" used two time steps to integrate from one analysis to the other, etc.

increases. The curve labeled "1" is equivalent to the adjustment by Thompson's (1969) method which makes the forecast from one analysis time to the other in a single step. The discrepancy between this curve and the corresponding curve in Lewis (1982, $16\Delta s$ curve in Fig. 2) which used Thompson's method, resulted from Lewis' use of a finite difference form of the Euler-Lagrange equation that was inconsistent with the differencing used in the constraint.

For the case of linear advection with spatially varying velocity, the system is no longer unitary so several iterations were generally necessary to reach the minimum. The associated reduction in mean square error is displayed by the solid curves in Fig. 2. When compared with Fig. 1 it is noted that the reduction is generally less than the constant velocity case, although the profiles are similar. The truncation error in the forecast with variable speed is greater than the constant velocity case, yet the approach to the 50% limit is not much different. Since the variable velocity case is not unitary, it is interesting to compare these optimal results with those obtained by the averaging process of Miyakoda-Talagrand. The results associated with averaging are displayed by the dashed line in Fig. 2 and can be seen to give good results but definitely inferior to the optimal results.

For the nonlinear advection equation, a time step of $\Delta t = 0.1$ is chosen to integrate between the analysis times of $t = 0$ and $t = 0.9$, the latter value chosen so that multiple valued analytic solutions can be avoided in the case of Burgers' equation. The spatial increment is $2\pi/8$ and thus $q_{max}\,\Delta t/\Delta x \sim 0.1$, which is sufficiently accurate for the implicit scheme. A quadratic line search is used to

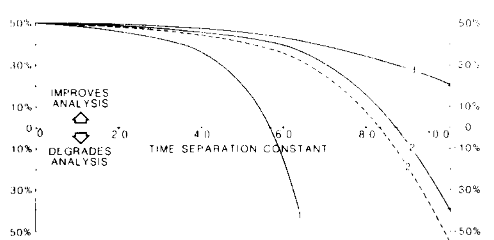


Fig. 2. Same as Fig. 1 except that the results pertain to the advection constraint with variable velocity rather than constant velocity. The dashed curve shows the error reduction resulting from the two time step case when the final fields are found by averaging the analysis with the forecast (hindcast).

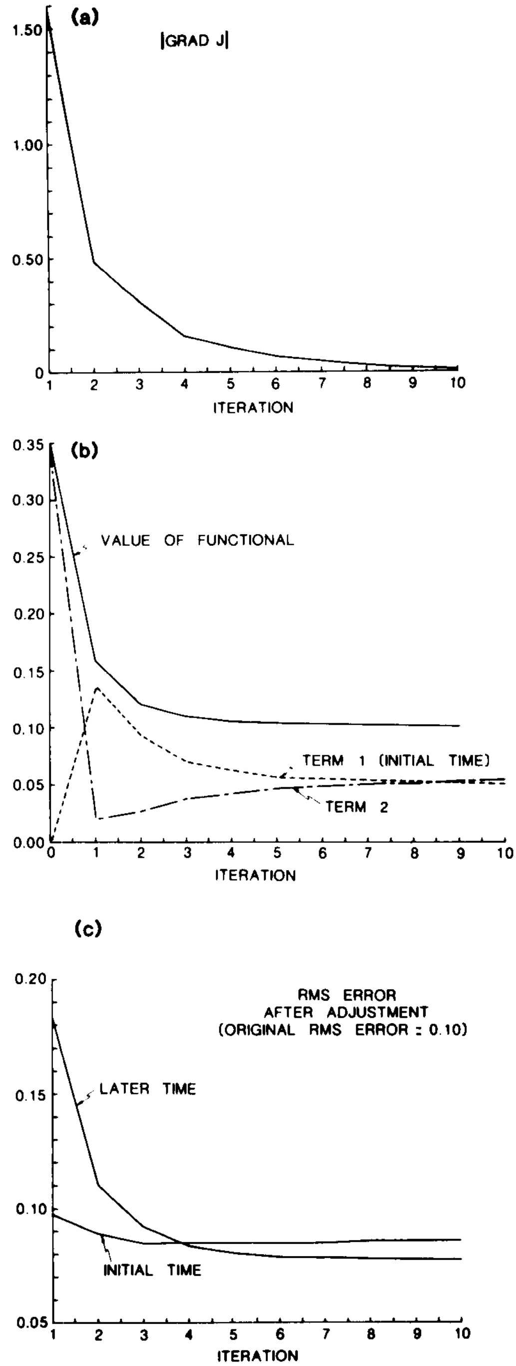


Fig. 3. Adjustment in the nonlinear case as a function of the number of iterations: (a) convergence of the scheme, (b) reduction in the functional value, and (c) error at each time. Terms 1 and 2 in (b) represent the squared deviations at the early and late time, respectively.

find the optimum step size on each iteration. The rate of convergence can be measured in terms of |Grad J| which is shown in Fig. 3a as a function of the number of iterations. The convergence is especially rapid within the first few iterations. The decrease in the value of the functional is shown in Fig. 3b and the reduction in rms error is shown in Fig. 3c. Unlike the linear advection, the functional reduction is greater than 50% and the mean square error reduction is less than 50%. The functional reduction is given as a percentage of the original functional value (iteration 0) calculated using $\tilde{q}(t_1)$ to initialize the forecast model. From Figs. 3b and 3c, the functional reduction is seen to be nearly 70% whereas the average rms error reduction is about 20% (from 0.10 to 0.08) which translates into a 36% reduction in mean square error. Since a change in q alters both the quantity being advected and the advecting velocity, a small change in the initial condition can result in a larger change in the final solution than for the linear case. Because of this, a solution closer to the observations can be found with an associated functional reduction greater than 50%. However, since the adjustments are smaller than those in the linear case, the error reduction is less than 50%.

(a)

(b)

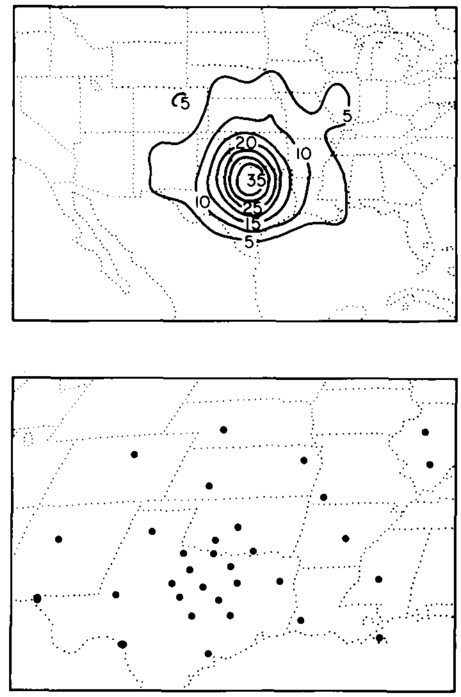


Fig. 4. Rawinsonde observations at each of the three analysis times on March 6, 1982. (a) The relative weighting distribution that is objectively determined from the density of observations and (b) the location of rawinsonde stations.

4. Case study

4.1. Synoptic situation and data distribution

Upper air data collected on March 6–7, 1982 over the central USA are used to test the adjustment process. This day was especially attractive for the evaluation since a special network of rawinsonde observations (RAOB) was operated by the National Aeronautics and Space Administration. These data were collected at three hour intervals as part of the Atmospheric Variability Experiment. Temperature retrievals from the VISSR Atmospheric (VAS) (Smith, 1983) aboard the geostationary satellite GOES-5, were valid 30 minutes before the RAOB data. The test is made on a sequence of three rawinsonde (satellite) data collection periods: 2100 (2030), 0000 (2330), and 0300 (0230) GMT on March 6–7.

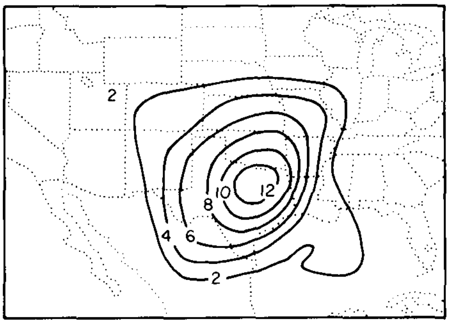
The synoptic pattern during this time period was characterized by a strong baroclinic zone extending from Mexico up through the northeastern USA. A cloud band paralleled the baroclinic zone with imbedded convective activity along the Texas and Lousiana coasts.

The distribution of RAOB and VAS data at 0000 GMT are shown in Figs. 4b and 5b, respectively. These distributions are typical of the other time periods where the number of observations are approximately 35 and 180 for RAOB and VAS, respectively. The average spatial separation between VAS (RAOB) sites is approximately 100 (250) km. Both sets exhibit a concentration of observations in north central Texas. The retrieval free corridor in the VAS data, extending northeastward from the Texas coast, corresponds to the cloud band where the VAS is unable to detect the temperature structure below cloud top level. The cloud band can be seen in the infrared imagery that is used as background for the vertical motion patterns displayed in Fig. 7.

4.2. Interpolation and weighting of observations

The analyses shown in this research were generated by Barnes' method (Barnes, 1973) on a 111 km (1 degree of latitude) grid. This is roughly

(a)



(b)

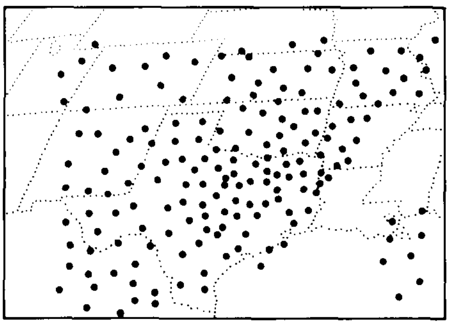


Fig. 5. Same as Fig. 4 except that these observations are from VAS at 2030 GMT. Note that the weighting distribution in (a) has been scaled by a factor of 0.1 relative to the distribution in Fig. 4a.

the average spacing for VAS observations and half the average spacing for rawinsonde observations. In addition to generating the heights at grid points, a by-product of the method is a field depicting the "density of observations" surrounding each grid point. This distribution is an accumulation of the weights accorded each observation used in defining the grid point value. These weighting distributions are shown in Figs. 4a and 5a for the observations of height at 700 mb for RAOB and VAS, respectively. The grid point values of the density distributions are used as the diagonal elements of the weighting matrix in the functional J (eq. 1).

4.3. *Dynamical model and solution of the adjustment equations*

The dynamical constraint used in the case study is the conservation of geostrophic potential vorticity. The two parameter model described by Lewis (1982) was used. The finite difference form of the constraint is given by (15) in Section 3.

Boundary conditions for the prediction have been assumed known from the analyses at the three times with values at intermediate times found by linear interpolation. A timestep, Δt, of 30 min is used for the prediction which produces values of approximately 0.20 and 0.10 for $V_{max}\,\Delta t/\Delta s$ at 250 and 700 mb, respectively. The iteration process for finding the minimum of functional J was considerably improved when the conjugate gradient method (e.g., Polak, 1971, pp. 53–54) was used instead of the steepest descent method.

4.4. *Height adjustments and vertical motion fields*

Our discussion will focus on the height adjustments at 700 mb since the results at 250 mb were very similar. Fig. 6 shows both the input (solid) and output (dashed) heights for RAOB and VAS. The most obvious differences between the VAS and RAOB inputs are:

(i) The VAS analyses of the 297 and 300 dm contours over the Oklahoma-Texas area extend substantially further to the west than the RAOB analyses at the first two times;
(ii) The appearance of some ridging over the Gulf of Mexico on the VAS analyses; and
(iii) the last VAS analysis is disorganized and inconsistent with the first two VAS analyses.

Examination of the adjustment pattern makes it evident that the RAOB heights are altered less, or stated alternatively, they are more consistent with the constraint. RMS adjustments for RAOB and VAS are 3.7 and 7.9 m (13.5 and 22.0 m at 250 mb), respectively. The larger figure for the VAS is contributed chiefly by the last time period where the input analysis is very inconsistent with the other analyses. The adjustments are reasonable considering the observational and analysis errors. A comparison of the ROAB and satellite analyses indicated deviations of 14 and 42 m at 700 and 250 mb for the input, and 12 and 53 m for the output. Thus, the 700 mb fields were statistically closer after adjustment whereas the 250 mb analyses were further apart. A positive bias of about 6 m (VAS higher than RAOB) existed at 250 mb before and after adjustment.

While the adjustment procedure appears to have improved the orientation and position of the Oklahoma-Texas trough in the VAS depiction, the 300 dm contour still extends substantially further to the west than the RAOB analysis indicates. Since this condition existed on two of the input analyses,

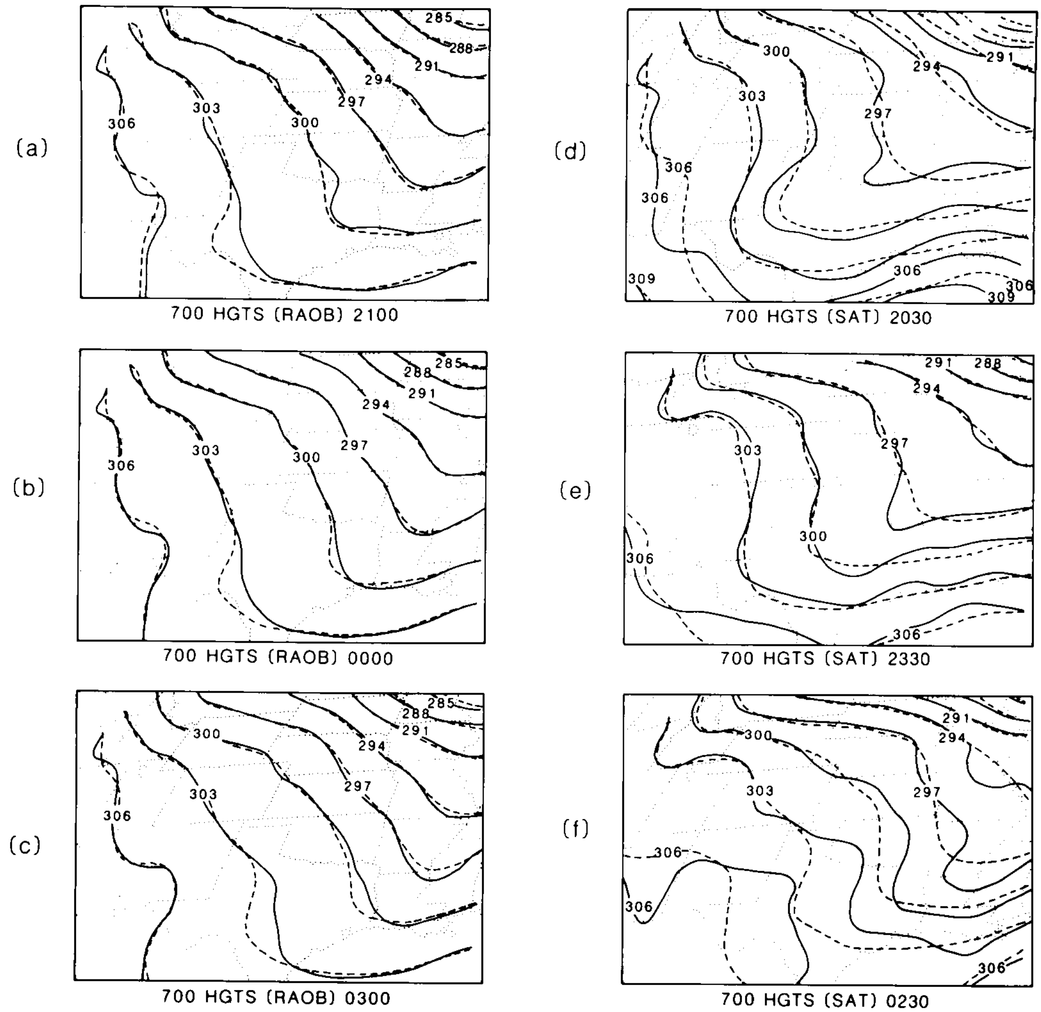
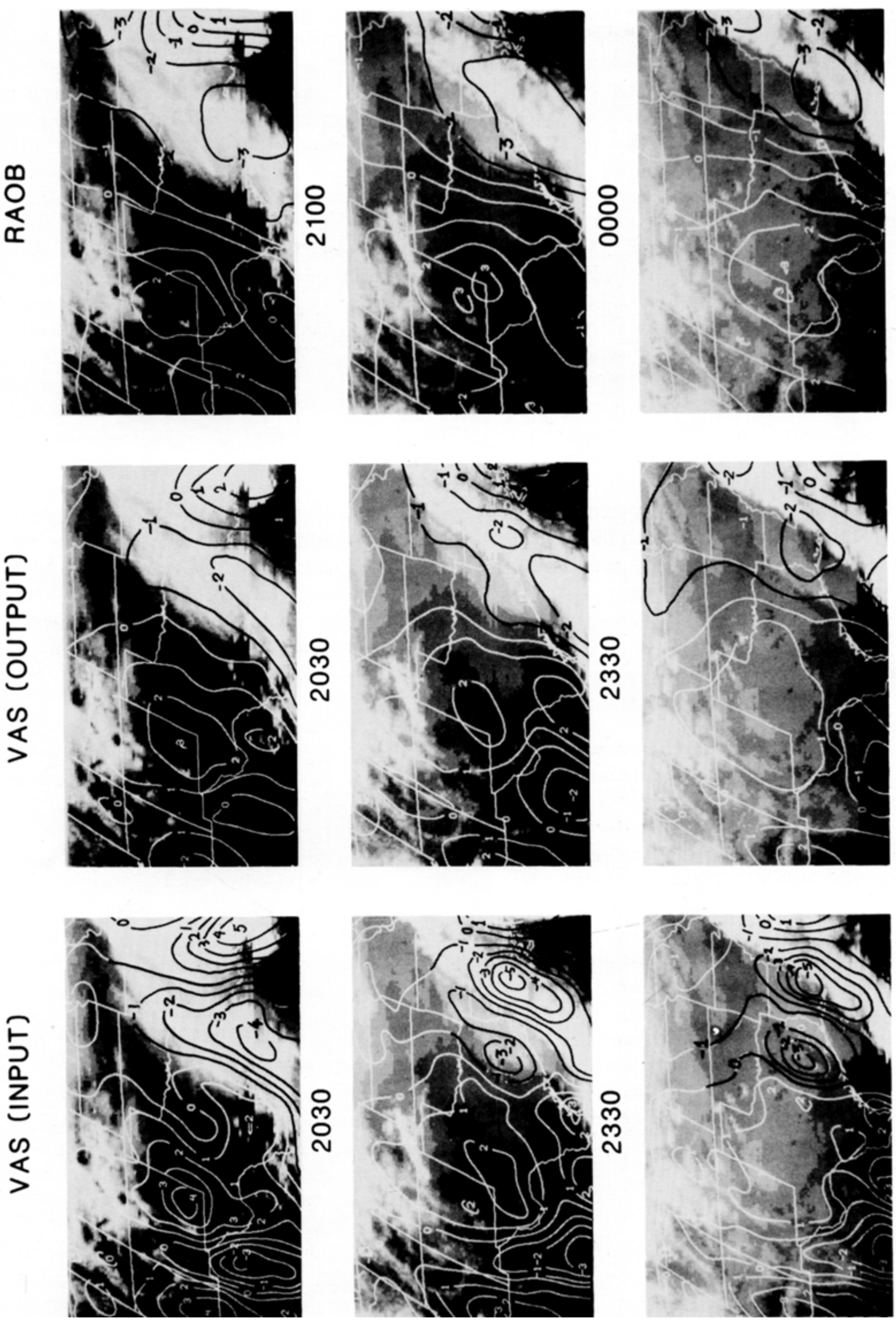


Fig. 6. Height analyses at 700 mb on March 6, 1982. Solid lines are contours of the input analysis whereas dashed lines are output that result from adjustment procedure. (a), (b), and (c) are obtained from rawinsonde observations alone, while (d), (e), and (f) use only VAS observations.

it can be inferred that the error was correlated in time and could not be removed. A perfectly correlated error in time implies that the solution to the governing constraints initialized with the various input analyses will produce the same error pattern at a fixed point in time. The spurious ridging over the Gulf of Mexico is removed by the adjustment process and reveals an advantage of the method, namely that the dynamical time continuity required by the adjustment can remove spurious features present in the input data.

The vertical motion fields at 500 mb were calculated from an omega equation consistent with the two parameter prediction model. These fields are shown in Fig. 7. Only the input is shown for the RAOB since the input and output omegas were very similar. The most obvious feature is the small scale structure evident in the VAS. The strong

Fig. 7. Vertical motion fields at 500 mb in $\mu b\ s^{-1}$, superimposed over infrared imagery that is valid at the corresponding times.


RAOB
2100
0000
0300
VAS (OUTPUT)
2030
2330
0230
VAS (INPUT)
2030
2330
0230

lifting in east Texas at 2330 and 0230 GMT is suspicious; this lifting is well behind the cloud band as depicted by the infrared imagery. Also, the lifting pattern in Mexico is questionable, well to the rear of the advancing baroclinic zone. The adjusted omegas are much better correlated with the ascent-descent regions as inferred from the infrared imagery. They also bear a strong resemblance to the omega pattern derived from the RAOB data. It appears from these results that the VAS data are noisy and the noise is quite successfully eliminated by the adjustment procedure.

In this case, the cloud cover was nearly fixed relative to the translating synoptic system. Accordingly, the Lagrangian advection of information from the data rich to data sparse regions was small relative to the situation where data sparse regions exhibit a more random distribution in space and time. Ideally, the adjustment scheme extrapolates information both forward and backward in time by using the model constraints in much the same way that forecast models are used to generate guess fields for analysis in data sparse regions.

5. Summary and discussion of extensions to more complicated models

Our purpose in this paper has been to develop a methodology for coupling successive analyses subject to forecast model constraints. We have used simple advection constraints in this study, however this method can be applied to more complicated constraints. The problem is cast as a minimization of weighted discrepancies in the analyses at the various times, subject to the model constraints. Traditionally, the solution to this variational problem has been obtained via direct solution of the Euler-Lagrange (E-L) equations to minimize the functional. The governing differential equations are higher order than the constraints and the number of equations in the coupled set increase with the number of time levels, making the solution to the coupled set difficult for even the simplest constraints.

The key to the simplification obtained here comes through the minimization of the functional (sum of weighted squared discrepancies) by finding its gradient with respect to the dependent variable at one of the time levels, e.g., the initial state. The gradient is shown to be given by solutions to a set of adjoint equations which are used to integrate backward in time to the initial state. Unlike the traditional variational approach, the adjoint equations are tractable even with complicated constraints. Non-advective processes can be included without difficulty when their formulation is continuously differentiable with respect to the initial conditions. For example, the inclusion of diffusion in the forecast equations would only result in the addition of a diffusion term in the adjoint equations. Knowledge of the gradient, coupled with standard algorithms such as steepest descent or conjugate gradients, permits minimization of the functional.

The principal deficiency in the method is the limitation imposed by assuming a perfect model. That is, all inconsistency between successive analyses are attributable to errors in the observations and/or objective methods used to interpolate observations to mesh points. Unfortunately, a straightforward procedure to generalize this method and incorporate the model as a weak constraint is not available. Nevertheless, the benefits accrued by amalgamating the analyses using an imperfect model as the constraint can partially offset the errors normally made in a static analysis.

When the advective constraint is linear with constant velocity and the weighting is uniform, the result of the adjustment process is identical to simple averaging of the forecasts. Miyakoda and Talagrand (1971) have demonstrated the utility of this approach with linear dynamics. Thompson's (1969) approach handles the case of varying advection velocities through the solution to the E-L equations associated with minimizing squared discrepancies subject to the advective constraint. Thompson's method results in solutions identical to the algorithm described here when the forecast and hindcast are made in a single timestep. The principal advantages of this method over Thompson's is the ease with which multiple timesteps are accommodated to reduce truncation error and the simplicity of including many analysis times. It should be noted that this method (or probably any dynamical method) cannot eliminate errors correlated in time through the dynamics of the model. It is unlikely that errors correlated to synoptic patterns, such as those that may be found from satellite soundings, can be substantially reduced since the errors will move with the system.

For nonlinear dynamics, the methodology is especially attractive because it avoids the drastic assumptions that are generally invoked to make the E-L equations tractable. Simulations with Burgers' equation indicates a robust algorithm that can significantly reduce mean square error. Multiple solutions to the minimization procedure are possible with the nonlinear constraints. If the initial guess is reasonable, it is very unlikely that the method will select a physically unrealistic state. However, it is impossible to determine if the solution is the global rather than local minimum to the functional. Nevertheless, in every case the solution will fit the constraining model at least as well as the guess.

The practicality of this technique has been demonstrated for the simple models in this paper. The analysis of heights for a sequence of three time periods during the winter of 1982 have shown that the method has potential to reduce the random error in the VAS height fields and make the analysis more compatible with the synoptic features depicted by the rawinsonde data. Vertical motions at 500 mb derived from the adjusted heights also show better correlation with cloud cover than the unadjusted fields.

Perhaps the greatest potential for this assimilation method lies not in the simple applications used in this paper, but rather in its extensions to more complex models. Currently, the method is being tested on a multilevel quasigeostrophic model with good results. The convergence rate for this model is still quite rapid. Results from this work will appear in a future paper.

The assimilation method can be used on a wide variety of forecast models with the restriction that the model must be continuously differentiable with respect to the initial conditions. Thus, models containing frictional dissipation and diabatic effects can be used if the parameterizations are properly formulated. Unfortunately, many parameterization schemes are written with inequalities dependent on forecast variables and accordingly are not differentiable.

The extension to a primitive equation model appears to be relatively straightforward assuming the parameterization schemes are differentiable. The major difficulty in using a primitive equation model is the large increase in computing time. The method relies on an iterative method to find the minimum with at least one integration of the forecast model and the adjoint model required per iteration. Also, at least one additional integration of the forecast model is required to determine the stepsize. In the multilevel quasigeostrophic model and the simple models used in this paper, it has been possible to determine the stepsize with a single model integration. With a primitive equation model, the functional may not change as smoothly in the search direction and several integrations of the forecast model may be necessary to determine the stepsize. However, even if some increase in the computing time occurs due to a decrease in the convergence rate or an increase in the number of integrations necessary to determine the stepsize, the method should at least be practical for research purposes.

Another potential difficulty in applying the method to a primitive equation model may be amplification of nonphysical modes in order to better fit the data. However, these nonphysical modes can be suppressed with little additional computation by introducing terms in the functional penalizing large amplitudes of the undesirable modes. This change to the functional would not entail any additional integrations of the adjoint or forecast model.

6. Acknowledgments

It was during the 14th Stanstead Seminar at Bishop's University in Quebec that Olivier Talagrand first mentioned the ideas of the "adjoint" method to us. We are grateful for the continued dialogue with both Olivier Talagrand and Francois Le Dimet. We appreciate the constructive comments on the manuscript made by Christopher Hayden and Thomas Koehler and the three referees. Jim Purser's suggestions on the simulated data study were helpful and Gary Wade assisted us in producing displays of data.

REFERENCES

Avriel, M. 1976. *Nonlinear programming: analysis and methods*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 512 pp.

Barnes, S. L. 1973. Mesoscale objective analysis using weighted time-series observations. NOAA Technical Memorandum ERL NSSL-62, National Severe Storms

Laboratory, 1313 Halley Circle, Norman Oklahoma, 73069, 60 pp., [NTIS COM-73-10781].

Bengtsson, L. 1975. Four-dimensional assimilation of meteorological observations. Geneva, World Meteorological Organization, International Council of Scientific Unions. Joint Organizing Committee. 75 pp (GARP Publications Series no. 15).

Bengtsson, L., Ghil, M. and Källén, E. (eds.) 1981. *Dynamic meteorology: data assimilation methods.* Springer-Verlag, Berlin, 330 pp.

Bloom, S. C. 1983. The use of dynamical constraints in the analysis of mesoscale rawinsonde data. *Tellus 35A*, 363–378.

Burgers, J. M. 1939. Mathematical examples illustrating relations occurring in the theory of turbulent fluid motion. Verhandelingen der Koninklijke Nederlandsche Akademie van Wetenschappen (Erste Sectie), 17, 53 pp.

Courant, R. and Hilbert, D. 1962. *Methods of mathematical physics*, vol. II. Interscience Publ., 830 pp.

Hovermale, J. 1983. Summary of a workshop on atmospheric profiling. *Bull. Amer. Meteorol. Soc. 64*, 1062–1066.

Kreyszig, E. 1978. *Introductory functional analysis with applications.* John Wiley and Sons, 688 pp (see pages 195–206).

LeDimet, F. X. 1982. A general formalism of variational analysis. CIMMS report 22, Cooperative Institute for Mesoscale Meteorological Studies, 815 Jenkins Street, Norman, OK, 73019, 34 pp.

Lewis, J. M. 1982. Adaptation of P. D. Thompson's scheme to the constraint of potential vorticity conservation. *Mon. Wea. Rev. 110*, 1618–1634.

Lewis, J. M. and Panetta, L. 1983. The extension of P. D. Thompson's scheme to multiple time levels. *J. Clim. Appl. Meteorol. 22*, 1649–1653.

Lions, J. L. 1971. *Optimal control of systems governed by partial differential equations.* Translated by S. K. Mitter. Springer-Verlag, Berlin-Heidelberg.

Miyakoda, K. and Talagrand, O. 1971. The assimilation of past data in dynamical analysis. I. *Tellus 23*, 310–317.

Platzman, G. W. 1964. An exact integral of complete spectral equations for unsteady one-dimensional flow. *Tellus 16*, 422–431.

Polak, E. 1971. *Computational methods in optimization, a unified approach.* Academic Press, New York, 329 pp.

Sasaki, Y. 1958. An objective analysis based on the variational method. *J. Meteorol. Soc. Japan II, 36*, 77–88.

Sasaki, Y. 1970. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev. 98*, 875–883.

Smith, W. L. 1983. The retrieval of atmospheric profiles from VAS geostationary radiance observations. *J. Atmos. Sci. 40*, 2025–2035.

Thompson, P. D. 1969. Reduction of analysis error through constraints of dynamical consistency. *J. Appl. Meteorol. 8*, 738–742