Q. J. R. Meteorol. Soc. (1992), **118**, pp. 1163–1186 551.509.313

An optimal nudging data assimilation scheme using parameter estimation

By X. ZOU
Supercomputer Computations Research Institute, Florida State University, Tallahassee, FL 32306.
I. M. NAVON
Dept. of Mathematics and Supercomputer Computations Research Institute, Florida State University, Tallahassee, FL 32306
and
F. X. LEDIMET
Laboratoire de Modelisation et Calcul, Université Joseph Fourier, B.P. 53*X*, 38041 *Grenoble Cedex, France*

(Received 2 October 1991; revised 3 June 1992)

Summary

A new optimal nudging dynamical relaxation technique is tested in the framework of 4-dimensional variational data assimilation, applied to an adiabatic T40 version of the National Meteorological Center (NMC) spectral model with 18 vertical layers. Several experiments are performed using the NMC operationally analysed data. The variational data assimilation algorithm is also employed in a parameter-estimation mode to determine the vector of optimal nudging coefficients. Results of data-assimilation experiments involving estimated nudging, optimal nudging and variational data assimilation are compared. Issues are addressed related to the dependence of the assimilation on the length of the assimilation period as well as to the ability of retrieving high-quality model initial conditions.

The study outlines the ability to obtain optimal nudging coefficients, which can vary in space, in the framework of a parameter-estimation approach using variational data assimilation. Based on our preliminary results the optimal nudging seems to be a most promising data-assimilation scheme.

1. Introduction

In recent years rapid advances in remote-sensing technology have led to the development of new observing systems capable of nearly continuous monitoring of the atmosphere. At the same time mainframe supercomputers have become more powerful, with improvements in memory size, the introduction of multi-processor capabilities as well as enhanced processor speed. These advances have created new opportunities to improve numerical weather prediction (NWP) capabilities significantly for both operational and research purposes. Increasing the accuracy of numerical forecasts critically depends upon initial conditions. Four-dimensional (4-D) data-assimilation schemes provide practical means for advanced model initialization that can successfully incorporate both synoptic and asynoptic data originating from a wide variety of observing systems, while accounting for the dynamic evolution of the atmosphere's state.

Among the 4-D data-assimilation approaches, nudging data assimilation (NDA) and variational data assimilation (VDA) methods are often considered to be two of the most promising techniques capable of utilizing the ever growing number of asynoptic observations. Briefly, the VDA method seeks to find an optimal initial condition which minimizes the differences between the model solution and observations in a certain assimilation time interval (see, for instance, LeDimet and Talagrand 1986; Navon *et al.* 1992). It uses an optimal control approach based on adjoint model integration to obtain the gradient of the cost function, with respect to the control variables for the minimization procedure, efficiently. This approach is cheaper than the explicit finite-difference approximation for the gradient calculation for large-dimensional models. Nevertheless, the cost of the VDA method for real distributed data is still prohibitive for operational applications. Additional research is required to improve the rate of convergence of the minimization part of the algorithm by proper scaling and weighting. Other issues related to VDA concern determining the optimal length of the assimilation window, the treatment of on–off physical processes such as large-scale precipitation and deep cumulus convection, and the control of high-frequency gravity-wave oscillations, to mention but a few. Moreover, the adjoint VDA method may also be used to perform parameter estimation and sensitivity analysis.

The NDA method relaxes the model state towards the observations during the assimilation period by adding a non-physical diffusive-type term to the model equations. The nudging terms are defined as the difference between the observation and model solution multiplied by a nudging coefficient. The size of this coefficient is chosen by numerical experimentation so as to keep the nudging terms small in comparison with the dominant forcing terms in the governing equations, in order to avoid the rebounding effect that slows down the assimilation process, yet large enough to impact the simulation. NDA techniques have been used successfully on the global scale by Lyne *et al.* (1982), Krishnamurti *et al.* (1991) and Lorenc *et al.* (1991), and in a wide variety of research applications on mesoscale models (Hoke and Anthes 1976; Davis and Turner 1977; Ramamurthy and Carr 1987, 1988; Wang and Warner 1988; Stauffer and Seaman 1990, to cite but a few). The NDA can be thought of as an iterative approximation to the Kalman filter (KF) (Lorenc 1986; Lorenc *et al.* 1991). The NDA method is a flexible assimilation technique which is computationally much more economical than the VDA method. However, results from NDA experiments are quite sensitive to the *ad hoc* specification of the nudging relaxation coefficient, and it is not at all clear how to choose a nudging coefficient so as to obtain an optimal solution.

In this study we aim to combine the aforementioned data-assimilation schemes in the most efficient way. A parameter-estimation approach is used in the framework of the VDA algorithm to determine optimally the coefficients for the NDA scheme. The goal is to find optimal nudging coefficients which best assimilate the given observations. It is well known that the best nudging coefficients are those related to a KF in a linear system of equations. Application of the KF technique to the assimilation of meteorological (or oceanographical) observations has been studied by several authors (e.g. Ghil and Malanotte-Rizzoli 1991), and present operational NDA procedures can be described as degraded forms of the KF. However the KF is very costly to implement in practice.

We can obtain optimal nudging coefficients in a much more economical way by using the VDA method; this allows the adjusting of variables other than the initial conditions for either linear or nonlinear systems and employs an adjoint model for gradient calculation. Parameters in the NWP model can easily be incorporated in the adjoint VDA procedure and serve as additional control variables. The variational algorithm is formulated here using the nudging coefficient as the control parameter. The nudging coefficient is estimated so that the model solution is as close as possible to the observations. Operationally analysed data from the National Meteorological Center (NMC) will be assimilated into the model. The plan of this paper is as follows. In section 2 the dynamical NDA scheme and the parameter estimation are briefly described for a general model; relations among NDA, optimal NDA and VDA are also discussed. In section 3 computational details and numerical results aimed at assessing the data-assimilation performance of estimated NDA, optimal NDA and VDA procedures are presented, using an adiabatic T40 version of the NMC spectral model with 18 vertical layers. Section 4 contains conclusions. In appendices A and B we cover the detailed derivation of the gradient calculation by using the adjoint model.

2. OPTIMAL NUDGING SPECIFICATION

(a) *Dynamical nudging*

We assume that the model equations have been discretized in space by a finite difference, finite element, or spectral discretization method. The time continuous model satisfies dynamical equations of the form

$$\frac{\partial X}{\partial t} = F(X) \tag{2.1}$$

$$X(t_0) = U \tag{2.2}$$

where X represents the discretized state variable of the model atmosphere, t is time and U represents the initial condition for the model. Say, for instance, $X^o(t)$ is a given observation, the objective of VDA is to find model initial conditions that minimize a cost function defined by

$$\mathcal{J}(U) = \int_{t_0}^{t_R} \langle W(X - X^o), X - X^o \rangle \mathrm{d}t \tag{2.3}$$

where $\langle\,\rangle$ is an inner product of two vectors and W is a diagonal weighting matrix. Note that $\mathcal{J}$ is only a function of the initial state U because X is uniquely defined by the model equations (2.1)–(2.2).

An implicit assumption made in VDA is that the model exactly represents the state of the atmosphere. However, this assumption is not true.

The NDA technique introduced by Anthes (1974) consists in achieving a compromise between the model and the observations by considering the state of the atmosphere to be defined by

$$\frac{\partial X}{\partial t} = F(X) + G(X^o - X) \tag{2.4}$$

where G is a diagonal matrix with $G_{\ln p_s}$, G_T, G_D, G_ζ and G_q as its diagonal submatrices representing adjustable nudging coefficients for the surface pressure, temperature, divergence, vorticity and moisture fields respectively.

Together with the initial conditions

$$X(t_0) = U \tag{2.5}$$

the system (2.4) has a unique solution $X(U, G)$.

The main difficulty in the NDA scheme resides in the estimation of the nudging coefficient G (Stauffer and Seaman 1990). If G is too large, the fictitious diffusion term will completely dominate the time tendency and will have an effect similar to replacing the model data by the observations at each time step. Should a particular observation have a large error that prevents obtaining a dynamic balance, an exact fit to the observation is not required since it may lead to a false amplification of observational errors. On the other hand, if G is too small, the observation will have little effect on the solution. In general, G decreases with increasing observation error, increasing horizontal and vertical distance separation, and increasing time separation. In the experiment of Anthes (1974) a nudging coefficient of 10^{-3} was used for all the fields for a hurricane model and was applied on all the domain of integration. In the experiment of Krishnamurti *et al.* (1991) the relaxation coefficients for the estimated NDA experiment were kept invariant both in space and time, and their values were simply determined by numerical experience. The following values were used:

$$\begin{aligned} G_\zeta, G_{\ln p_s} &= 1 \times 10^{-4}\mathrm{s}^{-1} \\ G_D &= 0.5 \times 10^{-4}\mathrm{s}^{-1} \end{aligned} \tag{2.6}$$

i.e. the vorticity, divergence and the pressure tendency fields are subjected to the Newtonian relaxation. The implicit dynamic constraints of the model then spread the updated information to the other variables (temperature and moisture) resulting eventually in a set of balanced conditions at the end of the nudging period.

We will now present a new parameter-estimation approach designed to obtain optimal nudging coefficients G^*. They are optimal in the sense that the difference between the model solution and the observations will be small. The value in (2.6) will be used both as the nudging coefficient for estimated NDA experiment (usual NDA using *ad hoc* nudging parameters) and as the initial guess in a variational parameter-estimation approach aimed at obtaining optimal nudging coefficients.

(*b*) *Parameter estimation*

By fitting the model solutions to the observational data, the unknown parameters of the model can be deduced simultaneously by minimizing a cost function that measures the misfit between the model results and observations, in which the model parameters are the control variables. For example, the barotropic gravity-wave speed in a two-dimension reduced-gravity, linear-transport model for the equatorial Pacific Ocean was used as a parameter control variable (Smedstad and O'Brien 1991). In the work of Panchang and O'Brien (1988) the friction coefficient for a one-dimension tidal-flow model was the parameter to be estimated from the observations.

The application of the variational approach to determine model parameters is conceptually similar to that of determining the initial conditions. In the following we will present a brief illustration of the method.

For the parameter estimation of the nudging coefficients, the cost function $\mathcal{J}$ can be defined as

$$\mathcal{J}(G) = \int_{t_0}^{t_R} \langle W(X - X^{\mathrm{o}}), X - X^{\mathrm{o}} \rangle \,\mathrm{d}t + \int_{t_0}^{t_R} \langle K(G - \hat{G}), G - \hat{G} \rangle \,\mathrm{d}t \tag{2.7}$$

where $\hat{G}$ denotes the estimated nudging coefficients and the Ks are specified weighting matrices. Here observations are assumed everywhere on model grid points for simplicity. For a more realistic case, a transform operator H from X to X^{o} should be included in (2.7) and hence G will no longer be a diagonal matrix with constant terms in each block. The second term plays a double role. On one hand it ensures that the new value of the nudging parameters is not too far away from the estimated quantity. On the other hand it enhances the convexity of the cost function since this term contributes a positive term, K, to the Hessian matrix of $\mathcal{J}$ (see also Smedstad and O'Brien 1991).

An optimal NDA procedure can be defined by the optimal nudging coefficients G^* such that

$$\mathcal{J}(G^*) \leqslant \mathcal{J}(G), \qquad \forall G. \tag{2.8}$$

The problem of extracting the dynamical state from observations is now identified as the mathematical problem of finding initial conditions or external forcing parameters that minimize the cost function.

Owing to the dynamical coupling of the state variables to the forcing parameters, the dynamics can be enforced through the use of a Lagrange function constructed by appending the model equations to the cost function as constraints, so avoiding the repeated application of the chain rule when differentiating the cost function. The Lagrange function is defined by

$$L(X, G, P) = \mathcal{J} + \left\langle P, \frac{\partial X}{\partial t} - F(X) - G(X^{o} - X) \right\rangle \tag{2.9}$$

where P is a vector of Lagrange multipliers. The Lagrange multipliers are not specified but computed in determining the best fit.

The gradient of the Lagrange function must be zero at the minimum point. This results in the following first-order conditions:

$$\frac{\partial L}{\partial X} = 0 \sim \text{adjoint model forced by } 2W(X - X^{o}) \tag{2.10}$$

$$\frac{\partial L}{\partial P} = 0 \sim \text{direct model (2.4)} \tag{2.11}$$

$$\frac{\partial L}{\partial G} = 0 \sim \int \left\langle P, -\frac{\partial(F(X) + G(X^{o} - X))}{\partial G} \right\rangle \mathrm{d}t + 2K(G - \hat{G}) = 0. \tag{2.12}$$

The solution of Eqs. (2.10)–(2.12) is called the stationary point of L. Even if the dynamical evolution operator is nonlinear, the equations ($\partial L/\partial X = 0$) will be the same as those derived by constructing the adjoint of the linear-tangent operator; the linearization is automatic owing to the Lagrange function L being linear in terms of the Lagrange multipliers P.

An important relation between the gradient of the cost function (2.7) with respect to parameters G and the partial derivative of the Lagrange function with respect to the parameters is

$$\nabla_G \mathcal{J}(G) = \left. \frac{\partial L}{\partial G} \right|_{\text{at stationary point}} \tag{2.13}$$

i.e. the gradient of the cost function with respect to the parameters is equal to the left-hand side of (2.12) which can be obtained in a procedure where the model state P is calculated by integrating the direct model forward and then integrating the adjoint model backwards in time with the Lagrange multipliers as adjoint variables.

Using this procedure we can derive the following expressions of the adjoint-model equation and gradient formulation (see appendix A)

$$\left. \begin{aligned} &\frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X}\right]^{\mathrm{T}} P - G^{\mathrm{T}} P = W(X - X^{o}) \\ &P(t_R) = 0 \end{aligned} \right\} \tag{2.14}$$

and

$$\nabla_G \mathcal{J} = -\int_{t_0}^{t_R} \langle (X^{o} - X), P \rangle \, \mathrm{d}t + 2K(G - \hat{G}). \tag{2.15}$$

We see that the adjoint equation of a model with a nudging term added is the same as that without a nudging term except that an additional term $-G^{\mathrm{T}}P$ was added to the left-hand side of the adjoint equation.

Having obtained the value of cost function $\mathcal{J}$ by integrating the model (2.4) forward, and the value of the gradient $\nabla_G \mathcal{J}$ by integrating the adjoint equation (2.14) backwards in time, any large-scale unconstrained minimization method may be employed to minimize the cost function and finally obtain an optimal parameter estimation. If both the initial condition U and the parameter G are controlled, the gradient of the cost function for performing the minimization would be

$$\nabla \mathcal{J} = (\nabla_U \mathcal{J}, \nabla_G \mathcal{J})^{\mathrm{T}} \tag{2.16}$$

where

$$\nabla_U \mathcal{J} = -P(t_0) \tag{2.17}$$

see Navon *et al.* (1992).

(c) *Kalman filter and optimal nudging*

In this section we give a brief description of the KF (see also Ghil and Malanotte-Rizzoli 1991) and the connection between the KF, nudging, optimal nudging and VDA.

Starting from the forecast model (2.1) which is advanced in discrete time steps Δt, $X_n = X(t_n)$, $t_n = n\Delta t$, i.e.

$$X_n^{\mathrm{f}} = \Psi_{n-1} X_{n-1}^{\mathrm{a}} \tag{2.18a}$$

where the superscript f stands for the forecast and the superscript a for the analysis, Ψ is the discretized form of the system matrix, describing the model dynamics.

A linear unbiased data assimilation scheme for the analysed model state can be written as

$$X_n^{\mathrm{a}} = X_n^{\mathrm{f}} + G_n(X_n^{\mathrm{o}} - H_n X_n^{\mathrm{f}}). \tag{2.18b}$$

where H represents the fact that only certain variables or combinations thereof are observed at a set of points smaller than the total number of model grid points. The weight matrix G_n is often called the gain matrix. The KF uses an optimal G_n to carry out such a linear unbiased data assimilation. The optimality is defined in the context of the following assumptions.

First assume the true evolution of the atmosphere, X_n^{t}, is governed by

$$X_n^{\mathrm{t}} = \Psi_{n-1} X_{n-1}^{\mathrm{t}} + b_{n-1}^{\mathrm{t}} \tag{2.19a}$$

where b_n^{t} is a Gaussian white-noise sequence, i.e.

$$Eb_n^{\mathrm{t}} = 0, \qquad Eb_n^{\mathrm{t}}(b_l^{\mathrm{t}})^{\mathrm{T}} = Q_n \delta_{nl} \tag{2.19b}$$

with E being the expectation operator and δ_{nl} being the Kronecker delta function. The second assumption used in optimizing the weight matrix G_n concerns the error model for the observations

$$X_n^{\mathrm{o}} = H_n X_n^{\mathrm{t}} + b_n^{\mathrm{o}} \tag{2.20a}$$

where b_n^{o}, the observational noise, satisfies

$$Eb_n^{\mathrm{o}} = 0 \qquad Eb_n^{\mathrm{o}}(b_l^{\mathrm{o}})^{\mathrm{T}} = R_n \delta_{nl}. \tag{2.20b}$$

The third assumption is that system noise and observational noise are uncorrelated with each other:

$$Eb_n^{\mathrm{f}}(b_n^{\mathrm{o}})^{\mathrm{T}} = 0. \tag{2.21}$$

Using (2.18)–(2.21), one can derive the time evolution of the error covariance matrix

$$W_n^{\mathrm{a}} \equiv E(X_n^{\mathrm{a}} - X_k^{\mathrm{t}})^{\mathrm{T}} (X_n^{\mathrm{a}} - X_k^{\mathrm{t}}) = (I - G_n H_n) W_n^{\mathrm{f}} (I - G_n H_n)^{\mathrm{T}} + G_n R_n G_n^{\mathrm{T}} \tag{2.22}$$

where

$$W_n^{\mathrm{f}} \equiv E(X_n^{\mathrm{f}} - X_k^{\mathrm{t}})^{\mathrm{T}}(X_n^{\mathrm{f}} - X_k^{\mathrm{t}}) = \Psi_{k-1} W_{n-1}^{\mathrm{a}} \Psi_{n-1}^{\mathrm{T}} + Q_{k-1}. \tag{2.23}$$

Hence, by advancing W_n^{f}, W_n^{a} along with X_n^{f}, X_n^{a}, one can know how well the true state X_n^{t} is estimated by X_n^{a} for any weight matrix G_n. This in turn permits one to determine the optimal G_n by minimizing

$$\mathcal{J}_{\mathrm{KF}}(G_n) \equiv \mathrm{tr} W_n^{\mathrm{a}} \tag{2.24}$$

where tr denotes the trace of the matrix.

The optimal weight matrix G_n at the nth time step is obtained by using (2.22) for the matrix W_n^{a} and setting the derivative of $\mathcal{J}$ with respect to each element of G_n equal to zero. The minimum is attained for

$$G_n^* = W_n^{\mathrm{f}} H_n^{\mathrm{T}} (H_n W_n^{\mathrm{f}} H_n^{\mathrm{T}} + R_n)^{\mathrm{T}}. \tag{2.25}$$

The above linear unbiased data assimilation scheme (2.18a, b) with the optimal gain matrix G_n^* (2.25) is called the KF (Kalman 1960).

There are two problems which arise in the KF. The first is the computational complexity of advancing in time the error covariance matrices. While Eqs. (2.18a, b) represent $O(M)$ computations per time step, Eqs. (2.22)–(2.23) represent, at face value, $O(M^2)$ computations. Second, the noise covariance matrices Q_n and R_n are assumed to be known in the subsequent derivation of the optimal G_n. This is not so in practice, and finding the actual magnitude of system and observational errors is an important function of the data-assimilation process.

The nudging scheme is carried out by the following procedure (see (3.2)):

$$X_n^{\mathrm{a}} = \Psi_{n-1} X_{n-1}^{\mathrm{a}} + G_n (X_n^{\mathrm{o}} - H_n X_n^{\mathrm{a}}). \tag{2.26}$$

The optimal nudging coefficients G_n, $n = 0, \ldots, R$, are obtained by minimizing a cost function measuring the distance between the analysis and the observations

$$\mathcal{J}_{\mathrm{NDA}}(G) = \sum_{n=0}^{R} (X_n^{\mathrm{a}} - X_n^{\mathrm{o}})^{\mathrm{T}} W (X_n^{\mathrm{a}} - Z_n^{\mathrm{o}}). \tag{2.27}$$

From (2.18) and (2.24) we see that the core of the KF is the optimal merging of observation and forecast information in the sense that the expected mean-square estimation error is minimized at every time step. The optimal NDA, on the other hand, is the optimal merging of observations and analysis in the sense that the total differences between them in a certain window of assimilation is minimized (see (2.26) and (2.27)). The main differences between the KF and the optimal NDA described in this paper are:

(i) observation errors at different times are assumed to be uncorrelated;
(ii) the weight matrix G_n at each time step is determined sequentially in the KF, while the nudging coefficients at every time step in the window of assimilation are obtained simultaneously. However, the two problems of the KF described above disappear in the optimal NDA. The computational cost is reduced by using an adjoint-model integration in the parameter-estimation mode of VDA. Moreover, the optimal NDA does not require any knowledge of the noise covariance matrices.

Therefore, the estimated NDA, the optimal NDA and the KF differ from each other mainly in the choice of the weight matrix G_n. The VDA, on the other hand, takes both the model forecasts and the observations as perfect, i.e. $b_n^{\mathrm{f}} = 0$ and $b_n^{\mathrm{o}} = 0$ when $n \neq 0$. It attempts to obtain an optimal initial condition (U^*) which minimizes the cost function

$$\mathcal{J}_{\mathrm{VDA}}(U) = \sum_{n=0}^{R} (X_n^{\mathrm{f}} - X_n^{\mathrm{o}})^{\mathrm{T}} W(X_n^{\mathrm{f}} - X_n^{\mathrm{o}}). \tag{2.28}$$

The theoretical framework of estimation and control theory provides the foundation of data-assimilation techniques. The estimated NDA and the KF are closer to the estimation theory, the VDA to the optimal control aspect, while optimal NDA is a combination of both (see also Lorenc 1986).

3. Numerical experiments

(a) *Algorithmic implementation*

Our procedure for obtaining the optimal nudging coefficient by parameter estimation is as follows:

(1) Start with an initial guess for the control variables: the initial state $U^{(0)}$ and the nudging parameter $G^{(0)}$, where $U^{(0)}$ is taken to be the NMC operationally analysed data at time t_0 and $G^{(0)}$ the estimated quantity defined in (2.6). Set the iteration number $k = 0$.

(2) Integrate the model (2.4) forward in time and calculate the value of the cost function defined by

$$\mathcal{J}(U, G) = W(U^{(k)} - X^0(t_0))^2 + W(X^{(k)}(t_R) - X^0(t_R))^2 \tag{3.1}$$

where $X^0(t_0)$ is the result of the normal mode initialization (NMI) of $U^{(0)}$, $X^0(t_R)$ is the 6 h or 12 h model integration by the forward nonlinear model (2.1) from the initial condition $X^0(t_0)$, $X^{(k)}(t_R)$ the model solution of equation (2.4) (i.e. the foreward nonlinear model augmented by the nudging terms) from the initial condition $U^{(k)}$, and the weight matrix W is defined as the inverse of the maximum square of the difference between two time-level observations. Store in memory the misfit between the model solution and the observations $W(X^{(k)} - X^{\mathrm{o}})$ at two times t_0 and t_R when observations are available.

(3) Integrate the adjoint equation (2.14) backwards in time and calculate the gradient of the cost function with respect to the control variables using (2.15)–(2.17).

(4) With both the cost function and its gradient being available, apply a limited-memory quasi-Newton (L-BFGS) unconstrained minimization method (Nocedal 1980; Liu and Nocedal 1988) to obtain a new value for U and G: $U^{(k+1)}$ and $G^{(k+1)}$.

(5) Check if the convergence criterion

$$\|\nabla\mathcal{J}\|/\|\nabla\mathcal{J}_0\| \leqslant \varepsilon \tag{3.2}$$

is satisfied, where ε is a predetermined adequately small number. Set $k + 1 = k$ and return to step (2) if the convergence criteria in (3.2) is not satisfied.

In each case, no prior information was presumed ($K = 0$).

For the nudging terms added to the model, an implicit time-differencing scheme was used in order to ensure the computational stability for any value assumed by the nudging coefficients. The time integrations are carried out in two steps, the tendency without the nudging term, $X^*(t + \Delta t)$ being calculated first. The Newtonian term is then expressed in finite difference form using the relation

$$\frac{X(t + \Delta t) - X^*(t + \Delta t)}{2\Delta t} = G(X^{\mathrm{o}}(t + \Delta t) - X(t + \Delta t)). \tag{3.3}$$

The aims of this study are: to determine the 'best' or 'optimal' nudging coefficients, and to gain insight into the differences in performance of the variational and NDA techniques applied to an adiabatic version of the NMC spectral model (Navon *et al.* 1992). To accomplish these dual objectives, a sequence of experiments were performed, all of them using NMC operationally analysed data or model-generated data.

(*b*) *Optimal estimation of the nudging coefficient*

It is important to obtain the correct gradient of the cost function with respect to the nudging coefficients in the parameter-estimation procedure before we carry out the minimization of the cost function. One way to check that a correct gradient has been obtained is to calculate the value of the function

$$\psi(\alpha) = \frac{\mathcal{J}(\hat{G} + \alpha\nabla_G \mathcal{J}) - \mathcal{J}(\hat{G})}{\alpha(\nabla_G \mathcal{J}(\hat{G}))^{\mathrm{T}}\, \nabla_G \mathcal{J}(\hat{G})} \approx 1 + \text{higher order terms.} \tag{3.4}$$

In Table 1 values of the function $\psi(\alpha)$ are given as a function of α. It is noted that for α between 10^{-10} to 10^{-19}, Eq. (3.4) is verified. A correct gradient has therefore been found.

TABLE 1. VERIFICATION OF THE GRADIENT OF THE COST FUNCTION WITH RESPECT TO THE NUDGING COEFFICIENT, USING THE ADJOINT MODEL

α	$\psi(\alpha)$
10^{-9}	1.520186555
10^{-10}	1.037451906
10^{-11}	1.003640547
10^{-12}	1.000363038
10^{-13}	1.000036316
10^{-14}	1.000003667
10^{-15}	1.000000842
10^{-16}	0.9999731466
10^{-17}	1.000061370
10^{-18}	1.000831928
10^{-19}	0.9927913232
10^{-20}	0.8040604642

Having obtained the correct gradient we minimize the cost function using the limited-memory quasi-Newton unconstrained minimization method (Liu and Nocedal 1989) in order to find the optimal nudging coefficient in the parameter-estimation procedure. In the NDA experiment one chooses the predicted variable fields to be assimilated. In the first NDA experiment, divergence, vorticity and surface pressure fields are adjusted (see Krishnamurti *et al.* 1991). In the second NDA experiment, divergence, vorticity and temperature fields are adjusted (Kuo and Guo 1989).

The first experiment involved nudging the surface pressure, divergence and vorticity in a 6-hour assimilation window. In Fig. 1, values of the cost function and the initial and final values (after 30 iterations) of the gradient norm are shown. The norm of the gradient decreases four orders of magnitude during the 30 iterations. We note a rapid decrease in the cost function during the first few iterations. After six iterations the value of the cost function has already decreased by about two orders of magnitude and the algorithm has practically converged.

Figure 2 shows the corresponding evolution of the nudging coefficients in the optimal-parameter-estimation procedure. During the first four to five iterations, all the nudging coefficients, $G_{\ln p_s}$, G_D and G_ζ, are observed to undergo a rapid increase. Then both $G_{\ln p_s}$ and G_ζ keep increasing slowly while G_D starts to decrease slowly. After 15 iterations

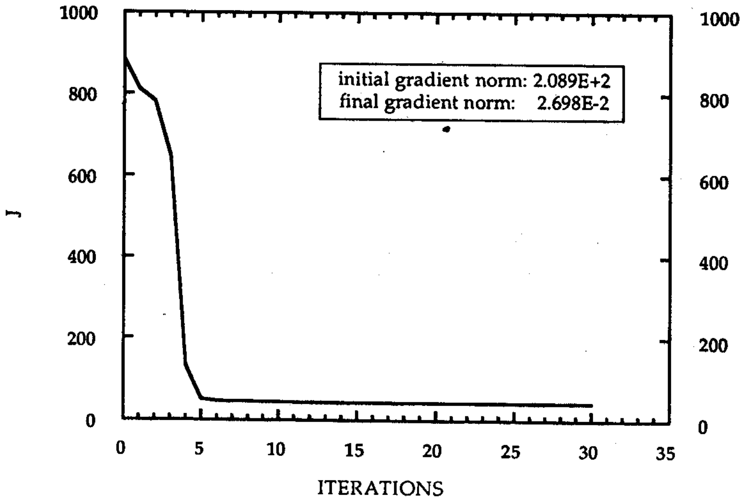

Figure 1. Variation of the cost function, $\mathcal{J}$, with the number of iterations. Both the initial conditions and the nudging coefficients, $G_{\ln p_s}$, G_D and G_ζ, serve as control variables. The norms of initial gradient, $\|g_0\|$, and final gradient, $\|g\|$, (30 iterations) are also given. The length of the assimilation window is 6 hours.

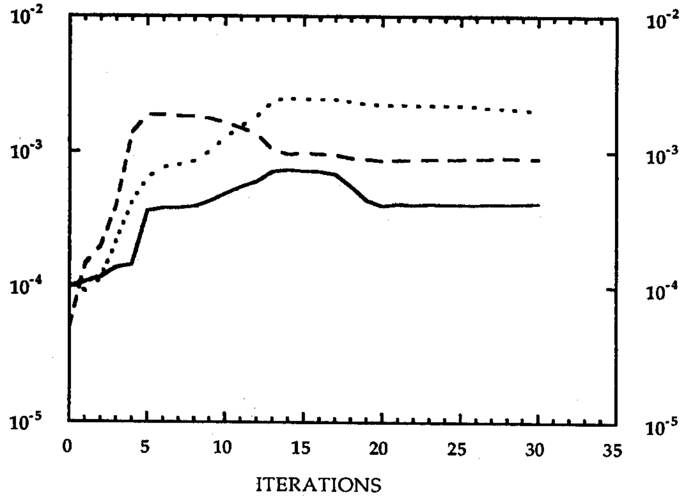

Figure 2. Variation of the nudging coefficients, $G_{\ln p_s}$ (——), G_D (– – –) and G_ζ (. . .), with the number of iterations. Both the initial conditions and the nudging coefficients $G_{\ln p_s}$, G_D and G_ζ serve as control variables. The length of the assimilation window is 6 hours.

an optimal value of the respective nudging coefficient was attained. The values for the optimal nudging coefficients are presented in Table 2. We found that the values of the optimal nudging coefficients are larger than the corresponding estimated nudging coefficients, as suggested by Krishnamurti *et al.* (1991). Noticing that the spectral truncation of our experiment (T40) is smaller than theirs (T106), the values obtained for the optimal nudging coefficients appear to be very reasonable since it is a general experience that the nudging coefficients increase with decreasing horizontal resolution.

A similar minimization for the parameter estimation was carried out where this time the length of the assimilation window was increased from 6 hours to 12 hours. In Figs. 3–4 we present the corresponding results from the above minimization. Similar results are obtained when we compare Figs. 3–4 with Figs. 1–2.

TABLE 2. OPTIMAL NUDGING PARAMETERS

Nudging coefficients	Nudge $\ln p_s$, D, ζ		Nudge T, D, ζ	
	6 hours	12 hours	6 hours	12 hours
$G_{\ln p_s}$	4.162E-4	4.276E-4		
G_T			3.366E-3	1.835E-3
G_D	8.889E-4	5.021E-3	1.021E-3	2.208E-3
G_ζ	1.992E-3	9.514E-4	8.245E-4	7.756E-4

See text for explanation of symbols.

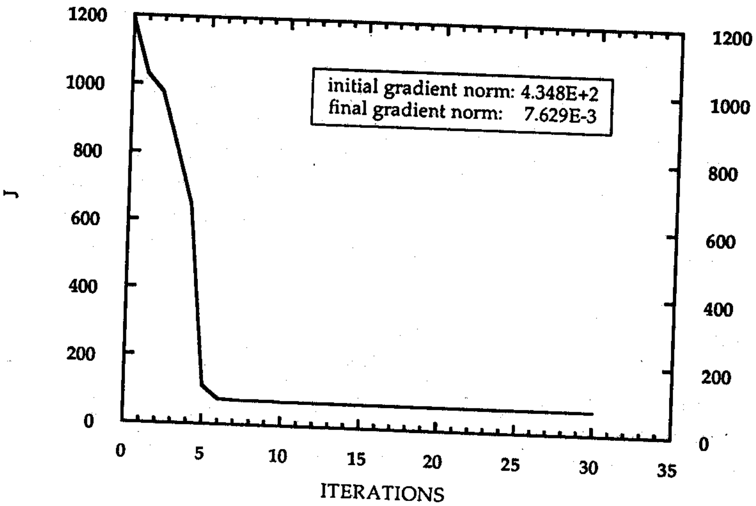

Figure 3. Same as Fig. 1 except that the length of assimilation window is 12 hours.

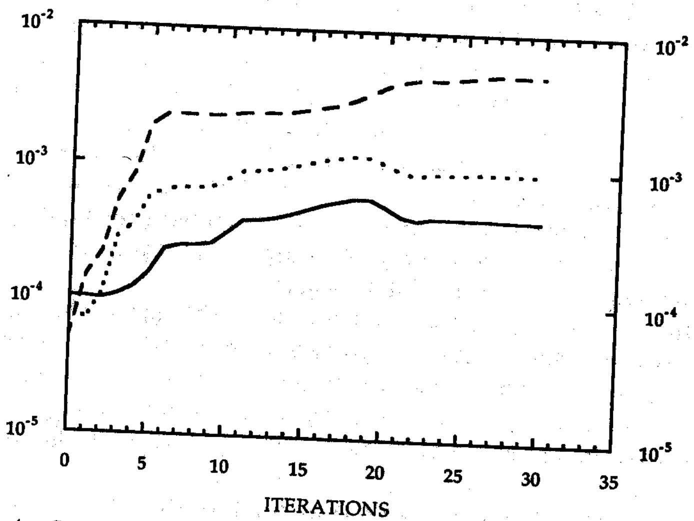

Figure 4. Same as Fig. 2 except that the length of assimilation window is 12 hours.

Instead of nudging the surface pressure which is important for data assimilation in the tropics (Krishnamurti *et al.* 1991; Brill *et al.* 1991), we decided to nudge the temperature field. The optimal nudging coefficients obtained by parameter estimation are shown in the last two columns of Table 2.

We note here that the optimal nudging coefficient does not tend to infinity as the minimization proceeds. With an infinite nudging coefficient G, X will be set to X^0 at time t_R, but $X(t_0)$ will tend to $X^0(t_0)$, i.e. the second term in (3.1) will be zero but the first term will remain non-zero. This explains why the infinite nudging coefficient G cannot minimize the cost function $\mathcal{J}$. Actually, when the minimization proceeds, both the initial state U and the nudging coefficient G are updated and the optimal solution of the minimization problem (3.1) is obtained because of the combined effect of the model dynamics and the nudging term.

(*c*) *Comparisons between estimated nudging, optimal nudging and VDA*

The NDA and VDA techniques are compared using observations at two time levels t_0 and t_R as described in section 3(a). Because observations were available at every model grid point and at the two time levels, $t = t_0$ and $t = t_R$, the predetermined estimation of the model variables at each time step for the NDA scheme was obtained by linear interpolation in time (Ramamurthy and Carr 1987). In this way information at both end points of the assimilation time interval is allowed to influence the assimilated value.

The experiment that follows consists of four types of assimilation:

(1) a control assimilation without nudging,

(2) an assimilation with estimated nudging coefficient,

(3) an assimilation with the optimal nudging coefficients determined by parameter estimation using VDA procedure, and

(4) VDA minimizing a cost function measuring the distance between the model solution and observations. An overview of the experiment is depicted in Fig. 5.

In this work, a value of G similar to that suggested by Krishnamurti *et al.* (1991) was used for the NDA experiment (called estimated NDA). Optimal nudging coefficients, determined in section 3(b) by a parameter-estimation procedure using the adjoint technique, were used for the optimal NDA experiment.

After obtaining both the estimated and optimal nudging coefficients for the four cases: (a) nudging the $\ln p_s$, D, ζ fields in a 6 h window, (b) nudging $\ln p_s$, D, ζ fields in a 12 h window, (c) nudging T, D, ζ fields in a 6 h window, and (d) nudging T, D, ζ in a 12 h window, we can perform eight parallel NDAs with both the estimated and optimal nudging coefficients. For the sake of comparison, and with a view to obtaining a better insight into the ability of the optimal NDA procedure, we also carried out four similar VDA experiments. The retrieved initial conditions after 30 iterations were taken as the final results of the VDA. The ensuing 6 h or 12 h integrations (depending on the length of the assimilation window) from the retrievals are used for carrying out a comparison with the corresponding results of the NDA schemes.

We do not intend to compare the NDA with the VDA approaches directly since the two methods have their own arbitrariness in the definition of the cost function, the choice of the minimization algorithm, the determination of stopping criteria, the length of assimilation window and the choice of nudging coefficient. However, the ability to reconstruct as accurately and economically as possible the state of the flow is of paramount importance.

Root-mean-square (r.m.s.) errors are computed for the aforementioned 12 assimilation experiments. Table 3 shows the total r.m.s. differences for all the model variables.

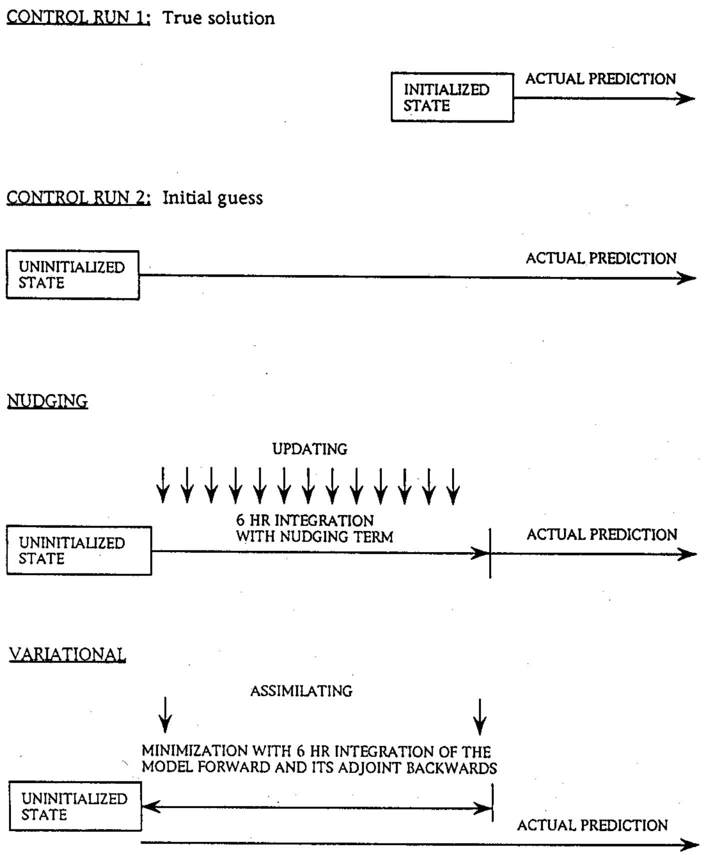

Figure 5. A schematic view of the assimilation-forecast cycle.

If we compare the results of the estimated NDA (nudging using estimated *ad hoc* nudging coefficients), the optimal NDA (nudging with optimal nudging coefficients) and the VDA procedures, we find that the estimated NDA approach yields the poorest results. Optimal NDA yields the best results, except in the case of temperature, divergence and vorticity field assimilation in a 6 h window where the VDA results in a slightly smaller r.m.s. value than the corresponding optimal NDA approach. Both the optimal NDA and the VDA perform much better than the estimated NDA procedure.

Figure 6 shows the vertical variation of the r.m.s. differences of the divergence field for the various aforementioned assimilation experiments. The r.m.s. difference of the control assimilation (dash-dotted line) is significantly reduced when an optimal NDA is applied for both the 6 h (Fig. 6(a)) and 12 h (Fig. 6(b)) assimilation windows. The 12 h optimal NDA yields a smaller r.m.s. difference than the corresponding 6 h optimal NDA. The opposite is true for the VDA. The estimated NDA in the 12 h window results in an even larger r.m.s. difference than the one obtained in the control assimilation experiment.

Since the available observations at time t_0 and the model-generated observations at time t_R are both free of gravity-wave oscillations while the initial guess for the assimilation contains gravity-wave oscillations, one way to estimate the quality of the assimilated fields is to find out the amount of the residual gravity oscillations for 24 h forecasts integrated from assimilated fields at time t_R. The results are shown in Figs. 7–10.

TABLE 3. DIFFERENCES BETWEEN THE ASSIMILATED AND THE TRUE MODEL INITIAL CONDITIONS BY ESTIMATED NUDGING, OPTIMAL NUDGING AND VARIATIONAL ASSIMILATIONS

Length of window	Assimilated variables	Fields	Estimated nudging	Optimal nudging	Variational assimilation
			Root-mean-square		
6 hours	$\ln p_s, D, \zeta$	$\ln p_s$			
		ζ	3.92E-4	1.37E-4	7.31E-5
		T	1.74E-1	2.18E-1	8.61E-2
		D	1.24E-6	2.95E-7	5.74E-7
		ζ	1.02E-6	2.06E-7	4.05E-7
		q	4.04E-5	4.40E-5	1.56E-5
	T, D, ζ	$\ln p_s$	7.11E-4	5.93E-4	2.71E-4
		T	1.06E-1	9.61E-3	1.82E-2
		D	1.28E-6	3.22E-7	2.72E-7
		ζ	1.04E-6	4.56E-7	1.41E-7
		q	4.32E-5	4.67E-4	3.33E-6
12 hours	$\ln p_s, D, \zeta$	$\ln p_s$	1.90E-3	1.68E-4	3.25E-4
		T	9.00E-1	3.96E-1	1.64E-1
		D	3.84E-6	5.27E-8	7.74E-7
		ζ	9.00E-6	4.26E-7	1.01E-6
		q	2.62E-4	1.12E-4	1.58E-5
	T, D, ζ	$\ln p_s$	5.07E-4	5.13E-4	5.36E-4
		T	1.54E-1	1.80E-2	9.67E-2
		D	1.55E-6	1.20E-7	6.97E-7
		ζ	1.94E-6	5.15E-7	9.20E-7
		q	8.00E-5	1.12E-4	2.82E-5

The evolution of the surface pressure at a point in the Indian Ocean (hereafter called Point 1) after 6 h assimilation is shown in Fig. 7. In Fig. 7(a) we present the results from a control run without nudging (dash-dotted line), an assimilation with estimated NDA of $\ln p_s$, D, and ζ fields (dotted line) and an assimilation experiment with optimal NDA of the same fields (solid line). The control assimilation exhibits the presence of a large amount of gravity oscillations while the oscillations in the forecast initialized by estimated NDA become relatively smaller. However, full damping of the gravity-wave oscillations occurs only when an optimal NDA scheme is applied. The results obtained using optimal NDA assimilation were then compared with the ones obtained from the VDA and the reference forecast (forecast from the observation at time 6 h), respectively. These results are displayed in Fig. 7(b). Only small differences may be perceived amongst them, the result from the VDA (dotted line) being somewhat closer to the reference forecast (dashed line) while the result from the optimal NDA experiment (solid line) being smoother.

Figure 8 presents a 24 h forecast ensuing from the 12 h assimilation. The forecast from the estimated NDA (dotted line) still contains some gravity-wave oscillations with opposite phase to the control assimilation (dash-dotted line). The optimal NDA yields a satisfactory damping of the gravity oscillations, and seems to perform better than the VDA (Fig. 8(b)), provided that the minimization of the cost function in the VDA was stopped after 30 iterations as before.

Finally, Figs. 9–10 display the forecast from assimilations when observations of the temperature, divergence and vorticity fields are available, i.e. when the surface pressure field is not nudged in the NDA scheme and the weighting coefficient for the surface pressure in the cost function is set to zero in the VDA. Figure 9 shows the variation of

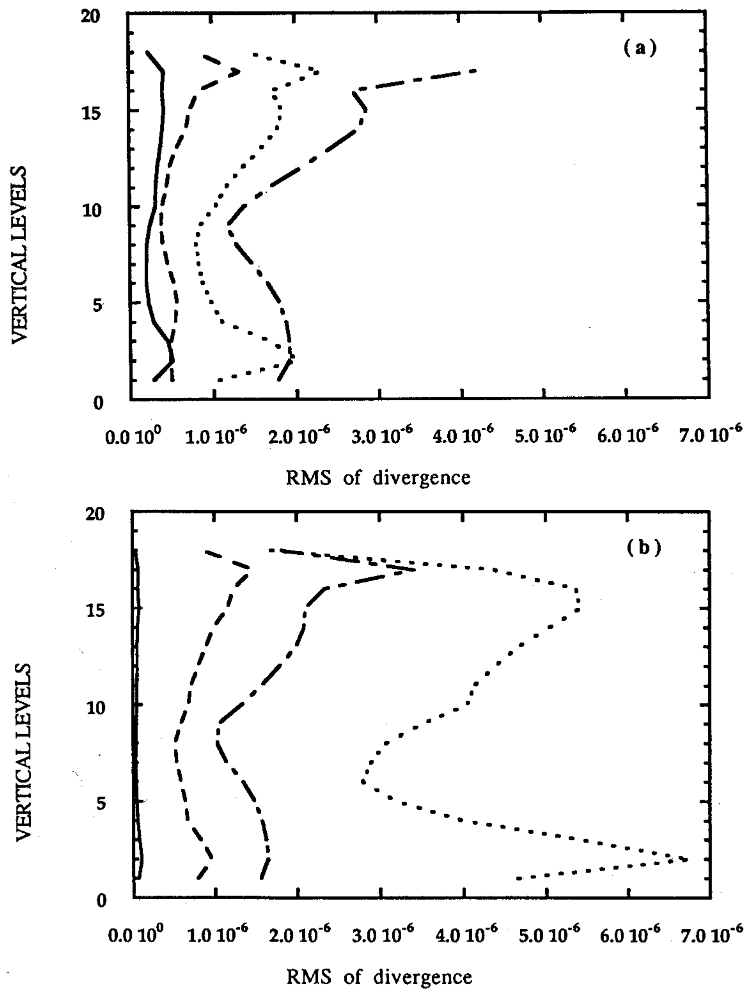

Figure 6. Distributions of the root-mean-square differences for the divergence field between the true solution and the assimilated fields at the end of the (a) 6-hour and (b) 12-hour assimilation intervals. Results are shown for different assimilation experiments: control run (—·—), estimated nudging (· · ·), variational data assimilation (– – –) and optimal nudging (—). The surface pressure, divergence and vorticity fields are nudged.

the surface pressure at Point 1, and Fig. 10 at Point 2, which is located near the Rocky Mountains. Again the optimal NDA procedure yields the strongest damping of the gravity oscillations while the VDA matches well the reference forecast. Here we found that even without directly nudging the surface pressure, gravity-wave oscillations present in the ensuing 24 h forecast in the surface pressure field have practically vanished owing to the mutual dynamical adjustment between the variables in the model.

In view of these initial experimental results, we are led to the conclusion that the optimal NDA procedure performs much better than the estimated NDA procedure and compares favourably with the VDA approach for the data set used here. It is worth noting that the computational cost (or the total CPU time) of NDA is much cheaper than the corresponding computational cost required by the VDA. Optimal nudging coefficients can be obtained in at most seven unconstrained minimization iterations of the cost function when both the initial conditions and the nudging coefficients serve as control variables for the minimization of the cost function. For real data assimilation,

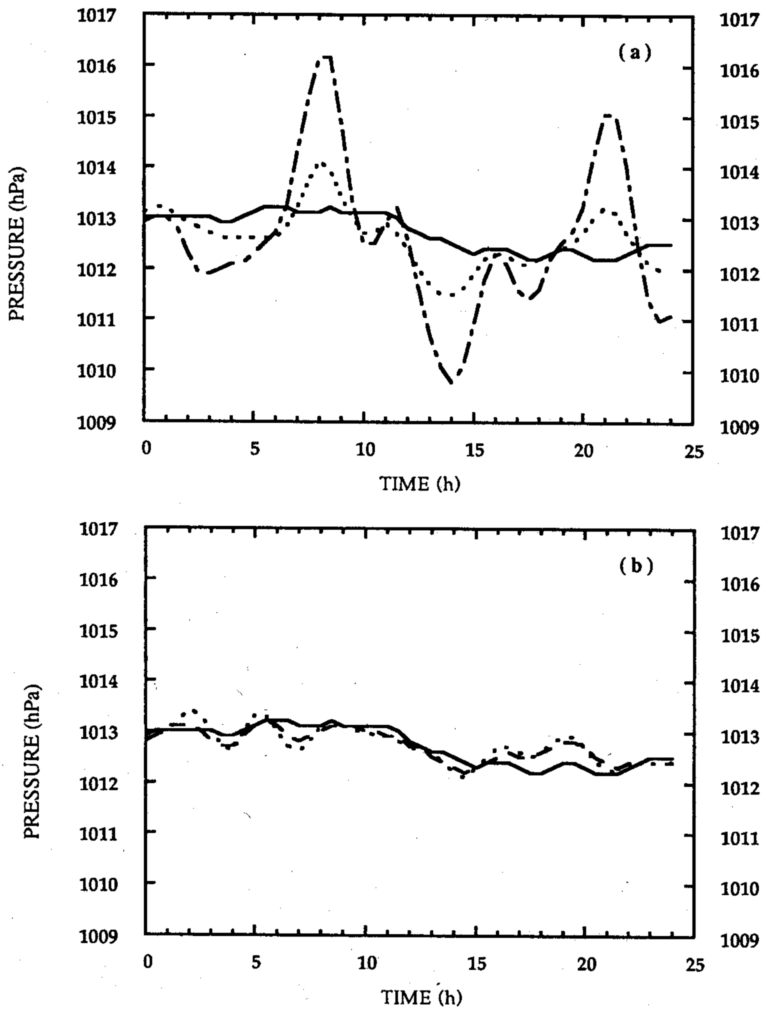

Figure 7. Temporal variation, over a 24-hour period, of the surface pressure field at Point 1 in the Indian Ocean from the assimilated fields with a 6-hour assimilation window. (a) Without nudging (—·—), with estimated nudging coefficients (· · ·) and optimal nudging coefficients (—). (b) Variational data assimilation (. . .), optimal nudging coefficients (—) and the true solution (---). The surface pressure, divergence and vorticity are nudged.

one has to carry out only a parameter estimation to obtain the optimal nudging coefficients which will result in a sizably smaller dimension of the control variable vector. Once the optimal nudging coefficient is obtained, a single run of the model is required to accomplish the optimal NDA. The idea put forward here is to combine the NDA procedure with the VDA approach in an optimal way in order to obtain a practical, satisfactory and implementable data-assimilation scheme for operational applications.

(*d*) *Space variable optimal nudging*

One factor governing the efficiency of the NDA process for damping of model errors is the ratio of the time-scale G^{-1} to that given by the Coriolis parameter, f^{-1} (Lorenc *et al.* 1991). Since the time-scales of the variables over different portions on the sphere and on different vertical levels may vary strongly, use of optimal nudging coefficients G,

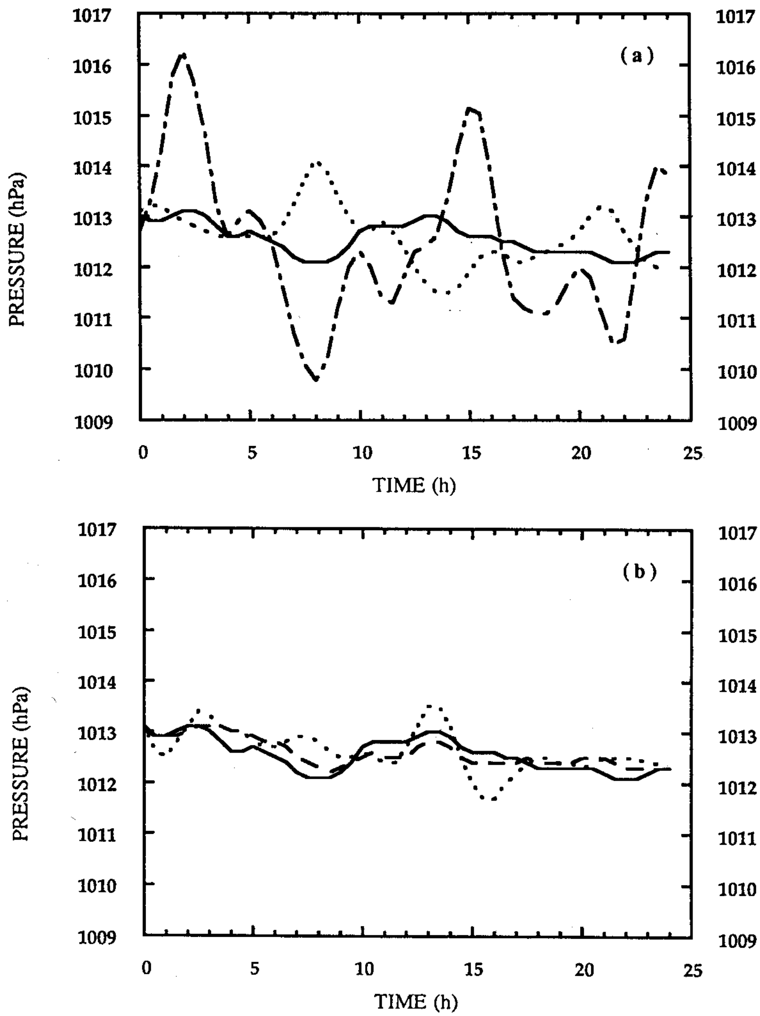

Figure 8. Same as Fig. 7 except the length of the assimilation window is 12 hours.

varying both in time as well as over the horizontal domain and on the different vertical levels, is desirable.

Lyne (1979) has shown that if G^{-1} is too small, assimilation of wind data can be hindered. This result is consistent with the empirically determined optimum value of a relaxation coefficient, λ, for equatorial latitudes being smaller than the value in the northern hemisphere, where λ represents the fraction of the analysed increment that is added to the model field in an analysis correction data-assimilation scheme (Lorenc *et al.* 1991). The analysis correction data-assimilation scheme is equivalent to the NDA scheme if

$$\lambda = G\Delta t/\{1 + G\Delta t\}.$$

Variations of the forcing rate for observational increments λ reflect the different forecast-error scales, data distributions and dynamical adjustment characteristics of the latitude band and different model atmosphere layers.

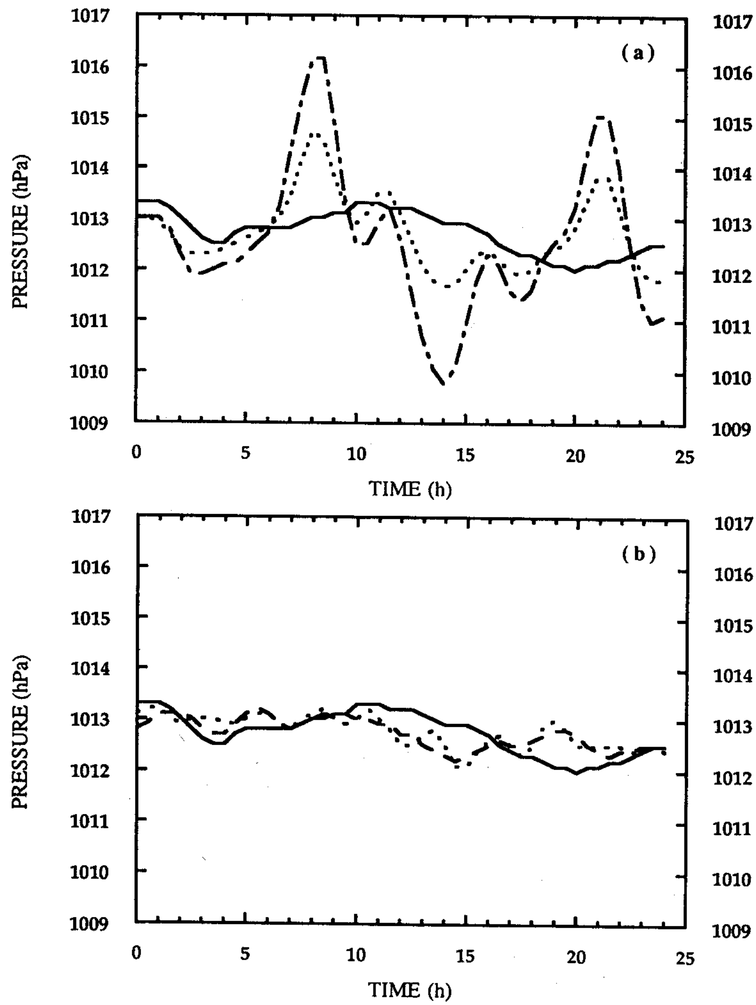

Figure 9. Same as Fig. 7 except that the temperature field is nudged instead of the surface pressure field.

In most of the conventional NDA experiments, estimated or experimental values of nudging coefficients were used, in which the nudging coefficient for a particular nudged variable is invariable in space. In a recent experiment by Ramamurthy and Xu (personal communication), nudging coefficients multiplied by a so-called confidence factor designed to take into account the three-dimensional distribution of new observations were used. The normalized confidence factors, which range from 0 at those points where no new data is available for the analysis to 1 where perfect observations are available, are based on the ratio of first-guess error estimate to analysis error estimate. However, their preliminary result shows that the use of the confidence factor, based on the analysis and background error estimates, did not have an appreciable impact when compared with nudging everywhere with a constant confidence factor of 1.

Here the optimal variable nudging coefficients are obtained by the same parameter-estimation procedure as used to obtain the previous optimal constant nudging coefficient. The only difference from the previous experiments is the increase in the dimension of the control variables vector. Since our observations are perfect, we know in advance that the spatial variations of the optimal variable nudging coefficients will be very small (i.e.

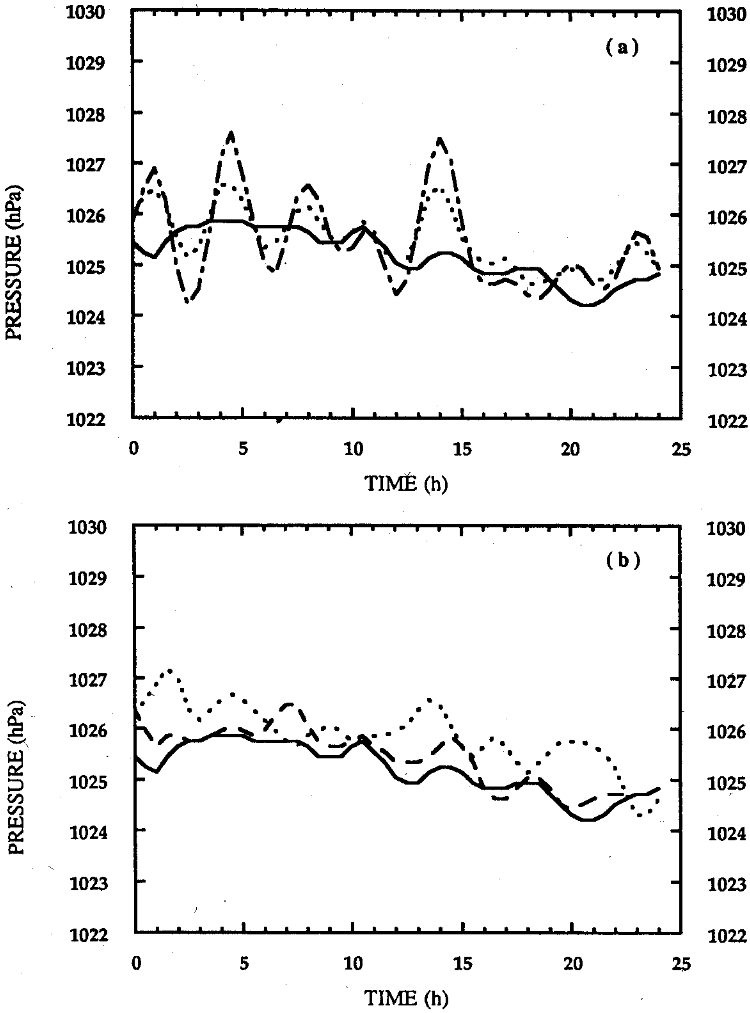

Figure 10. Same as Fig. 9 except at Point 2 near the Rocky Mountains.

Figure 11. Horizontal distribution of the variable optimal nudging coefficient $G_{\ln p_s}$ for the surface pressure field. Contour interval is 2×10^{-8} from 4.161×10^{-4} to 4.164×10^{-4}. Values in the figure are scaled by 10^{-5}.

the confidence factor is 1 everywhere). Figure 11 shows, for example, the horizontal distribution of the optimal nudging coefficient of surface pressure, $G_{\ln p_s}$. The spatial variability is less than 0.06%. Therefore, using a constant optimal nudging coefficient for each variable is satisfactory for perfect observations' assimilation.

From this experiment we are not able to draw the conclusion whether the optimal variable nudging will be more beneficial than constant optimal nudging because of the perfect observations we are using. However, the ability of determining variable optimal nudging coefficients by parameter estimation in the framework of VDA constitutes a major advantage of the optimal NDA over the estimated NDA approach, since it is extremely difficult to determine space-varying nudging coefficients by numerical experimentation. Moreover, the variable optimal NDA method does not require statistical information.

4. Conclusions

In this study we used the adjoint model of an adiabatic version of the NMC spectral multilayer model with diffusion and surface drag, developed in the VDA system framework, to determine optimal nudging coefficients (which can vary in 3-D space) effectively using a parameter-estimation approach to be applied in a NDA procedure. This procedure is much more economical than the VDA owing to the tremendous computational cost of the latter scheme.

Our experiments show that optimal NDA is practically implementable and performs very well in as far as the convergence and quality of the resulting assimilated state are concerned.

This study of VDA, NDA, and optimal NDA using the optimal parameter-estimation approach, indicates the potential future of optimal NDA application in operational data assimilation where observations in a certain time window can be effectively incorporated into the model so as to provide the best model initial conditions. However, it should be emphasized that a final assessment of the usefulness of the optimal NDA procedure has to await further experiments using more realistic conditions.

The idea here is to exploit the advantage of VDA's capability of performing a parameter estimation to obtain optimal nudging coefficients and then to revert to the computationally efficient NDA. The whole procedure of determining optimal nudging coefficients and carrying out nudging with these optimal coefficients is conceptually nearly equivalent to the implementation of the KF data assimilation. The main advantage of optimal nudging over the KF is that the cost of computation of the forecast-error covariance, which is the central ingredient of the KF, was avoided owing to the use of an adjoint model integration.

In a follow-up paper we will apply this optimal NDA procedure to more realistic situations, such as using observations which are not perfect, using the entire physical package of the NMC model.

Acknowledgements

The authors gratefully acknowledge the support of Dr Joseph Pandolfo, and Dr Pamela Stephens, Directors of GARP Division of the Atmospheric Sciences at the National Science Foundation (NSF). This research was funded by NSF Grant ATM-9102851. Additional support was provided by the Supercomputer Computations Research Institute at Florida State University, which is partially funded by the Department of Energy through contract No. DE-FC0583ER250000. The authors would like to thank Andrew Lorenc whose insightful remarks helped to clarify the paper.

APPENDIX A

Derivation of the continuous expression of $\nabla_G \mathcal{J}$

The dynamical model equations (2.1) viewed as strong constraints can be enforced by introducing a set of undetermined Lagrange multipliers. This leads to the formulation of the Lagrange function, given as

$$L(X, G, P) = \mathcal{J}(X, G) + \int_{t_0}^{t_R} \left\langle P, \frac{\partial X}{\partial t} - F(X) - G(X^o - X) \right\rangle \mathrm{d}t \tag{A.1}$$

where $\langle\ \rangle$ is the inner product of two vectors and P is the Lagrange multiplier vector, $\mathcal{J}$ is defined in (2.7), G is the nudging coefficient and $t_R - t_0$ is the time length of the assimilation window.

The constrained optimization problem is then replaced by a series of unconstrained minimization problems with respect to the variables U and G. By doing so, the problem of minimizing the cost function, subject to the model equations, becomes a problem of finding the stationary points of the Lagrange function. This in turn is equivalent to the determination of U and G subject to the condition that the gradient of the Lagrange function vanishes. This results in the following set of equations:

$$\frac{\partial L(X, G, P)}{\partial P} = 0 \tag{A.2}$$

$$\frac{\partial L(X, G, P)}{\partial X} = 0 \tag{A.3}$$

$$\frac{\partial L(X, G, P)}{\partial G} = 0. \tag{A.4}$$

Equation (A.2) recovers the original model equation (2.4), while (A.3) becomes

$$\frac{\partial \mathcal{J}}{\partial X} + \frac{\partial}{\partial X} \int_{t_0}^{t_R} \left\langle P, \frac{\partial X}{\partial t} - F(X) - G(X^o - X) \right\rangle \mathrm{d}t = 0. \tag{A.5}$$

Substituting (2.6) into (A.5) and using integration by parts we obtain

$$\int_{t_0}^{t_R} 2W(X - X^o)\,\mathrm{d}t + \frac{\partial}{\partial X}\left\{P(t_R)X(t_R) - P(t_0)X(t_0) - \int_{t_0}^{t_R} X\frac{\partial P}{\partial t}\right\} -$$
$$- \int_{t_0}^{t_R} \left\langle \left[\frac{\partial F}{\partial X}\right] - G, P \right\rangle \mathrm{d}t = 0. \tag{A.6}$$

Using initial conditions (2.2) and assuming

$$P(t_R) = 0 \tag{A.7}$$

(A.6) becomes

$$\int_{t_0}^{t_R} 2W(X - X^o)\,\mathrm{d}t - \int_{t_0}^{t_R} \left\{\frac{\partial P}{\partial t} + \left[\frac{\partial F}{\partial X}\right]^{\mathrm{T}} P - G^{\mathrm{T}}P\right\} \mathrm{d}t = 0 \tag{A.8}$$

for any length of the assimilation window $t_R - t_0$, where $(\ )^{\mathrm{T}}$ represents the transpose. Thus the integrated function should be zero, which results in the adjoint equation, given by

$$\left.\begin{aligned}&\frac{\partial P}{\partial t}+\left[\frac{\partial F}{\partial X}\right]^{\mathrm{T}}P-G^{\mathrm{T}}P=W(X-X^{\mathrm{o}})\\&P(t_{\mathrm{R}})=0\end{aligned}\right\}\tag{A.9}$$

which is the same as (2.14).

It is worth noting that the adjoint equation has a form similar to the original model equation, except for two important features. The first feature is that the nudging term in the adjoint equation has opposite sign to that in the model equation. The second feature relates to stability of the well-posed problem, thus requiring the integration of the adjoint equation to be backwards in time. In addition, the driving factor for the adjoint equation is the weighted difference between the model and the observations. The Lagrange multipliers carry the information about the data back to the initial time to influence the reconstruction of the model state.

Substituting (A.1) into (A.4), we have

$$2K(G-\hat{G})-\int_{t_0}^{t_R}\langle P,X^{\mathrm{o}}-X\rangle\,\mathrm{d}t=0.\tag{A.10}$$

Starting with a first guess for the unknown nudging forcing coefficient, the model equations (2.4)–(2.5) are integrated forward to obtain the state trajectory corresponding to this guess, and the adjoint equation (A.9) is then integrated backwards to compute the corresponding Lagrange multipliers, which provide information about the gradient of the cost function at the point of the first guess. The gradient of the cost function with respect to the nudging parameter is then given by the left-hand side of (A.10), i.e.

$$\nabla_G\mathcal{J}=2K(G-\hat{G})-\int_{t_0}^{t_R}\langle P,X^{\mathrm{o}}-X\rangle\,\mathrm{d}t.\tag{A.11}$$

APPENDIX B

Derivation of the discrete expression of $\nabla_G\mathcal{J}$

The time difference equations of (2.1)–(2.2) can be written as

$$X(t_0+\Delta t)=X(t_0)+\Delta tF(X(t_0))+\Delta tG(X^{\mathrm{o}}(t_0+\Delta t)-X(t_0+\Delta t))\tag{B.1}$$

$$\begin{aligned}X(t_i+\Delta t)=X(t_i-\Delta t)+2\Delta tF(X(t_i))+\Delta tG(X^{\mathrm{o}}(t_i+\Delta t)-X(t_i+\Delta t))\\ i=1,2,\ldots,R-1.\end{aligned}\tag{B.2}$$

The cost function defined in (2.7) now assumes the form

$$\mathcal{J}(G)=\sum_{i=0}^{R}\langle W(X(t_i)-X^{\mathrm{o}}(t_i)),X(t_i)-X^{\mathrm{o}}(t_i)\rangle+\langle K(G-\hat{G}),G-\hat{G}\rangle\tag{B.3}$$

For the leapfrog time difference scheme, the Lagrange function can be written as

$$\begin{aligned}L=&\sum_{i=0}^{R}\langle W(X(t_i)-X^{\mathrm{o}}(t_i)),X(t_i)-X^{\mathrm{o}}(t_i)\rangle+\langle K(G-\hat{G}),G-\hat{G}\rangle+\\&+\langle P_0,X(t_0+\Delta t)-X(t_0)-\Delta tF(X(t_0))-\Delta tG(X^{\mathrm{o}}(t_0+\Delta t)-X(t_0+\Delta t))\rangle+\\&+\sum_{i=1}^{R-1}\langle P_i,X(t_i+\Delta t)-X(t_i-\Delta t)-2\Delta tF(X(t_i))-2\Delta tG(X^{\mathrm{o}}(t_0+\Delta t)-\\&-X(t_0+\Delta t))\rangle.\end{aligned}\tag{B.4}$$

The equations for the Lagrange multipliers, or the adjoint equations, are found by letting the first-order variation of (B.4) with respect to $X(t_1)$, $X(t_2)$, . . ., $X(t_R)$ vanish. The equation resulting from differentiation with respect to $X(t_R)$ will be considered first, then the equation with respect to $X(t_{R-1})$, . . ., and finally the equation with respect to $X(t_1)$.

$$P_{R-1} + 2\Delta t G P_{R-1} + 2W(X(t_R) - X^{o}(t_R)) = 0 \tag{B.5}$$

$$P_{R-1} + 2\Delta t G P_{R-1} - 2\Delta t \left[\frac{\partial F}{\partial X(t_{R-1})}\right] P_{R-1} + 2W(X(t_{R-1}) - X^{o}(t_{R-1})) = 0 \tag{B.6}$$

$$P_{i-1} + 2\Delta t G P_{i-1} - 2\Delta t \left[\frac{\partial F}{\partial X(t_i)}\right] P_i - P_{i+1} + 2W(X(t_i) - X^{o}(t_i)) = 0$$
$$i = R-2, R-3, \ldots, 1. \tag{B.7}$$

Lastly, from the differentiation of L with respect to G one finds

$$2K(G - \hat{G}) - \Delta t P_0(X^{o}(t_0 + \Delta t) - X(t_0 + \Delta t)) - \sum_{i=1}^{R-1} 2\Delta t P_i(X^{o}(t_i + \Delta t) -$$
$$- X(t_i + \Delta t)) = 0 \tag{B.8}$$

The gradient of the cost function with respect to the nudging parameter is then given by the left-hand side of (B.8), i.e.

$$\nabla_G \mathcal{J} = 2K(G - \hat{G}) - \Delta t P_0(X^{o}(t_0 + \Delta t) - X(t_0 + \Delta t)) -$$
$$- \sum_{i=1}^{R-1} 2\Delta t P_i(X^{o}(t_i + \Delta t) - X(t_i + \Delta t)). \tag{B.9}$$

REFERENCES

Anthes, R. A. 1974 Data assimilation and initialization of hurricane prediction models. *J. Atmos. Sci.*, **31,** 702–719

Brill, K. F., Uccellini, L. W., Manobianco, J., Kocin, P. J. and Homan, J. H. 1991 The use of successive dynamic initialization by nudging to simulate cyclogenesis during GALE IOP 1. *Meteorol. Atmos. Phys.*, **45,** 15–40

Davies, H. E. and Turner, R. E. 1977 Updating prediction models by dynamic relaxation. An examination of the technique. *Q. J. R. Meteorol. Soc.*, **103,** 225–245

Ghil M. and Malanotte-Rizzoli, P. 1991 Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33,** 141–265

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

Kalman, R. E. 1960 A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.*, **82D,** 35–45

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

Kuo, Y. H. and Guo, Y. R. 1989 Dynamic initialization using observations from a network of profilers and its impact on short-range numerical weather prediction. *Mon. Weather Rev.*, **117,** 1975–1998

LeDimet F. X. and Talagrand, O. 1986 Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A,** 97–110

Liu, D. C. and Nocedal, J. 1989 On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, **45,** 1177–1194

Lorenc, A. C. 1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194

Lorenc, A. C., Bell, R. S. and Macpherson, B. 1991 The Meteorological Office analysis correction data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **117**, 59–89

Lyne, W. H. 1979 'Data assimilation by repeated correction of model fields—theoretical considerations'. Met. O.11 Tech. Note No. 130. Meteorological Office

Lyne, W. H., Swinbank, R. and Birch, N. T. 1982 A data assimilation experiment and the global circulation during the FGGE special observing periods. *Q. J. R. Meteorol Soc.*, **108**, 575–594

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

Nocedal, J. 1980 Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, **35**, 773–782

Panchang, V. G. and O'Brien, J. J. 1989 On the determination of hydraulic model parameters using the strong constraint formulation. Pp. 5–18 in *Modeling marine systems*, **I**. Ed. A. M. Davies. CRC Press, Inc.

Ramamurthy, M. K. and Carr, F. H. 1987 Four-dimensional data assimilation in the monsoon region. Part I: Experiments with wind data. *Mon. Weather Rev.*, **115**, 1678–1706

1988 Four-dimensional data assimilation in the monsoon region. Part II: Role of temperature and moisture data. *Mon. Weather Rev.*, **116**, 1896–1913

Smedstad, O. M. and O'Brien, J. J. 1991 Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model. *Prog. Oceanog.*, **26**, 179–241

Stauffer, D. R. and Seaman, N. L. 1990 Use of four-dimensional data assimilation in a limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Mon. Weather Rev.*, **118**, 1250–1277

Wang, W. and Warner, T. T. 1988 Use of four-dimensional data assimilation by Newtonian relaxation and latent heat forcing to improve a mesoscale-model precipitation forecast: A case study. *Mon. Weather Rev.*, **116**, 2593–2613