Tellus (2002), 54A, 125–137
Printed in UK. All rights reserved

TELLUS
ISSN 0280–6495

Data assimilation with a barotropically unstable shallow water system using representer algorithms

By LIANG XU* and ROGER DALEY†, *Naval Research Laboratory, 7 Grace Hopper Avenue, Monterey, CA 93943-5502, USA*

(Manuscript received 14 May 2001; in final form 16 October 2001)

ABSTRACT

The *cycling representer algorithm* of Xu and Daley (2000) is a weak constraint four-dimensional variational data assimilation algorithm. It was successfully applied to a one-dimensional transport problem and was able to successfully extract the signal from noisy and sparse observations. The algorithm, however, has not previously been applied to a multivariate, multidimensional system with dynamic instability. The algorithm is also very computationally demanding and awaits considerable enhancement in computer power before being practical for operational forecast models. We have two objectives in this paper. The first is to apply the cycling representer algorithm to a two-dimensional, multivariate barotropically unstable linear shallow water system. The second objective is to formulate and test an *accelerated representer algorithm* that is much more computationally tractable than the cycling representer algorithm itself. A linear shallow water system with a barotropically unstable basic state was used as a test bed to conduct data assimilation experiments. The evolution of a 'neutral' eastward-propagating singular vector was selected as the 'truth', against which all data assimilation experiments were to be evaluated. The results indicated that the cycling representer algorithm was capable of providing satisfying state estimates for a multivariate, multidimensional system. The results from the accelerated representer algorithm were very encouraging because it is sufficiently computationally tractable to be used on present day multi-processor machines for operational applications.

1. Introduction

Operational meteorological centers around the world are either implementing or developing advanced four-dimensional data assimilation systems. With respect to the present operational three-dimensional variational (3DVAR) data assimilation algorithms, the advanced four-dimensional systems have several advantages. Firstly, the observations are assimilated at the correct observation times, rather than being time-binned in the intermittent 3DVAR systems. Secondly, it is possible to properly account for serial error correlations in the observations. Thirdly, and perhaps most significantly, the all-important background error covariance can become, at least partially, flow-dependent.

The advanced four-dimensional algorithms fall into two classes. The first class contains sequential algorithms based on the Kalman filter and contains many variants such as the ensemble Kalman filter, extended Kalman filter, Kalman smoother etc. (e.g. Menard, 1994; Evensen, 1997; Evensen and van Leeuwen, 2000). The second class contains the variational algorithms (e.g. Courtier et al., 1994; Zupanski, 1997; Bennett, 1997; Courtier, 1998; Xu and Daley, 2000), which are all involved in some kinds of approximations (sub-optimal

* Corresponding author.
e-mail: xu@nrlmry.navy.mil
† Deceased August 2001.

algorithms) of the generalized inverse problem. In the variational algorithms, one seeks to minimize a four-dimensional cost function, which measures the fit to the initial conditions, the observations, the model and spatial boundary conditions. In a strong-constraint minimization, the prediction model is assumed to be perfect, thus eliminating one huge term from the cost function and reducing the control variables to only the initial state in analysis grid-space. In more sophisticated weak constraint algorithms, one does not make the perfect model assumption.

As for the 3DVAR algorithms, the four-dimensional variational algorithms can be posed in one of two different spaces — analysis grid-space and observation-space. Traditionally, analysis grid-space has been the preferred choice for atmospheric four-dimensional variational algorithms, but there exists a complete theory for four-dimensional variational assimilation in an observation-space. This theory is known as representer theory and has been extensively developed and applied to oceanographic data assimilation problems by Bennett and McIntosh (1982) and Bennett and Thorburn (1992).

In all four-dimensional variational data assimilation algorithms, the user must specify the initial background and observation error covariances, the model error covariance (in the weak case) and perhaps boundary error covariance (in the limited area model case). Currently, it is almost impossible to specify optimal error covariances for operational applications. From that point of view, it is fair to state that all the existing practical four-dimensional data assimilation algorithms are sub-optimal. For different applications (e.g. oceanographic hindcast and atmospheric data assimilation), the impacts on the state estimations by mis-specifying the error covariances may be different. For example, the oceanographic representer algorithms generally perform minimizations over a single (long) time period (about one year) to provide state estimates and derived quantities, leading to increased understanding of ocean circulations. Because of the minimization over the long time period, the initial background error covariance only influences the early part of the assimilation period and has little influence over most of the period. Thus, for oceanographic problems the specification of the initial background error covariance is not very critical. The specification of the model error covariance, on the other hand, may play a very important role due to the long model integration for the oceanographic problems. Conversely, data assimilation in the atmosphere is concerned with providing initial state estimates for numerical weather prediction models to estimate future atmospheric states. The operational numerical weather prediction model requires very short atmospheric data assimilation update cycles (3–12 h). The success of the atmospheric data assimilation for a relatively short period of time depends critically on the accurate specification of the initial background error covariances and may not depend so critically on the accurate specification of the model error covariance.

Consequently, Xu and Daley (2000) extended the representer theory to produce the cycling representer algorithm. The cycling representer data assimilation algorithm of Xu and Daley (2000) is a weak constraint four-dimensional variational data assimilation algorithm, which provides an internally consistent estimate of the initial background error covariance at the beginning of each cycle, based on the results of the previous cycle (and implicitly all previous cycles). In Xu and Daley (2000), the cycling representer algorithm was applied to a one-dimensional transport problem and was able to successfully extract the signal from noisy and sparse observations. However, the algorithm is very computationally demanding and awaits considerable enhancement in computer power before being practical for operational numerical weather prediction models. Furthermore, we recently found that Section 5.2 of Xu and Daley (2000) was in error and realized that the operation count there was no better than the one by using a Kalman filter.

We have two objectives in this paper. The first objective is to apply the cycling representer algorithm to a more realistic and computationally demanding data assimilation problem than the study of Xu and Daley (2000). Specifically, the problem is a two-dimensional, multivariate barotropically unstable shallow water system. The second objective is to formulate and test an accelerated representer algorithm that is much more computationally tractable than the cycling representer algorithm. The accelerated representer algorithm could be implemented for operational numerical weather prediction models on the

present day massively parallel multi-processor computers.

This paper is organized as follows. In Section 2, we introduce three representer algorithms — explicit representer, cycling and accelerated representers, respectively. The governing equations of the linear shallow water system and the associated numerical prediction and adjoint models can also be found in Section 2. In Section 3, we describe the design of the data assimilation experiments. We introduce a jet-like basic state and discuss the singular vectors of this system. We then choose a 'truth' based on the evolution of a 'neutral' singular vector, and discuss the observation networks and the specifications of the initial state and the initial background error covariance. The results from a number of experiments are described in Section 4, and the summary and conclusions are given in Section 5.

2. The data assimilation system

The data assimilation system used in this study consists of three components. They are the data assimilation algorithms, the prediction and the adjoint models of a linear shallow water system. In this section, we first briefly describe the three data assimilation algorithms related to this study. We then give a brief description of the numerical prediction and adjoint models as well as the associated parameters.

2.1. The data assimilation algorithms

Two different representer algorithms were used in this study. They were the cycling representer algorithm (Xu and Daley, 2000), and the accelerated representer algorithm to be presented in this section. The two algorithms are based on the original explicit representer algorithm by Bennett and McIntosh (1982) and Bennett and Thorburn (1992). In the following, we first give a brief description of the explicit and cycling representer algorithms. We then describe the formulation of the accelerated representer algorithm.

2.1.1. The representer algorithm. The representer algorithm was introduced into oceanography by Bennett and McIntosh (1982) and Bennett and Thorburn (1992). This algorithm is an effective way to solve the generalized inverse problem — i.e. to variationally minimize a four-dimensional cost function that measures the fit of the analysis to the observations, the forecast model, the spatial boundaries and any existing initial state estimate (initial background field). The forecast model is imposed as a weak constraint in this algorithm. It was designed to perform a single minimization over a long time period, thus achieving a state estimate making use of all the observations available from the period.

2.1.2. The cycling representer algorithm. The data assimilation problem in meteorology is somewhat different, because of the necessity of producing new forecasts every few hours. Thus, in meteorological operational practice, the forecast model is run from the previous analysis to produce a background state estimate and then the observations are combined with the background state estimate using the data assimilation algorithm to produce the new analysis. Long forecasts can then be obtained from this analysis. This cycling process is then carried on ad infinitum.

Xu and Daley (2000) designed a cycling version of the representer algorithm and tested it on a one-dimensional constituent transport problem. The cycling representer algorithm carries out a series of four-dimensional minimizations over short time periods. However, this requires that the initial background error covariance (as well as the state estimates) be updated at the beginning of every cycle. The representer mechanism was used in Xu and Daley (2000) to update the initial background error covariance by calculating the representer at every point at the end in the previous cycle. There were two main components in the cycling representer algorithm of Xu and Daley (2000). The first component was the explicit calculation of the representer matrix. The computational expenses were proportional to the number of observations used in the calculations. The second component was the calculation of the analysis error covariance at one time level (usually the end) of each cycle. It was clear that the calculations of explicit representers and the analysis error covariance were very expensive. It is impossible for the algorithm to be used in operational four-dimensional data assimilation with the present day computers.

2.1.3. The accelerated representer algorithm. Because of the expense of the cycling representer algorithm, an accelerated representer algorithm has been developed for the operational four-dimensional data assimilation. In the accelerated representer algorithm, the representer matrix is calculated implicitly and the initial background error covariance at the beginning of each cycle is specified rather than being calculated. Notice that the initial background field is still updated (cycled) using the analysis field from the previous cycle. Without calculating the analysis error covariance at end of each cycle, the accelerated representer algorithm is much more computationally tractable. It should be noted that the computational efficiency is achieved by sacrificing the accuracy of the initial background error covariance at each cycle. It is clear that the accelerated representer algorithm becomes sub-optimal as compared with the cycling representer algorithm when less accurate initial background error covariance is available.

Unlike the original explicit representer algorithm, the representer matrix in the accelerated representer algorithm is calculated implicitly using an iterative method described by Egbert et al. (1994), Amodei (1995), Courtier (1998), where it is referred to as the 4D-PSAS algorithm, and Chau and Bennett (2001). The accelerated representer algorithm is the dual formulation of the standard four-dimensional variational (4DVAR) algorithm that has been implemented at the European Center for Medium Range Forecasts (ECMWF). That is, while the standard 4DVAR algorithm is an analysis grid-space algorithm, the accelerated representer algorithm is an observation-space algorithm.

The specification of the initial background error covariance at the beginning of each cycle will be similar to the specification of the initial background error covariance in the current 3DVAR algorithm, but will not be discussed in detail in this study. In the following we will give a brief description of the minimization procedure of the accelerated representer algorithm. The derivation follows directly from Section 2 of Xu and Daley (2000). Equation (2.15) of Xu and Daley (2000), which is used to find the coupling vector $\vec{\beta}$, is re-written here as

$$[\mathbf{R}_e + \mathbf{R}]\vec{\beta} = [\mathbf{y} - \mathbf{H}\mathbf{x}^b], \tag{1}$$

where $\mathbf{H}$ is the forward operator that connects the analysis grid-space with the observation-space, $\mathbf{R}_e$ (calculated) is the $K \times K$ representer matrix and K is the number of observations, $\mathbf{R}$ (specified) is the observation error covariance matrix, $\mathbf{y}$ is the observation vector of length K, and $\mathbf{y} - \mathbf{H}\mathbf{x}^b$ is the innovation vector (sometimes referred to as the observation increment). $\mathbf{x}^b$ is the background state vector of dimension I. The most computationally intensive task in eq. (1) is to obtain the representer matrix $\mathbf{R}_e$ explicitly. It is, in general, not very practical in meteorological operations to calculate $\mathbf{R}_e$ explicitly and then directly invert $\mathbf{R}_e + \mathbf{R}$ due to the large values of K (200,000 or more).

Fortunately, we can solve eq. (1) using iterative descent procedures, such as standard conjugate gradient methods or more efficient preconditioned conjugate gradient methods (see Golub and Van Loan, 1996 or Strang, 1986). In these methods, it is necessary to perform the following matrix vector multiply for each iteration:

$$\mathbf{q} = [\mathbf{R}_e + \mathbf{R}]\mathbf{p}. \tag{2}$$

Here $\mathbf{p}$ is a known vector of length K and $\mathbf{q}$ is a vector of length K, which is the result of the matrix/vector multiply. There are two operations in eq. (2),

$$\mathbf{q}^+ = \mathbf{R}\mathbf{p} \quad \text{and} \quad \mathbf{q}^* = \mathbf{R}_e\mathbf{p}, \quad \text{with} \quad \mathbf{q} = \mathbf{q}^+ + \mathbf{q}^*. \tag{3}$$

Since $\mathbf{R}$ is specified, the operation $\mathbf{R}\mathbf{p}$ is straightforward. The second operation is much more complex, but $\mathbf{q}^*$ can be obtained from $\mathbf{p}$, using a similar procedure to that used in the explicit calculation of the representer matrix.

In the procedure, we first define a vector $\vec{\phi}$, of dimension I, that satisfies the following adjoint equation,

$$\vec{\phi}_n - \mathbf{M}_n^{\mathrm{T}}\vec{\phi}_{n+1} = [\mathbf{H}^{\mathrm{T}}\mathbf{p}]_n, \tag{4}$$

where $1 \leqslant n \leqslant N$; $\mathbf{M}_n^{\mathrm{T}}$ is the adjoint model at timestep n; $[\mathbf{H}^{\mathrm{T}}\mathbf{p}]_n$ is the projection of vector $\mathbf{p}$ onto the analysis grid-space at timestep n; and the initial condition is $\vec{\phi}_{N+1} = 0$. Equation (4) is very similar to eq. (2.6) of Xu and Daley (2000). We can now integrate eq. (4) backward from $n = N$ to $n = 1$ to obtain the adjoint field $\vec{\phi}$.

We then define a vector $\vec{\theta}$, of dimension I, that

satisfies the following prediction equation,

$$\vec{\theta}_n - \mathbf{M}_{n=1}\vec{\theta}_{n-1} = \sum_{n'=1}^{N} \mathbf{Q}_{nn'}\vec{\phi}_{n'}, \tag{5}$$

where $1 \leqslant n \leqslant N$; $\mathbf{M}_{n-1}$ is the prediction (forward) model at timestep $n-1$ and $\mathbf{Q}_{nn'}$ is the model error covariance between timestep n and timestep n'. The initial condition of eq. (5) is $\vec{\theta}_0 = \mathbf{B}_0^{\mathrm{b}}\mathbf{M}_0^{\mathrm{T}}\vec{\phi}_1$ with $\mathbf{B}_0^{\mathrm{b}}$ as the initial background error covariance. Equation (5) is very similar to eqs. (2.7) and (2.8) of Xu and Daley (2000). Comparing eqs. (4) and (5) with eqs. (2.10)–(2.13) of Xu and Daley (2000), it can be shown that

$$\mathbf{R}_{\mathrm{e}}\mathbf{p} = \mathbf{H}\vec{\theta}, \tag{6}$$

which is just what is needed, at a cost of one backward adjoint and one forward tangent linear integrations. Once $\vec{\beta}$ is obtained using an iterative descent procedure, we can then use eqs. (2.6)–(2.8) of Xu and Daley (2000) to obtain the final analysis as follows.

We first integrate the adjoint variable $\vec{\lambda} = (\vec{\lambda}_1 \cdots \vec{\lambda}_n \cdots \vec{\lambda}_{N+1})^{\mathrm{T}}$ backward in time using

$$\vec{\lambda}_n - \mathbf{M}_n^{\mathrm{T}}\vec{\lambda}_{n+1} = [\mathbf{H}^{\mathrm{T}}\vec{\beta}]_n, \tag{7}$$

where $1 \leqslant n \leqslant N$ and $\vec{\lambda}_{N+1} = 0$. The analysis at any time during the cycle is given by

$$\mathbf{x}_n^{\mathrm{a}} - \mathbf{M}(\mathbf{x}_{n-1}^{\mathrm{a}}) = \sum_{n'=1}^{N} \mathbf{Q}_{nn'}\vec{\lambda}_{n'}, \tag{8}$$

where $1 \leqslant n \leqslant N$, subject to the initial condition $\mathbf{x}_0^{\mathrm{a}} = \mathbf{x}_0^{\mathrm{b}} + \mathbf{B}_0^{\mathrm{b}}\mathbf{M}_0^{\mathrm{T}}\vec{\lambda}_1$.

It should be noted that the accelerated representer algorithm becomes sub-optimal due to the need to specify the initial background error covariance at each cycle.

2.2. The prediction and adjoint models

The shallow water system is one of the simplest and most frequently used test beds for applications in meteorology. Besides its popularity in testing advanced numerics used in NWP models, it is also very popular in testing advanced atmospheric data assimilation algorithms (e.g. Courtier and Talagrand, 1990; Cohn and Parrish 1991; Todling and Ghil 1994). Despite the overall simplicity of the system (only involving interactions among the wind components and geopotential in a single layer), it displays some of the complex multivariate, multidimensional interactions that are commonly observed in more comprehensive meteorological systems, such as the primitive equation system.

We applied the representer algorithms of Section 2.1 to the linearized shallow water equations in a β-plane channel with rigid walls at the northern and southern boundaries and periodic boundary conditions at the eastern and western boundaries, respectively. (We will address the issues associated with non-linearity in a future study). With this linear system we were able to examine the multivariate aspects of the cycling representer algorithm and to compare the results with the ones obtained using the accelerated representer algorithm introduced in the previous section.

The governing equations of the linearized shallow water equations in Cartesian coordinates are as follows:

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x} - \left(f - \frac{\mathrm{d}U}{\mathrm{d}y}\right)v = 0, \tag{9}$$

$$\frac{\partial v}{\partial t} + U\frac{\partial v}{\partial x} + \frac{\partial \phi}{\partial y} + fu = 0, \tag{10}$$

$$\frac{\partial \phi}{\partial t} + \Phi\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{\mathrm{d}\Phi}{\mathrm{d}y}v = 0, \tag{11}$$

where x and y point eastward and northward, respectively; u and v are the eastward and northward perturbation wind components, respectively; and ϕ is the perturbation geopotential; $U = U(y)$ is a time invariant basic state flow trajectory and is in geostrophic balance with the basic state geopotential $\Phi = \Phi(y)$ and $\Phi_0 = \Phi(y_0)$; finally y_0 is the value of y at the center of the channel. The geostrophic balance in the basic state is given by $(\mathrm{d}\Phi/\mathrm{d}y) + fU\,0$, where the Coriolis parameter $f = f_0 + \beta(y - y_0)$ with $f_0 = f(y_0)$ and $\beta \equiv (\mathrm{d}f/\mathrm{d}y)_{y_0}$. Note that x, y, ϕ, and β should not be confused with $\mathbf{x}$, $\mathbf{y}$, $\vec{\phi}$, and $\vec{\beta}$ in Section 2.1, respectively. For simplicity, we refer perturbation wind and geopotential as wind and geopotential for the rest of the paper.

The model was discretized on an Arakawa C-grid. The numerical techniques used to construct this shallow water model were the same as those used in the US Navy's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS) except that we only had one vertical layer in this study. Readers who are interested in the numerical techniques used in COAMPS can

find detailed descriptions in Xu (1995) or Hodur (1997). The adjoint numerical model of the linear shallow water system was obtained directly from the discretized numerical prediction model. The model domain was a 3850 km by 3850 km channel with cyclic boundary conditions at the eastern and western boundaries and rigid wall condition ($v = 0$) at the northern and southern boundaries. Due to the use of Arakawa C-grid, we had 11×11 gridpoints for v and 11×10 gridpoints for, u and ϕ, respectively. A uniform grid-length of 385 km was used in both x and y directions. The timestep was 600 s. The β-plane was centered at 45°N, with $f_0 = 10^{-4}\,\mathrm{s}^{-1}$, $\beta = 1.5 \times 10^{-11}\,\mathrm{m}^{-1}\,\mathrm{s}^{-1}$ and $\Phi_0 = 3 \times 10^4\,\mathrm{m}^2\,\mathrm{s}^{-2}$.

3. The data assimilation experiments

One of the objectives of this study was to apply the cycling representer algorithm of Xu and Daley (2000) to a multidimensional, multivariate shallow water system with dynamic instability. In our preliminary experimentations that are not reported here, we successfully applied the cycling representer algorithm to simpler shallow water data assimilation problems, such as Poincaré and Kelvin wave propagation, where analytical solutions were always available. While those simple tests were very important to validate the algorithm, they were incapable of mimicking the dynamic instabilities observed in more realistic meteorological flows.

In order to ensure dynamic instability in our shallow water β-plane channel system, we employed a cosine square jet profile (described in Section 3.1) as the basic state trajectory in eqs. (9)–(11). It is generally preferable to have available the 'true' solution of the problem in order to create artificial 'observations' to be used in the data assimilation experiments and to create a metric of the success or failure of the experiments. However, for the unstable jet, unlike the Poincare and Kelvin waves problems, we did not have a ready analytical solution to be used as the 'truth' in this unstable system. Instead, we used singular vector analysis (see details in Section 3.2), to separate various singular vectors of the barotropically unstable shallow water system. One of the 'neutral' singular vectors was used as the initial condition for the prediction model. The evolution of the 'neutral' singular vector was then chosen to be the 'true' solution (see details in Section 3.3). By assuming the evolution of the 'neutral' singular vector as the 'true' solution, we implicitly used the perfect model assumption. For consistency, we used the perfect model assumption ($\mathbf{Q} = \mathbf{0}$) throughout this study despite the representer algorithms' ability to handle model errors.

The data sampling strategy is always an important issue in data assimilation. To examine the impact of data sampling on the assimilation results, we used two very different observation networks with the number of observations being the same. For each observation network, we conducted 2 experiments using cycling and accelerated representer algorithms. A total of 4 experiments were conducted. The time period for the data assimilation was 96 h. A total of 8 short cycles were used to cover the 96-h time period with each cycle consisting of 12 h.

3.1. *A barotropically unstable shallow water system*

As discussed above, we conducted our data assimilation experiments with a barotropically unstable shallow water system. In his classic 1949 paper, Kuo thoroughly studied the dynamic instability of a two-dimensional nondivergent flow in a barotropic atmosphere from a theoretical point of view. He concluded that barotropic instability might occur in a shallow water system in a beta-plane channel with certain basic state jet profiles. Todling and Ghil (1994) studied the data assimilation under unstable atmospheric conditions in a barotropically unstable shallow water system using a Kalman filter. Their barotropically unstable basic-state flow was a cosine-squared jet [different from the one used in Kuo (1949)]. We tested the jet profiles used by Kuo (1949) and Todling and Ghil (1994), respectively and found that both jet profiles were capable of generating the needed barotropic instability. For convenience, we used a jet profile similar to the one used in Todling and Ghil (1994). The jet profile is given by

$$U = \begin{cases} U_0 & \text{if } |y| > \delta \\ U_0[1 + \cos^2(2\pi y/\delta)] & \text{if } |y| < \delta' \end{cases}, \qquad (12)$$

where $U_0 = 20\,\mathrm{m\,s}^{-1}$ was the minimum zonal wind speed and $2\delta = 1155$ km was the width of

the jet. The maximum zonal wind speed was $2U_0$. The cosine square jet profile is schematically illustrated in Fig. 1.

3.2. The singular vectors of the barotropically unstable system

One of the advantages of using the simple shallow water system (with only a modest number of degrees of freedom) was the ability to explicitly calculate all the singular vectors of the system. We first obtained the forward propagator matrix, which generated the state at $t = 96$ h (576 timesteps) forecast from the initial state ($t = 0$), using a technique similar to the one used by Reynolds and Palmer (1998). Once we had the forward propagator matrix, we then used a standard singular value decomposition program available from LAPACK (Anderson et el. 1995) to obtain the singular values and the associated singular vectors (initial and evolved) of the system. Since the length of the state vector was 341, we had 341 singular values and singular vectors, respectively. The leading singular value (associated with the most unstable singular vector) was 31.59. There were 73 singular values greater than 1.01 (unstable singular vector), 11 singular values greater than 0.99 but less than 1.01 ('neutral' singular vectors), and 257 singular values less than 0.99 (decaying singular vectors).

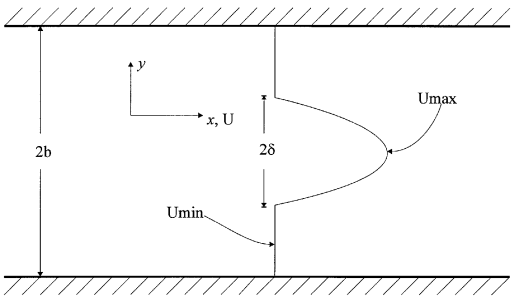

Fig. 1. A schematic drawing of the basic state jet profile in the channel. U_{max} and U_{min} are the maximum and minimum wind speeds in the x-direction, respectively. $2b = 3850$ km is the width of the channel and $2\delta = 1155$ km is the width of the jet. The basic state wind in the y-direction is always zero.

3.3. The 'true' solution

Although the singular vectors provided information about the relationship between the initial and final states, they did not provide the detailed evolutions in between. However, a model integration of the shallow water system from any initial singular vector provided us with the complete evolution that we desired. The 'true' solution for the problem was not chosen to be the leading unstable singular vector, but rather, was a 'neutral' singular vector. We felt that this choice would be a particular tough test of any data assimilation system, because any discrepancies or faults of the data assimilation system could easily excite an unstable singular vector and the error would grow rapidly.

Among the 11 'neutral' singular vectors, some were stationary during the 96-h time period and some were moving eastward. We chose an eastward-moving 'neutral' singular vector (singular vector No. 83 with a singular value of 1.003) as the initial condition and performed a 96-h forecast. The evolution of the 'neutral' singular vector then served as the 'true' solution, and provided the basis from which to create the observations. During the 96-h time period, the 'neutral' singular vector propagated eastward for about one evolution (3850 km), with some structural variations in the north-south direction. It should be noted that this singular vector was 'neutral' only in the sense that the amplitudes at $t = 0$ and $t = 96$ h were approximately the same. In fact, the evolution of this singular vector displayed a growth/decay cycle with rapid growth to about 25 times the initial amplitude at $t = 48$ h, followed by decay back to close to the original amplitude at the end of the period. The evolution of the 'neutral' and the leading unstable singular vectors at $t = 96$ h was plotted in Fig. 2(a) and (b), respectively. Figure 2(b) was served as a visual reference to see if the analysis solutions from both algorithms were stable.

3.4. Observations, observation networks, and observation error specification

The observations were created from the 'true' solution described in Section 3.3. The observations were assumed to have spatially and serially uncorrelated errors chosen from a normal distribution,

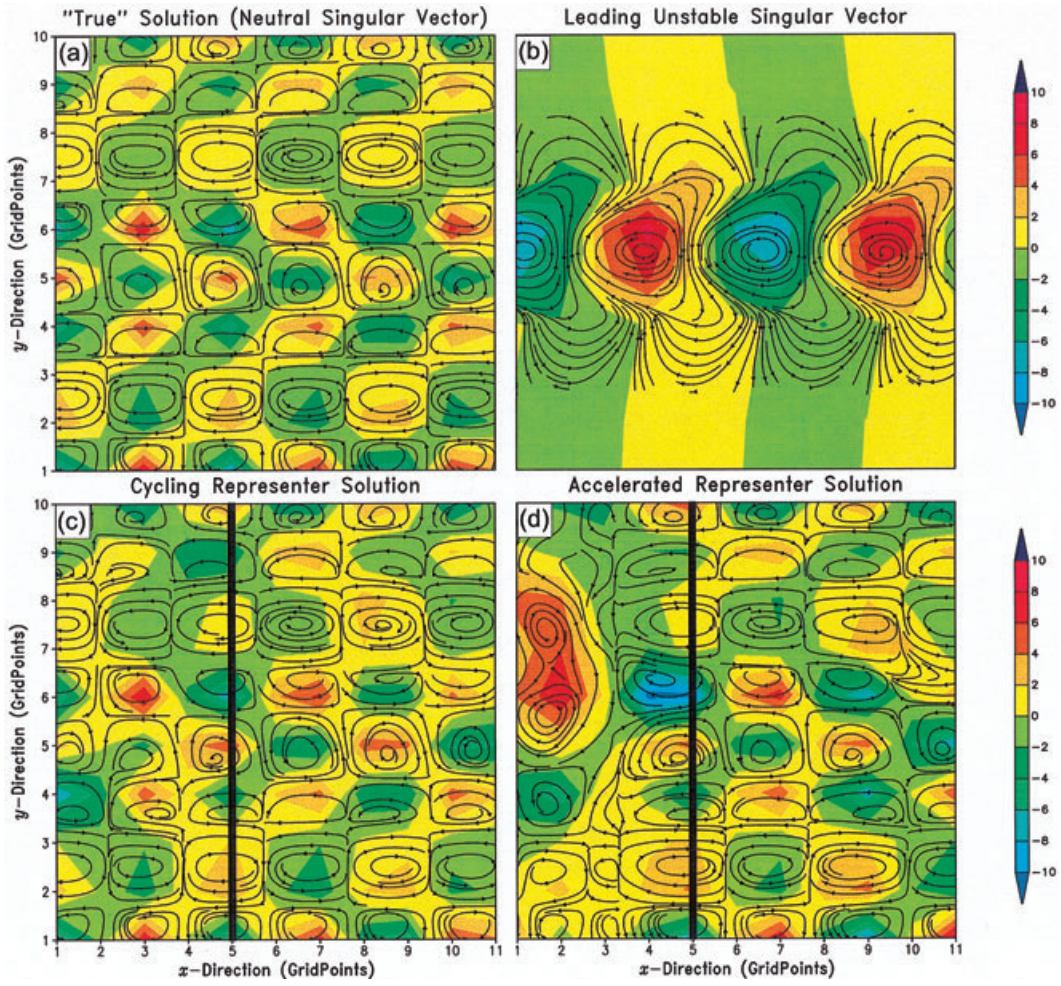

Fig. 2. The wind fields (in streamlines) and the geopotential fields (in color) at $t = 96$ h (end of the data assimilation period) from the 'true' solution and two different representer algorithms using observation Network A. Panels (a) and (b) show the evolution of the 'neutral' ('true' solution) and leading unstable singular vectors, respectively. Panels (c) and (d) show the cycling representer and the accelerated representer solutions, respectively. Notice that all the values in panel (b) have been scaled by 10^{-2} before being displayed. The observation Network A is represented by the thick black lines in panel (c) and (d), respectively.

with a standard deviation equal to 0.000 01% of maximum values of the corresponding fields. Thus, the observations were essentially perfect for practical purposes, but were mathematically imperfect. The imperfectness in the observations helped to ensure that the sum of the matrices $\mathbf{R}_e + \mathbf{R}$ was positive-definite in this study.

Two observation networks (A and B) were employed in this study and are described as follows.

3.4.1. Network A. Observations collocated with the 11 model gridpoints were available for the geopotential and wind components every 6 h (twice in each 12 h short cycle) along a fixed line in the north-south direction ($x = 5$), as indicated by the thick black lines in Fig. 2(c) and (d), respectively. This network had a maximum spatial coverage in the y-direction at $x = 5$, but with a very poor coverage in the x-direction. Since the 'true' solution moved eastward, we hoped that the

poor coverage in the x-direction would be compensated by the eastward propagation of information by the model and through the cyclical boundary conditions used in the x-direction. This network was created to see if the data assimilation systems were capable of spreading the sparse observed information as we hoped.

3.4.2. Network B. Observations collocated with 11 model gridpoints were available for geopotential and wind components every 6 h (twice in each 12 h short cycle) along a fixed line (not shown) in the east–west direction ($y = 6$). Unlike Network A, this network had a maximum spatial coverage in the x-direction at $y = 6$, but with a very poor coverage in the y-direction. Since the 'true' solution showed very little interaction in the y-direction, we suspected that data assimilation systems would have difficulty in propagating the observed information in the y-direction. Essentially, Network B was chosen to examine the effect of a 'poor' sampling strategy.

3.5. *Specification of the initial conditions and initial background error covariance*

In all the experiments, the initial ($t = 0$) background conditions for the first cycle were always specified to be equal to zero, the initial background conditions for the subsequent cycles were provided by the analysis at the end of the previous cycle (cycling). We arbitrarily specified homogeneous, diagonal, univariate initial ($t = 0$) background error covariance for the geopotential and wind components. We know from experience that such initial background error covariance is, in actuality, far from diagonal and is often multivariately coupled, but we wished to make the initial background error covariance simple and not very realistic. Although we could have found more realistic initial background error covariances through experiments, such as the χ^2-tests used in Xu and Daley (2000), we decided to use diagonal, univariate forms so that we could test the ability of the cycling representer algorithm in updating the initial background error covariance. We also wished to test the ability of the accelerated representer algorithm (where the initial background error covariance is not updated at the beginning of each cycle) in the case of a poor choice of the initial background error covariance. The initial (scalar) background error variances used in this study were $5 \times 10^{-3}\ \mathrm{m^2\,s^{-2}}$ and $1 \times 10^{-5}\ \mathrm{m^4\,s^{-4}}$ for wind and geopotential fields, respectively.

4. Results

For each observation network, we conducted 2 experiments, namely cycling with initial background covariance updating (cycling representer) and cycling with no initial background covariance updating (accelerated representer), respectively. In the following, we first examine the details of the analysis at the end of the data assimilation period ($t = 96$ h). We then examine the overall performance of each algorithm using an objective global measurement of error reductions for observation Network A and discuss the advantages and disadvantages of each algorithm. Finally, we demonstrate the impact of different sampling strategies by comparing the analysis from both Network A and B.

We first demonstrate the performance of the cycling and accelerated representer algorithms at $t = 96$ h using observation Network A, by displaying in Fig. 2 the wind and geopotential analysis and comparing them to the 'true' solution and the leading unstable singular vector. The geopotential fields are color shaded (see thermometer on the right) while the wind fields are plotted as streamlines. Figures 2(a) and (b) are the evolution of the 'neutral' and leading unstable singular vectors at $t = 96$ h, respectively. Figure 2(a) is served as the 'true' solution, while Fig. 2(b) is served as a visual reference to see if the two different representer algorithms were stable. It should be noted that all the variables in Fig. 2(b) have been scaled by 10^{-2} before being displayed to use the same color bars (thermometer) on the right. The wind and geopotential analysis produced by the cycling and accelerated representer algorithms are displayed in Figs. 2(c) and (d), respectively. Both algorithms appeared to be stable and capable of capturing the main features of the 'true' solution at the end of 96 h as indicated in Fig. 2. The cycling representer appeared to be very close to the 'true' solution [comparing panels (a) and (c)]. The results from the accelerated algorithm [panel (d)], on the other hand, were less accurate than the ones from the cycling representer algorithm. There were some distortions

near the left boundaries in the x-direction. We attribute the distortions to the mis-specification of the initial background error covariance in each cycle.

To further compare the overall accuracies of the analysis during the whole time period of data assimilation, we used a square root of total perturbation energy (SRTPE) as an indicator to represent the overall error at any given timestep. The SRTPE was defined as follows:

$$\text{SRTPE} = \frac{1}{2}\sum_{i=1}^{10}\sum_{j=1}^{10} \times \left((u_{ij}^t - u_{ij}^a)^2 + (v_{ij}^t - v_{ij}^a)^2 + \frac{1}{\Phi_j}(\phi_{ij}^t - \phi_{ij}^a)^2 \right)^{1/2}, \quad (13)$$

where i and j were gridpoint indexes in the x and y directions, respectively. Since the Arakawa C-grid was used in the model, we first obtained all the variables on the geopotential points and then calculated the SRTPE there. The smaller the SRTPE, the better the analysis was.

Figure 3 displays the plots of SRTPE against time for the two different assimilation algorithms using observation Network A. The dash-dot line represents the accelerated representer algorithm (no initial background covariance updates). The solid line represents the cycling representer algorithm (initial background covariance was updated every cycle). Both algorithms produced stable analysis, despite the dynamic model being barotropically unstable.

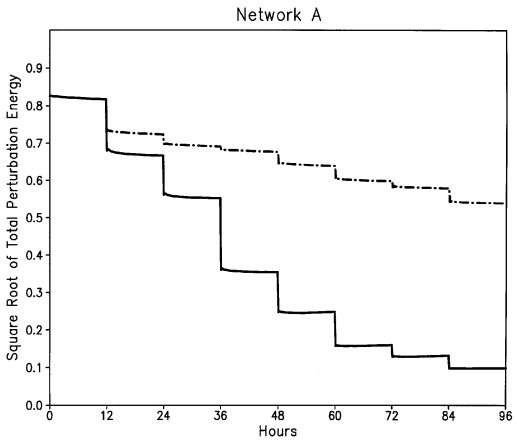

Fig. 3. The overall error level [from the SRTPE metric, eq. (12)] as a function of time (in hours) for the two different representer algorithms using observation Network A. The solid and dash-dot lines are for the cycling and accelerated representer algorithms, respectively.

The cycling representer algorithm (solid line) indicated a steady reduction of overall errors measured by the SRTPE metric. The error levels were essentially flat during each cycle, and the error reduction occurred at the end of each cycle (rather reminiscent of the behavior of a Kalman filter). The accelerated representer algorithm (dash-dot line) also showed a steady reduction of the overall errors but with a lower error reduction rate measured by the SRTPE metric. The accelerated and the cycling representer algorithms gave exactly the same results in the first cycle, but in later cycles, the initial background covariance update of the cycling representer is clearly advantageous. From computational a point of view, however, the accelerated representer algorithm was very affordable, while the cycling representer algorithm was much more expensive. It is unfortunate that initial background error covariance updating is so computationally burdensome.

One of the important issues in data assimilation is network design and sampling strategy. As discussed in Section 3.4, both observation Networks A and B had the same number of observations, but they were distributed very differently. Instead of using the SRTPE defined in eq. (13) as a function of time, we calculated a square root of the zonal average of the perturbation energy at the end of the data assimilation period ($t = 96$ h) for the two different representer algorithms with both observation networks. The results are plotted in Fig. 4, with the line conventions being the same as in Fig. 3. Panel (a) is for Network A and panel (b) is for Network B. The overall zonal averaged errors from observation Network A were clearly smaller than for Network B, except in the vicinity of $y = 6$ (where all the observations of Network B were located) for both cycling and accelerated algorithms.

The better results obtained with Network A were not unexpected because the 'true' solution was propagating eastwards instead of southwards or northwards. By using observation Network B, we effectively reduced the amount of useful information from observations and instead added

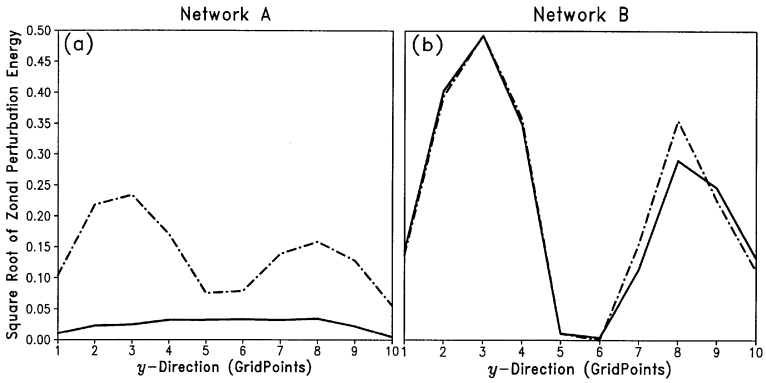

Fig. 4. The square root of zonal mean perturbation energy as a function of zonal gridpoint position (y-direction) at $t = 96$ h for observation Networks A (a) and B (b). The solid and dash-dot lines are for the cycling and accelerated representer algorithms, respectively.

redundant information from observations. The results for Network B clearly suggested that the analysis were very close to the 'true' solution in the vicinity of the abundant observations (near $y = 6$), but deteriorated rapidly to the north and south.

In Section 3.3, it was noted that although the 'truth' corresponded to the evolution of the 'neutral' singular vector, the predictive model of Section 2.2 also supported very rapidly growing barotropically unstable singular vectors. It was suggested in Section 2.2 that the spurious excitation of the leading unstable singular vectors [Fig. 2(b)], could lead to a catastrophic result in an improperly posed data assimilation system. Clearly, the results shown in Figs. 2–4 show no evidence of this.

5. Summary and conclusions

This study had two objectives. The first objective was to validate the cycling representer algorithm of Xu and Daley (2000) in a multidimensional, multivariate environment with dynamic instability. The second objective was to formulate and test the accelerated representer algorithm to be used in a future operational four-dimensional atmospheric data assimilation system.

A linear shallow water version of COAMPS with a barotropically unstable basic state was used as the test bed to conduct data assimilation experiments. The barotropic instability was produced through the use of a cosine square jet profile in the basic state. The system had a sufficiently small number of degrees of freedom that all of the singular values and singular vectors of this system over a given time interval could be obtained explicitly. The evolution of a 'neutral' eastward-propagating singular vector was selected as the 'truth', against which all data assimilation experiments were to be evaluated.

Almost perfect observations for two different observation networks were obtained by sampling the 'true' solution. For each network two data assimilation algorithms were tested. First, there was the cycling representer algorithm (which includes the initial background error covariance update at the beginning of each cycle). The second algorithm was the accelerated representer algorithm (which is much less computationally intensive but does not update the initial background error covariance).

Both the cycling and accelerated representer algorithms were found to produce stable results for observation Network A and B, despite the possibility of enormous error growth associated with the dynamically unstable system. The results clearly showed that Network A is much better than Network B in sampling useful information with the same number of observations. With the observation Network A, the cycling representer algorithm (which updates the initial background error covariance at the beginning of each cycle) produced the most accurate analysis at the end of the data assimilation period. The accelerated

representer produced the same result as the cycling representer algorithm in the first cycle, but it had slower error reductions for subsequent cycles. This was not surprising, because we had deliberately chosen a rather poor representation of the initial background error covariance to be specified at the beginning of each cycle.

The results from the accelerated representer algorithm were, however, very encouraging because the latter is sufficiently computationally tractable to be used on present day multi-processor machines for operational applications. It was clear from our experiments that the assimilation errors were steadily reduced after each cycle for this algorithm, although the poor specification of the initial background error covariance at the beginning of each cycle led to less error reduction than would have been desirable. It is likely that a more realistic initial background error covariance based on the statistical approaches used in 3DVAR algorithms would be helpful. Beyond that, there is always the possibility of using ensemble techniques to introduce more flow-dependence into the initial background error covariance.

An effort at the Naval Research Laboratory (NRL) in Monterey is currently underway to construct a four-dimensional variational global data assimilation system using the accelerated representer algorithm. This algorithm is being constructed for massively parallel machines and involves the parallel tangent linear and adjoint models of NOGAPS (Navy Operational Global Atmospheric Prediction System — Hogan and Rosmond, 1991). The observation and forward instrument modeling, as well as the construction of the initial background error covariance, all come from NAVDAS (NRL Atmospheric Variational Data Assimilation System — Daley and Barker, 2000, 2001). This new algorithm, which we refer to as NAVDAS-AR (NAVDAS–Accelerated Representer), is a natural extension of the three-dimensional observation-space NAVDAS system to four dimensions.

6. Acknowledgements

We thank Carolyn Reynolds of NRL in Monterey for discussions regarding several aspects of the singular vector analysis and Andrew Bennett of FNMOC for reviewing and discussing this paper. We also thank two anonymous reviewers for their constructive comments. Support of the sponsor, NRL, under Program Element 0601153N is gratefully acknowledged.

REFERENCES

Anderson, E., Bai, Z., Bischof, C., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., Ostrouchov, S. and Sorensen, D. 1995. *LAPACK users' guide*, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia. 325 pp.

Amodei, L. 1995. Solution approchee pour un problem d'assimilation de donnees meteorologiques avec prise en compte de l'erreur de modele. *C. R. Acad. Sci., Ser. IIa* **321**, 1087–1094.

Bennett, A. F. 1997. *Inverse methods and data assimilation*, 1997 Summer School Lecture Notes, Oregon State University, Corvallis, Oregon.

Bennett, A. F. and McIntosh, P. C. 1982. Open ocean modeling as an inverse problem: tidal theory. *J. Phys. Ocean.* **12**, 1004–1018.

Bennett, A. F. and Thorburn, M. A. 1992. The generalized inverse of a nonlinear quasigeostrophic ocean circulation model. *J. Phys. Ocean.* **22**, 213–230.

Chua, B. S. and Bennett, A. F. 2001. An inverse ocean modeling system. *Ocean Modeling* **3**, 137–165.

Cohn, S. E. and Parrish, F. P. 1991. The behavior of forecast error covariance for a Kalman filter in two dimensions. *Mon. Wea. Rev.* **119**, 1757–1785.

Courtier, P. 1998. Dual formulation of four dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* **123**, 2449–2461.

Courtier, P., Thepaut, J.-N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4D-Var using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387.

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

Daley, R. and Barker, E. 2000. *The NAVDAS source book*, Defense Printing Office for the Naval Research Laboratory, Washington, DC. 217 pp.

Daley, R. and Barker, E. 2001. NAVDAS — formulation and diagnostics. *Mon. Wea. Rev.* **129**, 869–883.

Egbert, G., Bennett, A. and Foreman, M. 1994. TOPEX/POSEIDON tides estimated using a global inverse method. *J. Geophys. Res.* **99**, 24,821–24,852.

Evensen, G. 1997. Advanced data assimilation for strong nonlinear dynamics. *Mon. Wea. Rev.* **125**, 1342–1354.

Evensen, G. and van Leeuwen, P. J. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.* **128**, 1852–1867.

Golub, G. H. and Van Loan, C. F. 1996. *Matrix computations*, 3rd edn. The Johns Hopkins Press, Baltimore and London. 694 pp.

Hodur, M. R. 1997. The Naval Research Laboratory's coupled ocean/atmosphere mesoscale prediction system (COAMPS). *Mon. Wea. Rev.* **125**, 1414–1430.

Hogan, T. and Rosmond, T. 1991. The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast system. *Mon. Wea. Rev.* **119**, 1786–1815.

Kuo, H. L. 1949. Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteorol.* **6**, 105–122.

Menard, R. 1994. Kalman filtering of Burgers' equation and its application to atmospheric data assimilation. Ph.D. Thesis, Department of Atmospheric and Ocean Sciences, McGill University, Montreal, Canada.

Reynolds, C. A. and Palmer, T. N. 1998. Decaying singular vectors and their impact on analysis and forecast correction. *J. Atmos. Sci.* **55**, 3005–3023.

Strang, G. 1986. *Introduction to applied mathematics.* Wellesley-Cambridge Press, 758 pp.

Todling, R. and Ghil, M. 1994. Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one-layer results. *Mon. Wea. Rev.* **122**, 183–204.

Xu, L. 1995. The study of mesoscale land-air-sea interaction processes using a nonhydrostatic model. Ph.D. Dissertation, North Carolina State University, Raleigh, 336 pp. (Available from Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC).

Xu, L. and Daley, R. 2000. Towards a true four dimensional data assimilation algorithm: application of a cycling representer algorithm to a simple transport problem. *Tellus* **52A**, 109–128.

Zupanski, D. 1997. A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.* **125**, 2274–2292.