Tellus (1993), 45A, 358–369
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

Optimal determination of nudging coefficients using the adjoint equations

By DAVID R. STAUFFER* and JIAN-WEN BAO, *Department of Meteorology, The Pennsylvania State University, 503 Walker Building, University Park, PA 16802, USA*

(Manuscript received 2 November 1992; in final form 1 June 1993)

ABSTRACT

The adjoint equations of a numerical model can be used for model-parameter estimation. In this study, a general computational procedure is developed to determine the size and distribution of any internal model parameter. The procedure is then applied to a one-dimensional shallow-water model in the context of analysis-nudging four-dimensional data assimilation (FDDA): the weighting coefficients used by the Newtonian nudging technique are determined such that the model error during the assimilation period is optimally reduced subject to some constraints. The sensitivity of these nudging coefficients to the optimal objectives and constraints is investigated using this simple grid-point model in an Observing Systems Simulation Experiments (OSSE) mode. The results show that in principle, it is feasible to determine a set of nudging weights which minimize the model error over the period covered by the observations. It is demonstrated, however, that the magnitude and distribution of these "optimal" nudging weights are sensitive to the prescribed estimate of the nudging weights and the corresponding coefficient matrix which define a penalty term in the cost function. The penalty term is a *weak* constraint on the size and distribution of the optimal nudging weights while the model is the *strong* constraint. The fit of the model to the data is greater when this constraint on the nudging weights is weaker, but then the nudging weights may be too large or even negative. Thus the "optimal" solution for this model parameter is not unique because specification of this penalty term in the cost function introduces a new uncertainty into the nudging FDDA framework. Nevertheless, this optimal-nudging approach does show promise, but the sensitivity of the technique to the penalty term requires further investigation under more realistic conditions.

1. Introduction

Increasing computer power and advanced observing systems which now provide near-continuous streams of information on evolving atmospheric processes have led to correspondingly finer resolution numerical weather prediction models in the hope of improving our understanding and our ability to simulate complex atmospheric processes. Continuous four-dimensional data assimilation (FDDA) schemes capable of effectively utilizing this data base are important not only for improving the initialization of mesoscale forecasts, but also for generating high-quality four-dimensional mesoscale analyses for diagnostic purposes and for input into air-quality models (Stauffer et al., 1993). These two FDDA applications are known as dynamic initialization and dynamic analysis, respectively.

Currently, Newtonian relaxation (nudging) and variational (adjoint) methods are two of the most promising techniques for handling these modern asynoptic data sets, and are thus being widely studied by both the research and operational scientific communities. The adjoint method (Le Dimet and Talagrand, 1986) uses a dynamic model as a strong constraint in a four-dimensional variational scheme. This variational four-dimensional analysis (VFDA) technique essentially fits

* Corresponding author.

the model to data distributed over a finite period of time. If model errors are neglected, this technique determines the optimal initial state such that the corresponding model evolution best fits the observations through the entire period. However, the scheme's complexity and extensive computational requirements currently limit its use for operational applications, and also make it unsuitable for most research applications on the mesoscale.

The nudging technique (Anthes, 1974), which is economically more efficient and also more flexible than the adjoint method, relaxes the model state toward the observations by adding to one or more of the prognostic equations, artificial tendency terms which are proportional to the difference between the two states. The "proportionality constants" are *positive* and usually specified based on scaling arguments so that the nudging tendencies are relatively small compared to the other terms in the governing equations. They also account for the time and space separation of the model solution from the data, as well as data quality and representativeness. The model solution can be relaxed toward gridded analyses of observations, or directly toward the observations which are distributed nonuniformly in space and time (Stauffer and Seaman, 1990; Stauffer et al., 1991). These two nudging techniques are hereafter referred to as "analysis nudging" and "obs nudging", respectively. Although the obs-nudging approach is especially attractive for assimilation of these asynoptic data sets, as a first step, this paper will focus on the simpler analysis-nudging technique and how the nudging weights defined over a model grid may be determined using optimal control theory.

The adjoint equations can be used to efficiently compute the gradient of a cost function with respect to a "control variable". When the adjoint method is used for model initialization, the control variable is the model initial state. For model-parameter estimation, the control variable is the vector of model parameters to be determined. Zou et al. (1993) have recently applied parameter estimation using the adjoint equations to the nudging coefficients in an adiabatic version of the National Meteorological Center T40 spectral model initialized with operationally analyzed data and run in an Observing Systems Simulation Experiments (OSSE) mode. That is, the observations are model-generated. In their approach, both the initial state and the analysis-nudging coefficients were control variables. They concluded that the technique works and shows promise when compared to standard analysis nudging and variational data assimilation methods. In this paper, which represents a parallel and independent research effort on this subject, the initial state is not included in the control-variable vector, the parameter-estimation technique is applied to a model in grid-point space, and most importantly, some of the *limitations and sensitivities* associated with optimal nudging are addressed for the first time.

A general computational procedure for determining the size and distribution of any model parameter using optimal control theory is developed in Section 2, and then applied to a one-dimensional shallow-water model in the context of analysis-nudging FDDA (Section 3). The nudging weights are determined such that the model error during the assimilation period is optimally reduced subject to given constraints. The main focus of the paper is to investigate the sensitivity of these nudging coefficients to the optimal objectives and constraints, which is made easier by using a simple grid-point model in an OSSE mode. This is a logical first step before applying this approach to real data and more complicated models. The results are summarized in Section 4, and future directions for continued research and development of the technique are also discussed.

2. Development of a general parameter-estimation procedure

The parameter-estimation problem in meteorology is essentially contained in the following: minimize a "cost function" or functional $J = J(\boldsymbol{x}(\boldsymbol{p}, t), \boldsymbol{p})$ subject to the model constraint

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = \boldsymbol{M}(\boldsymbol{x}, \boldsymbol{p}, t),$$

$$\boldsymbol{x}(0) = \boldsymbol{X}_0,$$

where $\boldsymbol{p}$ is the vector of parameters to be determined, which is assumed here to be constant in time, $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{p}, t)$ is the model-state vector, and $\boldsymbol{M}$ is the model vector function of $\boldsymbol{x}$, $\boldsymbol{p}$ and t. The functional, J, is defined in a quadratic form:

$$J(\boldsymbol{p})=\int_0^T D[\boldsymbol{x}(\boldsymbol{p},t),t]\,\mathrm{d}t + (\boldsymbol{p}-\tilde{\boldsymbol{p}})^{\mathrm{tr}}\boldsymbol{K}(\boldsymbol{p}-\tilde{\boldsymbol{p}}), \tag{1}$$

where $D[\boldsymbol{x}(\boldsymbol{p},t),t]$ is the measure of the "fit" or discrepancy between the model state and the observation at t, $\tilde{\boldsymbol{p}}$ is a constant vector that controls the size of $\boldsymbol{p}$ and thus represents prior knowledge about the magnitude of the parameter, the superscript, tr, denotes the transpose of a vector or a matrix, $\boldsymbol{K}$ is a symmetric matrix whose inverse represents the covariance matrix of estimation error on $\boldsymbol{p}$, and T is the assimilation time window. For simplicity, we assume that $\boldsymbol{K}$ is a diagonal matrix whose elements are constant. Eq. (1) represents a trade-off between two optimal objectives: (a) minimize the discrepancy between the model state and the observations, and (b) make the magnitudes of $\boldsymbol{p}$ as close to those prescribed as possible (e.g., when $\tilde{\boldsymbol{p}}=0$, objective (b) simply means to make $\boldsymbol{p}$ as small as possible). Assume that $D[\boldsymbol{x}(\boldsymbol{p},t),t]$ is quadratic as follows

$$D[\boldsymbol{x}(\boldsymbol{p},t),t]=[\boldsymbol{x}(\boldsymbol{p},t)-\boldsymbol{x}_{\mathrm{obs}}(t)]^{\mathrm{tr}} \times \boldsymbol{W}(t)[\boldsymbol{x}(\boldsymbol{p},t)-\boldsymbol{x}_{\mathrm{obs}}(t)], \tag{2}$$

where the subscript, obs, denotes the observation, and $\boldsymbol{W}(t)$ is the inverse of the covariance matrix of observation errors, and is assumed here to be a diagonal matrix whose elements are the weighting coefficients of the observed variables.

The gradient of J with respect to $\boldsymbol{p}$ is needed in order to efficiently minimize J. To obtain the gradient, the model constraint is appended to the functional to be minimized using the Lagrange multipliers. These multipliers are contained in the adjoint vector $\boldsymbol{a}(t)$. We define the Lagrange function J' as:

$$J'(\boldsymbol{p})=\int_0^T \left\{D[\boldsymbol{x}(\boldsymbol{p},t),t] + \boldsymbol{a}(t)^{\mathrm{tr}}\left[\frac{\mathrm{d}\boldsymbol{x}(\boldsymbol{p},t)}{\mathrm{d}t}-\boldsymbol{M}(\boldsymbol{x},\boldsymbol{p},t)\right]\right\}\mathrm{d}t + (\boldsymbol{p}-\tilde{\boldsymbol{p}})^{\mathrm{tr}}\boldsymbol{K}(\boldsymbol{p}-\tilde{\boldsymbol{p}}). \tag{3}$$

Defining

$$H\equiv D-\boldsymbol{a}^{\mathrm{tr}}\boldsymbol{M}, \tag{4}$$

we have

$$J'(\boldsymbol{p})=\int_0^T\left[H+\boldsymbol{a}(t)^{\mathrm{tr}}\frac{\mathrm{d}\boldsymbol{x}(\boldsymbol{p},t)}{\mathrm{d}t}\right]\mathrm{d}t + (\boldsymbol{p}-\tilde{\boldsymbol{p}})^{\mathrm{tr}}\boldsymbol{K}(\boldsymbol{p}-\tilde{\boldsymbol{p}}). \tag{5}$$

After integration by parts, the second integral becomes

$$[\boldsymbol{a}(t)^{\mathrm{tr}}\boldsymbol{x}(\boldsymbol{p},t)]_{t=T}-[\boldsymbol{a}(t)^{\mathrm{tr}}\boldsymbol{x}(\boldsymbol{p},t)]_{t=0} - \int_0^T\frac{\mathrm{d}\boldsymbol{a}(t)^{\mathrm{tr}}}{\mathrm{d}t}\boldsymbol{x}(\boldsymbol{p},t)\,\mathrm{d}t.$$

Thus, J' may be rewritten as

$$J'(\boldsymbol{p})=[\boldsymbol{a}(t)^{\mathrm{tr}}\boldsymbol{x}(\boldsymbol{p},t)]_{t=T} - [\boldsymbol{a}(t)^{\mathrm{tr}}\boldsymbol{x}(\boldsymbol{p},t)]_{t=0} + \int_0^T\left[H-\frac{\mathrm{d}\boldsymbol{a}(t)^{\mathrm{tr}}}{\mathrm{d}t}\boldsymbol{x}(\boldsymbol{p},t)\right]\mathrm{d}t + (\boldsymbol{p}-\tilde{\boldsymbol{p}})^{\mathrm{tr}}\boldsymbol{K}(\boldsymbol{p}-\tilde{\boldsymbol{p}}). \tag{6}$$

Because the initial state $\boldsymbol{X}_0$ is not a function of $\boldsymbol{p}$, the first variation of $J'(\boldsymbol{p})$ is

$$\delta J'=[\boldsymbol{a}(t)^{\mathrm{tr}}\delta\boldsymbol{x}]_{t=T}+2(\boldsymbol{p}-\tilde{\boldsymbol{p}})^{\mathrm{tr}}\boldsymbol{K}\,\delta\boldsymbol{p} + \int_0^T\left\{\left[\left(\frac{\partial H}{\partial x}\right)^{\mathrm{tr}}-\left(\frac{\mathrm{d}\boldsymbol{a}(t)}{\mathrm{d}t}\right)^{\mathrm{tr}}\right]\delta\boldsymbol{x} + \left(\frac{\partial H}{\partial p}\right)^{\mathrm{tr}}\delta\boldsymbol{p}\right\}\mathrm{d}t. \tag{7}$$

By introducing the adjoint system:

$$\frac{\mathrm{d}\boldsymbol{a}(t)}{\mathrm{d}t}=\frac{\partial H}{\partial\boldsymbol{x}}=-\left(\frac{\partial\boldsymbol{M}}{\partial\boldsymbol{x}}\right)^{\mathrm{tr}}\boldsymbol{a}(t)+\nabla_x D, \tag{8}$$

$$\boldsymbol{a}(T)=0, \tag{9}$$

and noting $J=J'$, we can obtain the gradient of J with respect to $\boldsymbol{p}$:

$$\frac{\partial J}{\partial\boldsymbol{p}}=2\boldsymbol{K}(\boldsymbol{p}-\tilde{\boldsymbol{p}})+\int_0^T\left(\frac{\partial H}{\partial\boldsymbol{p}}\right)\mathrm{d}t = 2\boldsymbol{K}(\boldsymbol{p}-\tilde{\boldsymbol{p}})-\int_0^T\boldsymbol{a}(t)^{\mathrm{tr}}\frac{\partial\boldsymbol{M}}{\partial\boldsymbol{p}}\,\mathrm{d}t. \tag{10}$$

In summary, the general computational procedure for parameter estimation using the adjoint method is:

(i) specify an initial guess of the unknown parameters $\boldsymbol{p}$ and the size-control vector $\tilde{\boldsymbol{p}}$;

(ii) integrate the numerical model from $t = 0$ to $t = T$;

(iii) compute D, and thus J, within $[0, T]$ using the forward model output;

(iv) compute the gradient of J with respect to $\boldsymbol{p}$ by integrating the adjoint system backward within $[0, T]$;

(v) use a descent-direction algorithm to obtain a new estimate of $\boldsymbol{p}$ such that J is reduced toward the minimum;

(vi) if the convergence criteria are satisfied, terminate the procedure and output the estimate of $\boldsymbol{p}$ which is supposed to be optimal; otherwise use the new estimate of $\boldsymbol{p}$ as the initial guess and return to (ii) to start another iteration.

In Section 3, a one-dimensional shallow-water model is used to illustrate the feasibility of the parameter-estimation procedure applied to the Newtonian nudging coefficients used in analysis-nudging FDDA. In this parameter-estimation problem, the functional to be minimized involves the first-guess of the parameters, the penalty coefficient matrix $\boldsymbol{K}$, and the size-control vector $\tilde{\boldsymbol{p}}$. By assuming that the first guess of the parameters is equal to the size-control vector, we investigate the sensitivity of the parameter estimation to the trade-off between the measure of fit of the model and the size-control constraint in the cost function.

3. Determination of nudging coefficients in a one-dimensional shallow-water model

The governing equations of the one-dimensional shallow-water model with analysis nudging are written in "flux form" as follows

$$\frac{\partial(\phi u)}{\partial t} = -\frac{\partial}{\partial x}(\phi u u) - \phi\frac{\partial \phi}{\partial x} + f\phi v + \phi G_{\mathrm{u}}(u_{\mathrm{obs}} - u) + u G_{\phi}(\phi_{\mathrm{obs}} - \phi), \tag{11}$$

$$\frac{\partial(\phi v)}{\partial t} = -\frac{\partial}{\partial x}(\phi u v) - f\phi u + \phi G_{\mathrm{v}}(v_{\mathrm{obs}} - v) + v G_{\phi}(\phi_{\mathrm{obs}} - \phi), \tag{12}$$

$$\frac{\partial \phi}{\partial t} = -\frac{\partial}{\partial x}(\phi u) + G_{\phi}(\phi_{\mathrm{obs}} - \phi), \tag{13}$$

where $\phi = gh$, where g is the gravitational acceleration and $h =$ depth of fluid, f is the Coriolis parameter and G_{u} (G_{v} or G_{ϕ}) is the spatially varying analysis-nudging coefficient for u (v or ϕ). All variables are assumed to be independent of y. Although this model is simple, it can be used to simulate various phenomena in the atmosphere from long-wave propagation to hydraulic jumps.

The model is discretized in a domain that has 121 grid points with $\Delta x = 60$ km. Cyclic boundary conditions are used in space. The finite differencing and the staggered horizontal grid (Arakawa B) are the same as those in the Pennsylvania State University/National Center for Atmospheric Research (PSU/NCAR) mesoscale model (Anthes et al., 1987). A leap-frog scheme is used for the temporal integration with a time step of 6 min, and the time interval over which the data-assimilation experiments are performed is 12 h. The adjoint of

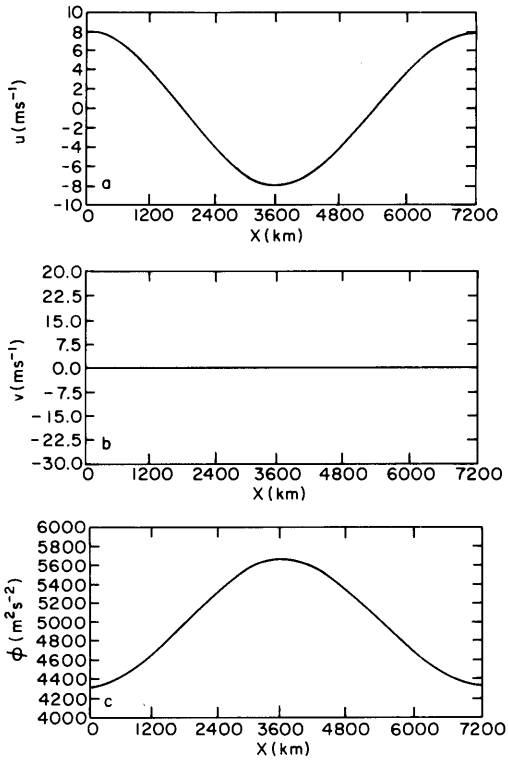

Fig. 1. The perfect initial conditions (based on Parrett and Cullen, 1984) used to generate the reference atmosphere and the OSSE "observations" used for assimilation. (a) u-component ($\mathrm{m\,s^{-1}}$), (b) v-component ($\mathrm{m\,s^{-1}}$) and (c) geopotential ($\mathrm{m^2\,s^{-2}}$).

the variational model and the boundary conditions for the adjoint variables are derived directly from the discrete form of the governing equations of the model using Lagrange multipliers (Thacker and Long, 1988), and are verified using the procedure suggested in Sun et al. (1991).

The discrete form of (1), the functional J to be minimized, is specified as follows:

$$
\begin{aligned}
J = K \sum_{i=1}^{I} & [(G_u^i - \tilde{G}_u^i)^2 \\
& + (G_v^i - \tilde{G}_v^i)^2 + (G_\phi^i - \tilde{G}_\phi^i)^2] \\
& + \sum_{i=1}^{I} \sum_{n=1}^{N} [W_u^{i,n}(u^{i,n} - u_{\text{obs}}^{i,n})^2 \\
& + W_v^{i,n}(v^{i,n} - v_{\text{obs}}^{i,n})^2 \\
& + W_\phi^{i,n}(\phi^{i,n} - \phi_{\text{obs}}^{i,n})^2],
\end{aligned}
\tag{14}
$$

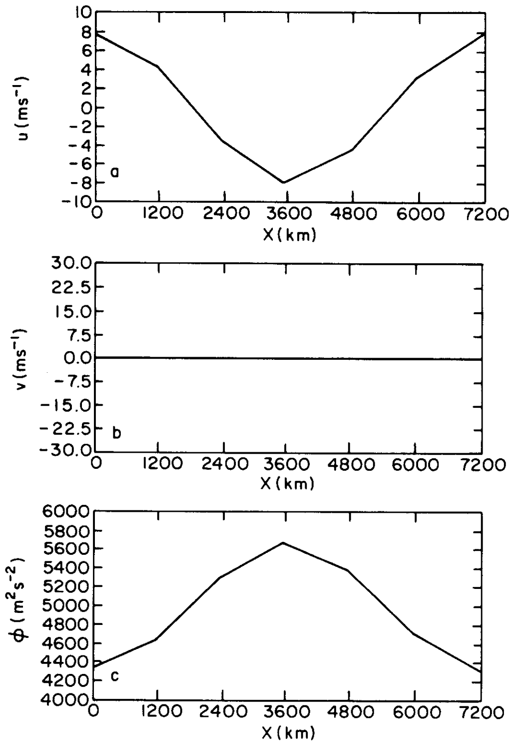

Fig. 2. The objectively analyzed (degraded) initial conditions used for all experiments. Observations from the perfect reference atmosphere were made available every 20 grid points (1200 km), and a linear interpolation was used between observations. (a) u-component (m s^{-1}), (b) v-component (m s^{-1}) and (c) geopotential (m^2 s^{-2}).

where W_u, W_v and W_ϕ are weighting coefficients reflecting the quality and distribution of the data, the superscripts, i and n, represent grid point i and time step n, and subscript "obs" denotes observations.

Experiments are conducted for two flow regimes: non-rotational flow ($f = 0$) and rotational flow ($f = 1.0 \times 10^{-4}$ s^{-1}). Fig. 1 shows the perfect initial conditions of u, v and ϕ (based on Parrett and Cullen, 1984) used to generate the "observations" throughout the 12-h assimilation periods for both flow regimes. The initial conditions for all experiments were generated by objective analysis (linear interpolation) of these perfect observations which were made available every 20 grid points (Fig. 2). For simplicity, the observations to be assimilated were assumed perfect and made available at every grid point and time step. Thus W_u, W_v and W_ϕ in (14) are set to unity. The role of the analysis nudging is simply to offset the imperfect model initialization. Although we consider the use of FDDA here to be for dynamic analysis, the results are also applicable to the dynamic-initialization problem.

As mentioned in Section 2, the adjoint of the first variation of the model is integrated as an efficient approach for obtaining the gradient of a cost function with respect to the control variable. Because the nudging coefficients are at least 5 to 7 orders of magnitude smaller than the corresponding model variables, the Newtonian nudging coefficients are first scaled by a large constant factor, $S_G = 10^8$, in order to make the minimization problem well-conditioned. For example, G_u in (14) is replaced by $G_u^* = G_u S_G$. The gradient of the discrete form of J' (3) with respect to the scaled nudging coefficients G_u^*, G_v^* and G_ϕ^* is computed, and then a limited-memory quasi-Newton method (Gill et al., 1981) is applied using the NAG© FORTRAN Library routine, E04DGF, to perform the minimization of (14). Note that the nudging terms in the equations defining both the forward model and corresponding adjoint system still use $G_u = G_u^* S_G^{-1}$, for example, to preserve the relative magnitude of the artificial nudging terms to the other physical processes in the governing equations.

Fig. 3 shows the model-simulated u, v and ϕ for Exp. NCTL (see Table 1) after 12 h of simulation for the *non-rotational* ($f = 0$) *case without nudging* ($\boldsymbol{G} = 0$). Due to the lack of horizontal diffusion

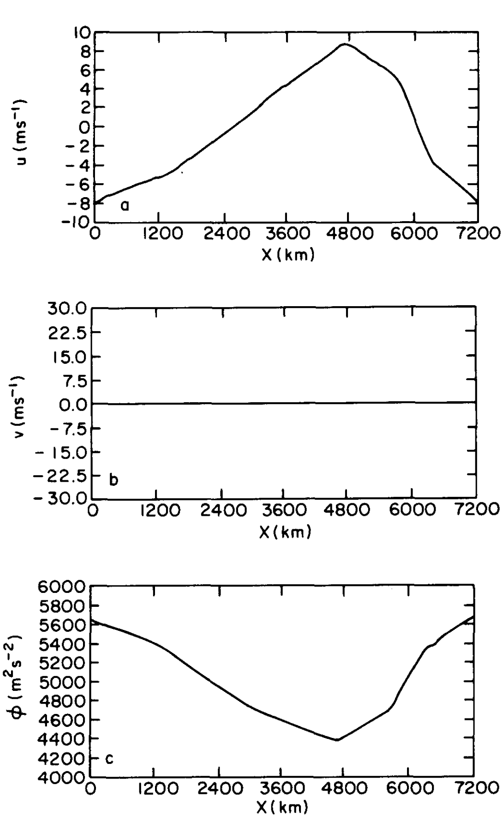

Fig. 3. The no-nudging ($G = 0$) control simulation (Exp. NCTL) for the non-rotational ($f = 0$) case at 12 h. (a) u-component (m s^{-1}), v-component (m s^{-1}) and (c) geopotential (m^2 s^{-2}).

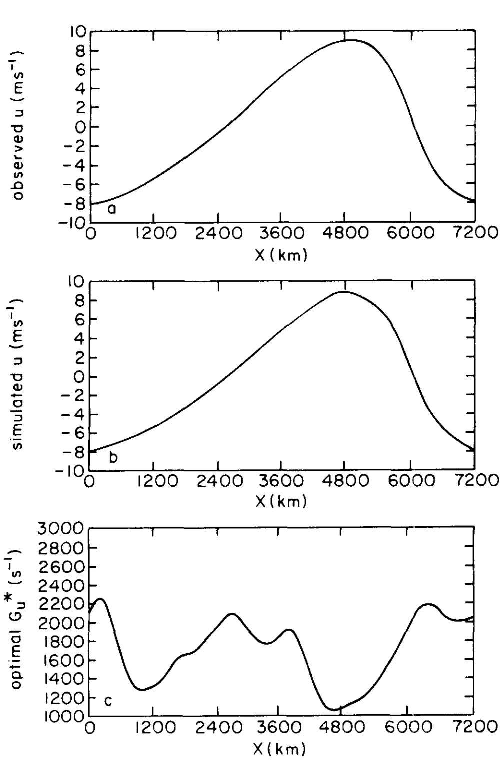

Fig. 4. Optimal nudging results for u-component of the wind for the non-rotational case (Exp. N3, Table 1). (a) Observed (reference atmosphere) u-component (m s^{-1}) at 12 h, (b) model-simulated u-component (m s^{-1}) at 12 h, and (c) scaled optimal nudging weights (s^{-1}) for u-component. (Note: $G_u^* = 1000$ s^{-1} corresponds to $G_u = 10^{-5}$ s^{-1}.)

in the model (11–13), this no-nudging control simulation of a developing hydraulic jump still shows local extrema in u and ϕ related to the observation separation in the objectively analyzed initial state (Fig. 2). Since $v = 0$ initially and $f = 0$, the tendency for v in (12) is also zero; thus, v remains zero throughout the simulation and is not shown for any of the non-rotational optimal-nudging experiments.

Fig. 4 presents the "observed" u-component of the wind for this non-rotational case at 12 h, the model-simulated value obtained using analysis nudging throughout the 12-h assimilation period (Exp. N3 in Table 1) and the corresponding optimally determined nudging coefficients G_u which vary in space across the grid. Fig. 5 is the same as Fig. 4, except for ϕ. The nudged u and ϕ fields at the end of the 12-h assimilation period agree very well with the observed fields (compare with Exp. NCTL in Fig. 3). In this case, the spreading of information is accomplished by advection and pure gravity-wave adjustment. The nudging weights vary by about a factor of two across the domain, and their extrema also appear to be related to the observation separation in the objectively analyzed initial conditions. After all, in this demonstration the nudging is simply compensating for the degraded initialization.

As mentioned earlier, the cost function J (14) represents a trade-off between reducing the model error and constraining the size of the optimal nudging coefficients. The size constraint is important because nudging coefficients that are too large (a) disrupt model adjustment processes that

Table 1. *Sensitivity of the optimal nudging weights to the coefficient K in the penalty term of the cost function for the non-rotational case*

			$u^{(1)}$			$\phi^{(2)}$		
EXP	K	Itn	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f ($m\,s^{-1}$)	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f ($m^2\,s^{-2}$)
NCTL	—	—	—	—	0.60	—	—	44.8
N1	10^0	3	1012	6	0.39	1014	5	28.9
N2	10^{-1}	8	1110	59	0.37	1132	47	27.5
N3	10^{-2}	18	1702	352	0.26	1907	323	20.3
N4	10^{-3}	65	3053	988	0.17	4306	1233	12.0
N5	10^{-4}	2772	3841	4702	0.16	8242	5410	10.0

The number of iterations required for convergence of the minimization algorithm is given by Itn. The nudging weights G^* are scaled by $S_G = 10^8$ such that $G = S_G^{-1} G^*$, and $G^* = 1000\,s^{-1}$ corresponds to $G = 10^{-5}\,s^{-1}$. The domain mean and standard deviation of the scaled G for each variable are denoted by $\bar{G}^*$ and σ_{G^*}, respectively. The initial and final rms errors for each variable are given by rms_i and rms_f, respectively.

(1) $rms_i = 0.54\ m\,s^{-1}$.

(2) $rms_i = 45.7\ m^2\,s^{-2}$.

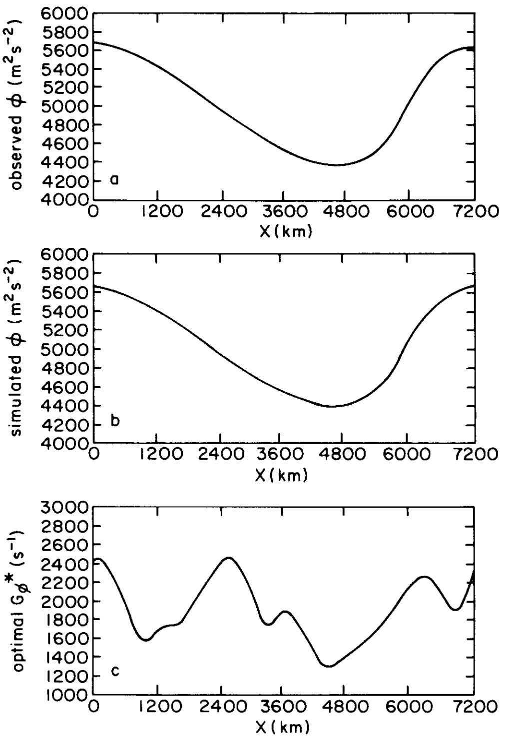

Fig. 5. Optimal nudging results for geopotential for the non-rotational case (Exp. N3, Table 1). (a) Observed (reference atmosphere) geopotential ($m^2\,s^{-2}$) at 12 h, (b) model-simulated geopotential ($m^2\,s^{-2}$) at 12 h, and (c) scaled optimal nudging weights (s^{-1}) for geopotential. (Note: $G_\phi^* = 1000\,s^{-1}$ corresponds to $G_\phi = 10^{-5}\,s^{-1}$.)

spread observational information throughout the model domain, (b) cause the artificial nudging effects to overwhelm the real physical effects in the governing model equations, and (c) can lead to numerical instabilities. (For numerical stability in this model, $G \leqslant 1/\Delta t$ where Δt is the model time step in seconds.) Thus the minimization of J depends on the penalty term coefficient K and the size-control vector $\tilde{G}$. The main focus of this paper is to investigate the sensitivity of the parameter-estimation technique to these two factors comprising the penalty term.

Table 1 shows the sensitivity of the optimal nudging weights in the non-rotational flow regime to values of K ranging from 10^0 to 10^{-4}. The scaled value of the size-control vector $\tilde{G}^*$ and the initial-guess of the nudging coefficients is set to $1000\,s^{-1}$. The initial root-mean-square (rms) errors in u and ϕ for all experiments are $0.54\,ms^{-1}$ and $45.7\,m^2\,s^{-2}$, respectively. Note that the no-nudging control simulation (Exp. NCTL) has rms wind and geopotential errors of $0.60\,ms^{-1}$ and $44.8\,m^2\,s^{-2}$, respectively, after 12 h of model integration. When the penalty coefficient $K = 10^0$ the mean scaled optimal-nudging coefficients over the domain for u and ϕ are $1012\,s^{-1}$ and $1014\,s^{-1}$, respectively. The corresponding standard deviations are $6\,s^{-1}$ and $5\,s^{-1}$, and the final rms errors at the end of the 12-h assimilation period have been reduced to $0.39\,ms^{-1}$ and $28.9\,m^2\,s^{-2}$. As K decreases, the number of iterations for convergence of the minimization algorithm increases,

Table 2. *Same as Table 1, except for the rotational case*

			$u^{(1)}$			$v^{(2)}$			$\phi^{(3)}$		
EXP	K	Itn	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f $(m\ s^{-1})$	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f $(m\ s^{-1})$	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f $(m^2\ s^{-2})$
RCTL	—	—	—	—	0.44	—	—	0.43	—	—	41.8
R1	10^0	3	1003	2	0.28	1003	3	0.28	1012	7	27.0
R2	10^{-1}	8	1028	23	0.28	1028	25	0.27	1112	65	26.2
R3	10^{-2}	17	1234	165	0.25	1205	165	0.23	1791	431	21.2
R4	10^{-3}	43	2005	598	0.17	1833	534	0.14	4060	1507	13.3
R5	10^{-4}	1114	2508	3502	0.13	3344	1133	0.01	8205	5505	9.5

(1) $rms_i = 0.54$ m s^{-1}.
(2) $rms_i = 0.00$ m s^{-1}.
(3) $rms_i = 45.7$ m^2 s^{-2}.

and both the mean and standard deviations of the nudging coefficients increase. The final rms errors decrease as K decreases and the size-constraint on the nudging weights is reduced. Thus when the nudging weights are allowed to increase, the model fit to the data improves. For reference, the results that were presented in Figs. 4 and 5 were for the median value of $K = 10^{-2}$ (Table 1, Exp. N3). For $K = 10^{-4}$, the mean nudging coefficients in Exp. N5 are about four ($\bar{G}_u^*$) and eight ($\bar{G}_\phi^*$) times larger than those in Exp. N1 ($K = 10^0$). However, the standard deviation of G_u^* in Exp. N5 exceeds the mean, which indicates that over some portions of the domain, the optimal nudging weights for u have become *negative* because there is no explicit constraint on positiveness. It is especially interesting to note that the ratio of $\bar{G}_\phi^*/\bar{G}_u^*$ increases as the size-constraint on the nudging weights (K) is reduced. For $K = 10^0$, this ratio is close to one, but for $K = 10^{-4}$, this ratio exceeds two.

For the rotational case, Fig. 6 shows the observed and Exp. R3 (Table 2) simulated geopotential field at the end of the 12-h analysis-nudging period, and the corresponding optimal nudging weights. Again, the experiment using the median value of $K = 10^{-2}$ was chosen from Table 2 for plotting. The initial conditions for all rotational experiments were the same as those used in the non-rotational experiments (Fig. 2), and the observations were similarly generated using the model with non-zero Coriolis and the perfect initial state shown in Fig. 1. Comparison of Figs. 5 and 6 shows the effect of Coriolis effects on the optimal nudging of geopotential. Although the locations of the extrema for G_ϕ are similar between

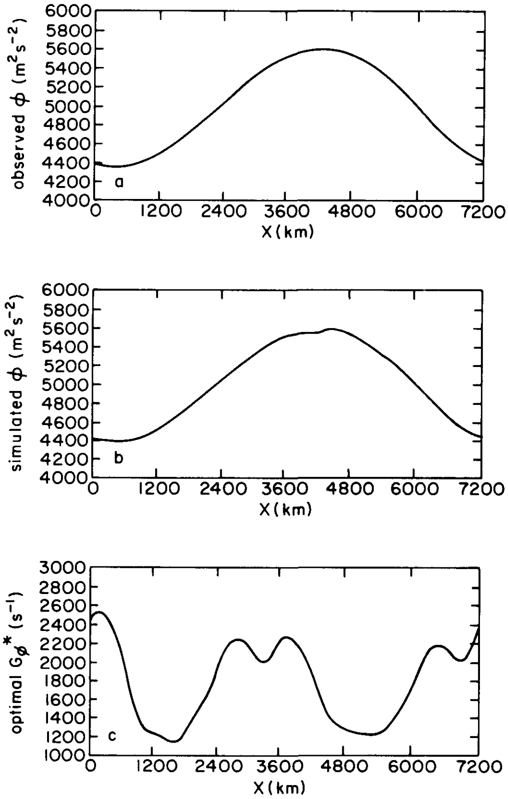

Fig. 6. Optimal nudging results for geopotential for the rotational case (Exp. R3, Table 2). (a) Observed (reference atmosphere) geopotential (m^2 s^{-2}) at 12 h, (b) model-simulated geopotential (m^2 s^{-2}) at 12 h, and (c) scaled optimal nudging weights (s^{-1}) for geopotential. (Note: $G_\phi^* = 1000$ s^{-1} corresponds to $G_\phi = 10^{-5}$ s^{-1}.)

the two cases, the domain mean value is greater in the non-rotational case while its spatial variation is greater in the rotational case. Tables 1 and 2 show this to be true for all values of K, and for the wind, both the mean and standard deviation of the optimal nudging coefficients are greater in the non-rotational case for all K. In the rotational case, model adjustments involve inertia-gravity modes rather than pure gravity modes, but the conclusions are the same as those for the non-rotational case: the mean and standard deviation of the optimal nudging weights increase with decreasing K, and the model fit to the data is closer for smaller values of K. As in the non-rotational case, when $K = 10^{-4}$, the nudging coefficients for u become negative, and the mean scaled nudging weight for geopotential is two to three times that for the u- and v-components of the wind.

The sensitivity of the optimal-nudging technique to the size-control vector in the penalty function is addressed as follows: the optimal nudging scheme is used to refine the specification of the size-control vector by performing successive minimizations of the cost function. The optimal nudging weights from Exp. N3 are used to define the size-control vector for a subsequent minimization, from which its optimal nudging weights are then used to define the size-control vector for the next minimization step. This process is repeated for six minimization steps (Table 3). The mean and standard deviations of the nudging weights increase with each successive minimization of the cost function. The number of iterations required for convergence within each minimization step did not vary significantly from one minimization step to the next (not shown). After 6 minimization steps, the rms errors for u and ϕ have been reduced by about one-third. Interestingly, this result is similar to that obtained in Table 1 when reducing K from 10^{-2} (Exp. N3) to 10^{-3} (Exp. N4). The mean and standard deviations of the nudging coefficients after six minimization steps (Table 3) are also similar to those for Exp. N4 (Table 1).

Table 4 is the same as Table 3 but for the rotational case using Exp. R3. Again, the mean and standard deviations of the nudging weights increase with each successive minimization step, and the results after six minimizations of the cost function are very similar to those of Exp. R4 in Table 2. Thus it appears that similar results may be obtained by performing a single minimization with a reduced value of K. This may suggest that the uncertainty in specifying the size-control vector can be transferred to the specification of the penalty coefficient K. If this is true, is spatial variability within the size-control vector important, and can it be used to represent confidence in the analyzed data? For example, the analysis-nudging weights should be smaller in magnitude over the ocean where there are fewer data, or in

Table 3. *Sensitivity of the optimal nudging weights to the size-control vector $\tilde{G}$ in the penalty term of the cost function for the non-rotational case*

EXP. N3	$u^{(1)}$			$\phi^{(2)}$		
$N_{\min}$	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f ($m\ s^{-1}$)	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f ($m^2\ s^{-2}$)
1	1702	352	0.26	1907	323	20.3
2	2187	585	0.21	2580	566	16.0
3	2551	754	0.18	3115	762	13.6
4	2833	880	0.17	3555	923	12.6
5	3054	976	0.17	3922	1054	12.5
6	3224	1047	0.18	4231	1163	13.0

The results after one minimization step, $N_{\min} = 1$ are those of Exp. N3 ($K = 10^{-2}$), in which the scaled size-control vector $\tilde{G}^* = 1000\ s^{-1}$, which corresponds to $\tilde{G} = 10^{-5}\ s^{-1}$. For each subsequent minimization step and determination of optimal nudging weights, the size-control vector is defined to be the optimal nudging weights from the previous minimization step. The domain mean and standard deviation of the scaled G for each variable are denoted by $\bar{G}^*$ and σ_{G^*}, respectively. The initial and final rms errors for each variable are given by rms_i and rms_f, respectively.

$^{(1)}$ $rms_i = 0.54\ m\ s^{-1}$.

$^{(2)}$ $rms_i = 45.7\ m^2\ s^{-2}$.

Table 4. *Same as Table 3, except with Exp. R3 for the rotational case*

EXP. R3	$u^{(1)}$			$v^{(2)}$			$\phi^{(3)}$		
N_{min}	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f (m s^{-1})	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f (m s^{-1})	$\bar{G}^*$ (s^{-1})	σ_{G^*} (s^{-1})	rms_f (m^2 s^{-2})
1	1234	165	0.25	1205	165	0.23	1791	431	21.2
2	1431	290	0.22	1362	282	0.20	2392	742	17.9
3	1599	391	0.20	1491	371	0.18	2880	986	15.9
4	1742	472	0.19	1601	443	0.16	3291	1187	14.7
5	1864	539	0.18	1697	503	0.15	3644	1355	14.1
6	1966	593	0.18	1783	555	0.14	3953	1499	13.8

$^{(1)}$ $rms_i = 0.54$ m s^{-1}.
$^{(2)}$ $rms_i = 0.00$ m s^{-1}.
$^{(3)}$ $rms_i = 45.7$ m^2 s^{-2}.

areas where the data may not be representative. This information can be included in the optimal nudging framework via a nonhomogeneous size-control vector, but it is not clear how an appropriate value for K will be determined.

4. Summary and recommendations for future research

In this study, a general approach has been developed for optimal determination of model parameters applied to a one-dimensional shallow-water model in the context of analysis-nudging four-dimensional data assimilation. The purpose of this application is to overcome the uncertainty in the traditional method of specifying the nudging coefficients. The cost function to be minimized is composed of two terms: one is to measure the discrepancy between the model state and the corresponding observations, and the other is a scale-constraint on the optimal nudging weights.

The adjoint equations represent an efficient way to calculate the gradient of the cost function with respect to the control variable. In this application, the control variable is the analysis-nudging weights defined over a uniform model grid. That is, the model initial state is *not* included in the control-variable vector. The adjoint system is the same as that used in the variational four-dimensional analysis (VFDA) method where the control variable is the initial state, except that the nudging terms have been added to both the forward model and the adjoint system. Numerical experiments are conducted for both non-rotational and rotational flow regimes using model-generated (OSSE) observations to test the feasibility of this parameter estimation technique using the adjoint method. The main focus of this paper is to investigate the sensitivity of the optimal-nudging weights to the penalty coefficient matrix $\boldsymbol{K}$ and the size-control vector $\tilde{\boldsymbol{G}}$ comprising the penalty term added to the cost function as a size-constraint on the optimal-nudging weights. This constraint is of general importance because nudging coefficients which are too large (a) disrupt model adjustment processes that spread observational information throughout the model domain, (b) cause the artificial nudging effects to overwhelm the real physical effects in the governing model equations, and (c) can lead to numerical instabilities.

It is shown that this approach is able to obtain a solution for optimal-nudging coefficients that vary in space and minimize the cost function. Because the experimental design uses analysis nudging to offset an imperfect model initialization, there is strong correlation between the spatial variability of the nudging coefficients in space and the observation separation in the model initial conditions. It is found that the weaker the size constraint in the cost function, the larger the domain mean and standard deviations of the nudging weights. Correspondingly, with decreasing K, the model fit to the data becomes closer, and the number of iterations required for convergence of the minimization algorithm increases. It is also shown that the sensitivity of the technique to the size-control vector may be passed to the specification of K in these cases. Thus further research is needed to determine if the size-control vector can be used effectively to represent confidence in the data analysis in terms of data distribution, representativeness, etc. Although the technique works, there is this *new* uncertainty K which has been introduced into the nudging framework, and proper specification of K under more realistic conditions is definitely an area for future research.

Another very important issue for this optimal-nudging approach is how to best measure success; that is, what makes a "good analysis"? In this paper, the cost function was defined by a penalty term plus a measure of the fit of the model to the data during the assimilation period. In general, improving the fit to the data does not guarantee a better analysis by other subjective and objective criteria (Andrew Lorenc, personal communication). Also, confidence in the analysis would be greater if this "fit" is measured to independent data (data not directly assimilated into the model). Furthermore, if your definition of a good analysis is one which produces an improved forecast, determination of the optimal nudging weights could be made based on the model fit to the data during a subsequent forecast (Lorenc et al., 1991). Regardless of how you define the cost function to measure "goodness of fit", the main point of this paper is that there is a sensitivity to the penalty term which is needed to control the size of the nudging weights. Any nudging method must limit the size of these weights for the reasons mentioned above, and failure to do so violates the method's most fundamental underlying assumptions.

As mentioned earlier, in this application of optimal nudging, the control variable was defined as the vector of optimal-nudging weights, and did not include the model initial state as done in the VFDA method. The VFDA method has great theoretical attractions because it can make good use of the knowledge embodied in the model forecast equations, but at a considerable computational cost. The VFDA technique allows variables which are not directly observed to be altered to improve the fit to parameters that are observed. Lorenc (1988), using the identical shallow-water model example in this paper, showed that in the non-rotational case, observations of v could be used to improve the analysis of u and ϕ. Successive v fields provide information about the movement of a tracer, and hence about u. The nudging methods, "optimal" or otherwise, cannot do this in this case. Since $f = 0$, the alterations made to v by the G_v term remain uncoupled from u and ϕ. So the optimal-nudging method, which has much of the computational cost of VFDA, cannot improve the v fields in this non-rotational example. However, it must be remembered that the VFDA technique assumes the model is perfect, while optimal nudging does not by introducing an artificial non-physical term into the model equations. Thus there may be situations where it may be beneficial to combine the two variational approaches by defining the control variable as both the initial state and the nudging coefficients. Nevertheless, the VFDA approach is still the most attractive alternative since it does not introduce any artificial terms into the governing model equations.

The next logical question is the following: although a very simple model with idealized data was used here, how general are these conclusions for more realistic models and data sets? Although the sensitivity of the optimal-nudging approach to the choice of cost function is certainly fundamental and requires further research using more realistic models, the additional sensitivity of the technique to the penalty term which controls the size of the nudging weights is of paramount importance since it must preserve the method's underlying assumption that the nudging terms play a secondary role in the model's governing equations. Otherwise, the advantage of using the model will be greatly diminished. This is the first paper to address this very important issue and demonstrates that this should be a concern when using this approach in more realistic conditions.

Also, once the nudging coefficients have been determined for one particular period by this technique, to what extent can they be used again for another period? It is our hope that this technique may be used to gain basic insight into how to best define the nudging weights, say, in the vertical, or near a coastline, and then use this information in the standard nudging approach. In many circumstances, however, the nudging weights must depend on the dynamics, as well as the data quality, density and frequency (Stauffer and Seaman, 1993). Thus, the nudging weights would not be readily applicable to other time periods when any of these factors have changed substantially. Furthermore, the length of the assimilation period becomes especially important when using a technique which assumes the nudging weights do not vary in time.

Finally, this approach currently uses *analysis nudging* where the observed data are defined and assimilated over a regularly spaced grid. For data assimilation on the mesoscale, a nudging approach which uses asynoptic observational data directly has been shown to be very effective by using real asynoptic, mesobeta-scale data sets in the PSU/NCAR mesoscale model (Stauffer and Seaman, 1991; Stauffer et al., 1993). It is becoming increasingly clear that effective assimilation of these modern asynoptic data sets (e.g., profilers, acoustic sounders, etc.) requires an approach that uses the observational data directly for continuous assimilation rather than as gridded analyses based on observation at a fixed time. Thus a nudging-to-observations ("obs nudging") technique is especially attractive for assimilation of these data sets, and an optimal control method designed to help specify the spatial and temporal influence that each *observation* has in the model atmosphere would prove to be invaluable.

5. Acknowledgements

This research was supported by NSF Grant No. ATM-9116176, with additional support from NASA Grant No. NGT30040, NSF/FAA Grant No. ATM-9203317 and ONT Contract No. N00039-92-C-0100. The authors wish to thank Andrew Lorenc and one anonymous reviewer whose insightful remarks helped to improve the paper.

REFERENCES

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

Anthes, R. A., Hsie, E.-Y. and Kuo, Y.-H. 1987. Description of the Penn State/NCAR mesoscale model version 4 (MM4). NCAR Tech. Note NCAR/TN-282 + STR, 66 pp. [Available from the National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO, USA 80307].

Gill, P. E., Murray, W. and Wright, M. H. 1981. *Practical optimization*, 401 pp., Academic Press.

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

Lorenc, A. C. 1988. Optimal nonlinear objective analysis. *Quart. J. Roy. Meteor. Soc.* **114**, 205–240.

Lorenc, A. C., Bell, R. S. and Macpherson, B. 1991. The meteorological office analysis correction data assimilation scheme. *Quart. J. Roy. Meteor. Soc.* **117**, 59–89.

Parrett, C. A. and Cullen, M. J. P. 1984. Simulation of hydraulic jump in the presence of rotation and mountains. *Quart. J. Roy. Meteor. Soc.* **110**, 147–165.

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. Wea. Rev.* **118**, 1250–1277.

Stauffer, D. R. and Seaman, N. L. 1991. On mesobeta-scale four-dimensional data assimilation. Preprints, *Ninth Conference on Numerical Weather Prediction*, AMS, Denver, CO USA October 14–18, 1991, 595–598.

Stauffer, D. R., Seaman, N. L. and Binkowski, F. S. 1991. Use of four-dimensional data assimilation in a limited-area mesoscale model. Part II: Effects of data assimilation within the planetary boundary layer. *Mon. Wea. Rev.* **119**, 734–754.

Stauffer, D. R. and Seaman, N. L. 1993. On multi-scale four-dimensional data assimilation. *J. Appl. Meteor.*, in press.

Stauffer, D. R., Seaman, N. L., Warner, T. T. and Lario, A. M. 1993. Application of an atmospheric simulation model to diagnose air-pollution transport in the Grand Canyon region of Arizona. *Chemical Engineering Communications* **121**, 9–26.

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

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

Zou, X., Navon, I. M. and Le Dimet, F. X. 1993. An optimal nudging data assimilation scheme using parameter estimation. *Quart. J. Roy. Meteor. Soc.*, in press.