The University of Reading

Inner loop stopping criteria for incremental four-dimensional variational data assimilation

A.S. Lawless and N.K. Nichols

NUMERICAL ANALYSIS REPORT 5/05

Department of Mathematics

The University of Reading

Whiteknights, PO Box 220

Reading

Berkshire RG6 6AX

Department of Mathematics

Abstract

Incremental four-dimensional variational data assimilation is a method that solves the assimilation problem by minimizing a sequence of approximate 'inner loop' functions. In any implementation of such a scheme a decision must be made as to how accurately to solve each of the inner minimization problems. In this paper we apply theory that we have recently developed to derive a new stopping criterion for the inner loop minimizations, that guarantees convergence of the outer loops. This new criterion is shown to give improved convergence compared to other commonly used inner loop stopping criteria.

Contents

1 Introduction

Four-dimensional variational data assimilation (4D-Var) is a method to estimate the model trajectory most consistent with the available observational data over a particular time window through the minimization of a nonlinear cost function. A major advantage of this technique over previously used asssimilation methods is that the dynamical forecasting model is used as a constraint on the assimilation. Early studies showed how this allows a 4D-Var assimilation to infer information from observations in a dynamically consistent way (for example, Courtier and Talagrand 1987, Thépaut and Courtier 1991, Rabier and Courtier 1992, Thépaut *et al.* 1993). However, in its full, nonlinear formulation 4D-Var is very computationally demanding and so methods of simplification are needed before operational implementation is possible. The most successful of these used currently is the incremental approach (Courtier *et al.*, 1994), which has recently been shown to be equivalent to using a Gauss-Newton iteration method to minimize the nonlinear cost function (Lawless *et al.* 2005a, 2005b). Such a method has been implemented operationally for several weather and ocean forecasting systems (see, for example, Rabier *et al.* 2000, Weaver *et al* 2003) and is under development for others. In the incremental approach the minimization of the 4D-Var nonlinear cost function is replaced by a series of minimizations of convex quadratic cost functions constrained by a linear model. Each of these quadratic cost functions is minimized using an inner iteration or 'inner loop' and the solution is then used to update the nonlinear model trajectory in an 'outer loop'. These outer loop trajectories can then be used to define new quadratic cost functions and so the method cycles through several outer loops, each of which requires the minimization of an inner loop cost function. An advantage of this method is that it allows the inner loop cost functions to be simplified, for example by solving the problem at a lower spatial resolution, and this makes the 4D-Var problem tractable for operational use.

One problem with the incremental 4D-Var method is deciding how many iterations to perform in each inner loop minimization. It is clear that the inner loop problem must be solved to sufficient accuracy to ensure an accurate update to the nonlinear trajectory. However, we may expect that since the inner loop cost function is an approximation to the full nonlinear cost function, then it is not worth solving it too accurately. In a study using a model of barotropic flow, Laroche and Gautier (1998) found that it was preferable to choose a lower number of inner iterations and update the outer loop trajectory more often. Hence it is important to choose carefully the stopping criterion for the inner loop minimization. In many operational implementations of incremental 4D-Var the inner loop has been stopped using criteria based on a fixed number of iterations or on the norm of the gradient.

In this paper we propose a new stopping criterion for the inner loop minimization, based on the relative change in the inner loop gradient. The theory for this criterion arises from our recent theoretical work on the convergence of approximate Gauss-Newton iterations (Gratton *et al.* 2004). Preliminary application of these results showed a potential benefit from using such a stopping criterion within an incremental 4D-Var system (Lawless *et al.*, 2005c). Here we present the theoretical basis for this criterion and show how it can be used to guarantee the convergence of the outer loops. Numerical experiments are presented to illustrate how a stopping criterion based on this theory leads to faster and smoother convergence of the 4D-Var assimilation.

The paper is arranged as follows. In Section 2 we present the incremental 4D-Var algorithm and review its relationship to the Gauss-Newton iteration. In Section 3 we discuss different stopping criteria for the inner loop minimization and show how the theory of Section 2 naturally leads to a particular choice. The subsequent sections present a series of numerical experiments which illustrate the advantages of this new criterion under different conditions. The model and assimilation system is described in Section 4. In Section 5 we compare the performance of the new stopping criterion to other commonly used criteria. Section 6 illustrates how the tolerance of the new criterion may be tuned to improve the convergence of incremental 4D-Var. In Section 7 we examine the performance of the new criterion when the inner loop is at a lower spatial resolution than the outer loop. Finally, in Section 8, we summarize our findings and highlight the implications for operational assimilation systems.

2 Mathematical formulation of incremental 4D-Var

2.1 Incremental 4D-Var

In the full nonlinear formulation of 4D-Var we aim to find the initial model state $\mathbf{x}_0$ that minimizes the cost function

$$\mathcal{J}[\mathbf{x}_0] = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{n}(H_i[\mathbf{x}_i] - \mathbf{y}_i^o)^{\mathrm{T}}\mathbf{R}_i^{-1}(H_i[\mathbf{x}_i] - \mathbf{y}_i^o) \quad (1)$$

subject to satisfying the discrete nonlinear forecast model

$$\mathbf{x}_i = S(t_i, t_0, \mathbf{x}_0). \quad (2)$$

In this expression $\mathbf{x}^b$ is the background field with error covariance matrix $\mathbf{B}_0$, $\mathbf{y}_i^o$ are the observations at time t_i, with error covariance matrices $\mathbf{R}_i$ and H_i is the observation operator which maps the field from model space to observation space. The operator $S(t_i, t_0, \mathbf{x}_0)$ represents the solution operator of the nonlinear model, which maps the initial state $\mathbf{x}_0$ at time t_0 to the state at time t_i. The incremental formulation of 4D-Var, introduced by Courtier *et al.* (1994), replaces a direct minimization of (1) with a series of minimizations of linearly constrained cost functions, each of which is solved for an increment to the current estimate. This gives the following algorithm:

- Set first guess $\mathbf{x}_0^{(0)} = \mathbf{x}_b$.
- Repeat for $k = 0, \ldots, K-1$
  - Minimize

$$\begin{aligned}\tilde{\mathcal{J}}^{(k)}[\delta\mathbf{x}_0^{(k)}] &= \frac{1}{2}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}])^{\mathrm{T}}\mathbf{B}_0^{-1}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}]) \\ &+ \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)}) \quad (3)\end{aligned}$$

    with

$$\mathbf{d}_i^{(k)} = \mathbf{y}_i^o - H_i[\mathbf{x}_i{}^{(k)}], \quad (4)$$

$$\delta\mathbf{x}_i^{(k)} = \mathbf{L}(t_i, t_0, \mathbf{x}^{(k)})\delta\mathbf{x}_0^{(k)}, \quad (5)$$

where $\mathbf{H}_i$ is the linearization of the observation operator H_i around the state $\mathbf{x}_i^{(k)}$ at time t_i and $\mathbf{L}(t_i, t_0, \mathbf{x}^{(k)})$ is the solution operator of a linear model linearized around the nonlinear model trajectory.

  - Update $\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta\mathbf{x}_0^{(k)}$

- Set analysis $\mathbf{x}^a = \mathbf{x}_0^{(K)}$

For an exact method the linear model $\mathbf{L}(t_i, t_0, \mathbf{x}^{(k)})$ is equal to the linearization of the discrete nonlinear model $S(t_i, t_0, \mathbf{x}_0)$, but in practice an approximate linearization is often used (for example, Lorenc *et al.* 2000, Mahfouf and Rabier 2000). It was shown by Lawless *et al.* (2005a, 2005b) that incremental 4D-Var with an exact tangent linear model is equivalent to a Gauss-Newton method applied to minimize the nonlinear cost function (1). We now review this equivalence.

2.2 Gauss-Newton algorithm

The Gauss-Newton method is an approximation to a Newton iteration, in which the second order terms of the Hessian are neglected (Dennis and Schnabel, 1996). To illustrate this we consider a general nonlinear least squares problem

$$\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2} \parallel \mathbf{f}(\mathbf{x}) \parallel_2^2 = \frac{1}{2}\mathbf{f}(\mathbf{x})^{\mathrm{T}}\mathbf{f}(\mathbf{x}), \tag{6}$$

with $\mathbf{x} \in \mathbb{R}^n$, which we assume to have an isolated local minimum $\mathbf{x}^*$. We write

$$\begin{aligned} \nabla\mathcal{J}(\mathbf{x}) &= \mathbf{J}^{\mathrm{T}}\mathbf{f}(\mathbf{x}), & (7)\\ \nabla^2\mathcal{J}(\mathbf{x}) &= \mathbf{J}^{\mathrm{T}}\mathbf{J} + Q(\mathbf{x}), & (8) \end{aligned}$$

where $\mathbf{J}$ is the Jacobian of $\mathbf{f}(\mathbf{x})$ and $Q(\mathbf{x})$ are the second order derivative terms. Then the Gauss-Newton iteration for minimizing (6) is given by

$$\text{Solve for } \delta\mathbf{x}^{(k)}: \quad (\mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}}\mathbf{J}(\mathbf{x}^{(k)}))\delta\mathbf{x}^{(k)} = -\mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}}\mathbf{f}(\mathbf{x}^{(k)}), \tag{9}$$

$$\text{Update:} \quad \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}. \tag{10}$$

We note that for large systems it may be impossible to solve (9) directly. In such case the solution $\delta\mathbf{x}^{(k)}$ may be found by an 'inner' minimization of the function

$$\tilde{\mathcal{J}}(\delta\mathbf{x}^{(k)}) = \frac{1}{2} \parallel \mathbf{J}(\mathbf{x}^{(k)})\delta\mathbf{x}^{(k)} + \mathbf{f}(\mathbf{x}^{(k)}) \parallel_2^2 . \tag{11}$$

To see the link with incremental 4D-Var, we note that (1) can be written in the form (6) by putting

$$\mathbf{f}(\mathbf{x}_0) = \begin{pmatrix} \mathbf{B}_0^{-1/2}(\mathbf{x}_0 - \mathbf{x}^b) \\ \mathbf{R}_0^{-1/2}(H_0[\mathbf{x}_0] - \mathbf{y}_0^o) \\ \vdots \\ \mathbf{R}_n^{-1/2}(H_n[\mathbf{x}_n] - \mathbf{y}_n^o) \end{pmatrix}. \tag{12}$$

Then, as was shown by Lawless *et al.* (2005a, 2005b), if we apply the Gauss-Newton iteration to the nonlinear cost function (1), the step given by (11) is equivalent to the minimization of the inner loop cost function (3). Thus incremental 4D-Var is equivalent to a Gauss-Newton iteration and we can use the theory for the Gauss-Newton method to analyse the behaviour of incremental 4D-Var.

2.3 The truncated algorithm

In practice the inner loop cost function of incremental 4D-Var is minimized using an iterative procedure, such as a conjugate gradient method. Such minimization methods employ a stopping criterion to determine when the solution has been found to sufficient accuracy. Hence the minimum of the inner cost function is not found exactly, but only to within the degree of accuracy determined by the stopping criterion. In the context of the Gauss-Newton algorithm this can be considered as a method in which the step (9) of the algorithm is solved inexactly. Thus we obtain the truncated Gauss-Newton (TGN) algorithm

$$\text{Solve for } \delta\mathbf{x}^{(k)}: \quad (\mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}}\mathbf{J}(\mathbf{x}^{(k)}))\delta\mathbf{x}^{(k)} = -[\mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}}\mathbf{f}(\mathbf{x}^{(k)}) + \mathbf{r}^{(k)}], \tag{13}$$

$$\text{Update:} \quad \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}, \tag{14}$$

where $\mathbf{r}^{(k)}$ is the residual arising from the premature termination of the inner loop minimization. Convergence of the TGN algorithm is guaranteed by the following theorem, first stated in Lawless *et al.* (2005b) and proved in Gratton *et al.* (2004):

Theorem 1 *Assume that* $\hat{\beta} < 1$ *and that on each iteration the Gauss-Newton method is truncated with*

$$\| \mathbf{r}^{(k)} \|_2 \leq \beta_k \| \mathbf{J}(\mathbf{x}^{(k)})^T \mathbf{f}(\mathbf{x}^{(k)}) \|_2, \tag{15}$$

where

$$\beta_k \leq \frac{\hat{\beta}- \parallel (\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{J}(\mathbf{x}^{(k)}))^{-1}Q(\mathbf{x}^{(k)}) \parallel_2}{1+ \parallel (\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{J}(\mathbf{x}^{(k)}))^{-1}Q(\mathbf{x}^{(k)}) \parallel_2}. \quad (16)$$

Then there exists $\eta > 0$ such that, if $\parallel \mathbf{x}_0 - \mathbf{x}^* \parallel_2 \leq \eta$, the truncated Gauss-Newton iteration (TGN) converges to the solution $\mathbf{x}^*$ of the nonlinear least squares problem (6).

Proof: See Gratton *et al.* (2004). □

In the next section we demonstrate how this theorem may be used to determine a suitable stopping criterion for the inner loop of incremental 4D-Var.

3 Defining a suitable stopping criterion

As stated in the introduction, it is important to choose carefully the stopping criterion for the inner loop minimization of incremental 4D-Var. If too few inner iterations are performed, then the outer loop iterations may not converge to the solution of the original problem. Although this may not appear to be a problem if we are not running many outer loops, it does mean that with too few inner iterations the system may diverge and the final analysis may be further from the truth than the first guess. Hence even in the case where only a few outer loops are performed, adequate minimization of the inner loop is necessary. However, we also wish to avoid performing too many iterations on the inner loop. If the current outer iterate $\mathbf{x}^{(k)}$ is far from the true solution, then iteration of the inner minimization to too high an accuracy may result in extra computational work which does not lead to increased accuracy in the outer loops. We discuss some common ways in which the inner loop minimization is stopped and then use the theory of Section 2 to propose a new stopping criterion.

3.1 Choice of criterion

Since we know that the gradient of a function is zero at its minimum, a natural method for stopping the inner iteration is to stop when the norm of the gradient of the inner loop cost function falls below a specified tolerance. If we use ϵ to denote a user-set tolerance and use subscript m to denote the iteration count of the inner loop, then we can write this criterion as

(C1) Absolute norm of gradient

$$\| \nabla \tilde{\mathcal{J}}^{(k)}_{(m)} \|_2 < \epsilon. \tag{17}$$

An alternative stopping criterion from optimization theory is to stop when the relative change in the inner loop cost function itself is less than a given tolerance (Gill *et al.* 1986, p.306). Such a measure ensures that extra iterations are not performed if they will not significantly decrease the value of the cost function. We can write such a criterion

(C2) Relative change in function

$$|\tilde{\mathcal{J}}^{(k)}_{(m+1)} - \tilde{\mathcal{J}}^{(k)}_{(m)}| < \epsilon(1 + \tilde{\mathcal{J}}^{(k)}_{(m)}). \tag{18}$$

Often other criteria are chosen for practical reasons, for example a fixed number of inner iterations may be used to stop the inner loop (Rabier *et al.* 1998, Weaver *et al.*, 2003).

We now derive a new stopping criterion for incremental 4D-Var using the theory for truncated Gauss-Newton methods presented in the previous section. We suppose that a total of M inner iterations are performed in the inner minimization, after which the solution to the inner problem (13) is given by $\delta\mathbf{x}^{(k)}_{(M)}$ with residual $\mathbf{r}^{(k)}_{(M)}$. Then Theorem 1 proves convergence of the truncated Gauss-Newton method if the ratio

$$\mathcal{R} = \frac{\| \mathbf{r}^{(k)}_{(M)} \|_2}{\| \mathbf{J}(\mathbf{x}^{(k)})^T \mathbf{f}(\mathbf{x}^{(k)}) \|_2} \tag{19}$$

is bounded, where (16) provides the bound. Thus a natural stopping criterion is to stop the inner iteration when this ratio is less than a given tolerance. We examine this formula in more detail to understand how we may calculate it in practice.

Firstly we note that the gradient of the general inner loop cost function (11) at the point $\delta\mathbf{x}^{(k)}_{(M)}$ is given by

$$\nabla \tilde{\mathcal{J}}^{(k)}(\delta\mathbf{x}^{(k)}_{(M)}) = (\mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}} \mathbf{J}(\mathbf{x}^{(k)}))\delta\mathbf{x}^{(k)}_{(M)} + \mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}} \mathbf{f}(\mathbf{x}^{(k)}). \tag{20}$$

A comparison of this with (13) reveals that $\mathbf{r}^{(k)}_{(M)} = \nabla\tilde{\mathcal{J}}^{(k)}(\delta\mathbf{x}^{(k)}_{(M)})$ and so the norm $\| \mathbf{r}^{(k)}(\delta\mathbf{x}^{(k)}_{(M)}) \|_2$ in the numerator of $\mathcal{R}$ is simply the final norm of the gradient of the inner loop cost function. Whereas criterion (C1) stops the iteration based on the absolute norm of this gradient, Theorem 1 implies that the inner loop should be stopped on the size of this norm relative to $\| \mathbf{J}(\mathbf{x}^{(k)})^T\mathbf{f}(\mathbf{x}^{(k)}) \|_2$. Furthermore, from (7) we see that the norm of $\mathbf{J}(\mathbf{x}^{(k)})^T\mathbf{f}(\mathbf{x}^{(k)})$ is identical to the norm of the gradient of the outer loop cost function evaluated at $\mathbf{x}^{(k)}$. Hence $\mathcal{R}$ is simply the relative sizes of the gradients of the inner loop and outer loop cost functions. The theorem for the convergence of the truncated Gauss-Newton algorithm indicates that this ratio should form the basis of the stopping criterion for the inner loop minimization.

We can simplify this further by noting that where $\delta\mathbf{x} = 0$, we have $\nabla\mathcal{J}(\mathbf{x}^{(k)}) = \nabla\tilde{\mathcal{J}}^{(k)}$. Since, within an incremental 4D-Var system, we assume that the first guess to $\delta\mathbf{x}$ is zero at the start of each inner loop, then $\nabla\mathcal{J}(\mathbf{x}^{(k)})$ on each outer loop is equal to the value of $\nabla\tilde{\mathcal{J}}^{(k)}_{(0)}$ at the start of the inner minimization. Thus $\mathcal{R}$ can be calculated using gradient information from only the inner loop cost function. This implies a stopping criterion which is a simple relative criterion on the norm of the inner loop gradient. We can write such a criterion in the form

(C3) Relative change in gradient

$$\frac{\| \nabla\tilde{\mathcal{J}}^{(k)}_{(m)} \|_2}{\| \nabla\tilde{\mathcal{J}}^{(k)}_{(0)} \|_2} < \epsilon, \tag{21}$$

where the subscripts indicate the index of the inner loop iteration. From the theory for the convergence of the truncated Gauss-Newton method it appears more natural to use this relative stopping criterion on the norm of the inner loop gradient, rather than criteria (C1) or (C2). Provided that we choose $\epsilon \leq \min_k \beta_k$, with β_k bounded as in (16), then Theorem 1 ensures the convergence of incremental 4D-Var using stopping criterion (C3).

3.2 Choice of tolerance

Besides determining the type of stopping criterion to use, we must also decide what value of the tolerance ϵ to choose. If we use the relative change in the inner loop gradient as the criterion, then a suitable tolerance is provided by the right hand side of (16). However in practice this bound will be hard to calculate. The terms $Q(\mathbf{x})$ are the second order terms of the Hessian and these include the second derivative terms of the nonlinear model, which cannot be calculated easily for a complex forecast model. Nevertheless it may be possible to tune the tolerance of the assimilation system using a few experiments. To gain insight into how to do this, we note that the bound in (16) is essentially a measure of the nonlinearity of the problem. The norm $\| (\mathbf{J}(\mathbf{x}^{(k)})^T\mathbf{J}(\mathbf{x}^{(k)}))^{-1}Q(\mathbf{x}^{(k)}) \|_2$ measures the size of the second order terms of the Hessian relative to the first order terms. The more nonlinear the problem, the higher this norm will be. This will lead to a smaller bound on the right hand side of (16), implying that the inner loop must be solved more accurately. If we assume that within the assimilation system we have a similar degree of nonlinearity for a given time window, then we expect that this bound will not vary too widely and simple experiments should be sufficient to provide a suitable tolerance. We now demonstrate the application of this theory in some idealised assimilation experiments.

4 Experimental system

4.1 Numerical model and assimilation system

The model we use for this study is that described in Lawless *et al.* (2003). We consider a one-dimensional shallow water model for flow over an obstacle in the absence of rotation. The continuous system is described by a momentum equation and a mass-continuity equation,

$$\frac{\mathrm{Du}}{\mathrm{Dt}} + \frac{\partial \phi}{\partial x} = -g\frac{\partial \bar{h}}{\partial x}, \tag{22}$$

$$\frac{\mathrm{D}(\ln \phi)}{\mathrm{Dt}} + \frac{\partial u}{\partial x} = 0, \tag{23}$$

with

$$\frac{\mathrm{D}}{\mathrm{Dt}} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x}. \tag{24}$$

In these equations $\bar{h} = \bar{h}(x)$ is the height of the bottom orography, u is the velocity of the fluid and $\phi = gh$ is the geopotential, where g is the gravitational constant and $h > 0$ the depth of the fluid above the orography. The problem is defined on the domain $x \in [0, L]$, with periodic boundary conditions, and we let $t \in [0, T]$. The model is discretized using a semi-implicit semi-Lagrangian scheme, as described in Lawless *et al.* (2003).

An incremental 4D-Var assimilation scheme is set up for this model, as described in Section 2.1. For the numerical experiments reported in this paper, no background term is included and sufficient observations are made available to determine a unique solution. Further details of the assimilation scheme can be found in Lawless *et al.* (2005a). In that study two different linearizations of the nonlinear model were compared in assimilation experiments. Here we follow the standard method and use the exact tangent linear model. The inner loop cost function is minimized using a conjugate gradient method. For this study we compare the three different inner loop stopping criteria introduced in Section 3.

4.2 Details of experiments

Identical twin experiments are used to compare the different stopping criteria. The model is set up using a periodic domain with 200 grid points, with a spacing of $\Delta x = 0.01\ m$, so that $x \in [0\ m, 2\ m]$. An orography is defined in the centre of the domain by

$$\bar{h}(x) = \bar{h}_c \left(1 - \frac{x^2}{a^2}\right) \quad \text{for} \quad 0 < |x| < a, \tag{25}$$

and $\bar{h}(x) = 0$ otherwise, with $\bar{h}_c = 0.05\ m$ and $a = 40\Delta x = 0.4\ m$. The gravitational constant g is set to $10\ ms^{-2}$ and the model time step Δt is $9.2 \times 10^{-3}\ s$.

For the assimilation experiments we use an assimilation window of 50 time steps unless stated otherwise. The initial conditions at time $t = 0$ are the same as those of Case II of Lawless *et al.* (2005a). This is a highly nonlinear test case in which a shock solution is formed in both the u and ϕ fields. The true solution at the start and end of the assimilation window is shown in Figure 1. The first guess field is taken to be the truth with a phase error of $0.5\ m$. The setting of the convergence criteria for the inner loop are detailed in the description of each experiment. For all experiments the outer loops are stopped when the absolute norm of the gradient of the updated outer loop cost function is less than the value of ϵ specified for the inner loop criterion. The number of outer loops is limited to a maximum of 12.

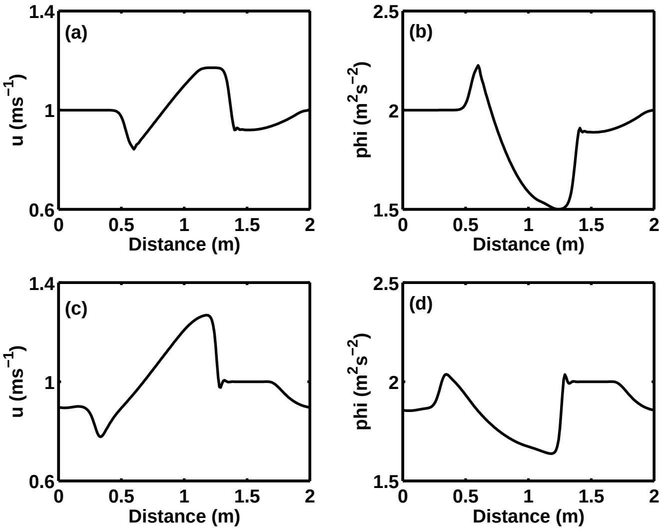

Figure 1: True solution for (a) u and (b) ϕ at time $t = 0$ and for (c) u and (d) ϕ after 50 time steps.

5 Comparison of different stopping criteria

In this section we compare the three stopping criteria introduced in Section 3. For these experiments the convergence tolerance ϵ is set to a value of 0.1. This is larger than is usually used in practice, but is chosen to provide a clear comparison without too many inner iterations.

5.1 Perfect observations

We begin with an experiment in which we have perfect observations of both u and ϕ on every time step and at every spatial point. The convergence of the cost function and its gradient is shown in Figure 2 for each of the three different convergence criteria. In this figure and other convergence plots in this paper, the values are plotted for each inner iteration and the value before the first inner iteration in each outer loop (when the perturbation is zero) is also shown. Since the inner and outer loop cost functions are identical for a zero perturbation, the function and gradient values before the inner loops are equal to the outer loop values. In the figures these iterates which correspond to the start of a new outer loop are indicated by circles. The number of iterations shown is therefore the total number of outer and inner loops.

We first consider the relative computational costs of the three experiments, noting that for this system, with the inner loop at the same spatial resolution as the outer loop, the inner and outer iterations are of comparable cost. We see from Figure 2 that the experiment using criterion (C1) takes the most iterations to converge. After the first few iterations of each inner loop the function value remains almost constant, while the gradient norm continues to decrease. Thus many inner iterations are performed which have little effect on reducing the inner loop cost function. On the other hand, when using the new stopping criterion (C3), the assimilation converges in many fewer iterations. In the initial outer loops the inner loop is truncated earlier than with criterion (C1). This has the effect of avoiding plateaus in the convergence of the cost function and its overall convergence is faster.

We note that the convergence as measured by the norm of the inner loop gradient appears to be slowed down initially when using criterion (C3) instead of (C1). However, the gradient convergence is much smoother with criterion (C3), whereas with criterion (C1) large jumps are observed in the gradient at the start of each new outer loop. Moreover, if we fix attention on the iterations corresponding to new outer loops (indicated by the circles on the figure), we see that the outer loop gradient is in fact converging faster with criterion (C3) than with criterion (C1). This indicates that when using criterion (C1), although the inner loop cost function is being minimized more accurately, this cost function is not a good approximation to the nonlinear problem. Thus, when the approximation is refined by performing a new outer loop and redefining the inner loop cost function, the gradient norm increases again by two orders of magnitude. This does not occur in the assimilation using stopping criterion (C3), which does not solve the approximate problem too accurately and so avoids large jumps in the gradient. Hence the gradient convergence also indicates criterion (C3) as the better option to choose.

For the experiment using criterion (C2), the convergence is initially very similar to that using criterion (C3), but after approximately 12 iterations, the convergence slows down. After this point the stopping criterion causes the inner loop to be truncated too soon, before much progress can be made in reducing the cost function and its gradient. In fact, for this experiment, the maximum number of 12 outer

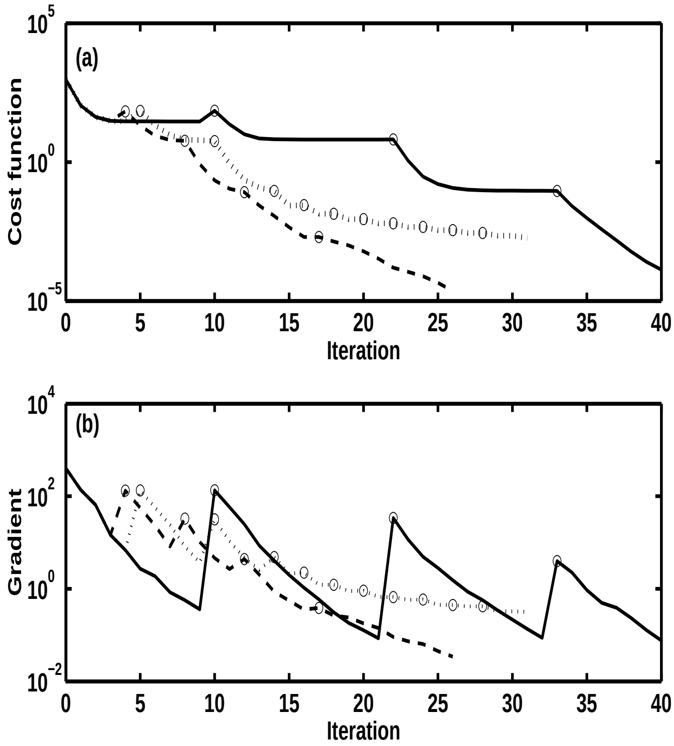

Figure 2: Convergence of (a) cost function and (b) gradient for stopping criteria (C1) (solid line), (C2) (dotted line) and (C3) (dashed line). The circles indicate a new outer loop iteration.

loops is used, but the final values of the cost function and gradient are higher than in the other two experiments.

Further experiments were performed using different values of the tolerance ϵ with criteria (C1) and (C2), to see whether the convergence could be improved (figures not shown). For all cases tested, however, the convergence was still slower than the experiment using criterion (C3) with $\epsilon = 0.1$. Thus criterion (C3) appears to show the best performance, giving convergence in the least number of iterations and a smooth convergence of both the cost function and its gradient.

5.2 Imperfect observations

To understand how well this conclusion holds with imperfect observations, we add random Gaussian noise to the observations of both u and ϕ, with a standard deviation of 5% of the mean value of the variable. The convergence for all three criteria is

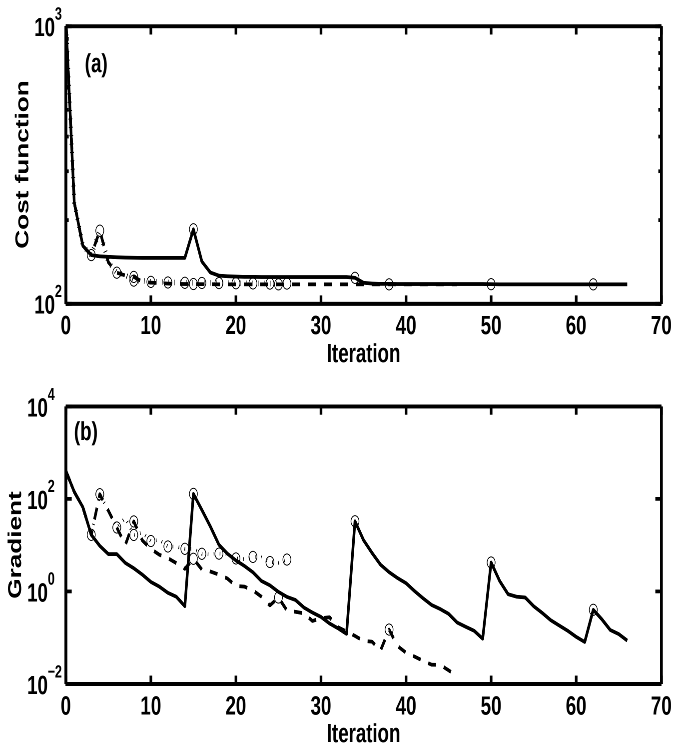

Figure 3: As Figure 2, but for imperfect observations.

shown in Figure 3. In this case the cost function itself has a large decrease initially, with subsequent iterations slowly providing further reduction. For this case perhaps the convergence of the gradient norm is more revealing. Again we find that the relative gradient criterion (C3) provides the fastest and smoothest overall convergence. As for the case with perfect observations, with criterion (C1) too many iterations are performed in the initial inner loops, which gives a large decrease in the gradient of the inner loop cost function, but an almost stagnated convergence of the cost function itself. Instead, when stopping criterion (C3) is used, the minimizations of the initial inner loop cost functions are solved less accurately, so that a faster overall convergence of the cost function and a smoother convergence of the gradient are obtained. With stopping criterion (C2) the assimilation again shows good convergence initially, but then convergence slows down and, as before, the maximum number of outer loops is performed. By comparison, the experiments with criteria (C1) and (C3) use 5 and 6 outer loops respectively. Thus the experiments using imperfect observations show the same qualitative behaviour as those using perfect observations, with the new stopping criterion (C3) showing the best convergence with the least amount of work.

5.3 Premature termination of outer loops

In practical data assimilation the outer loops are not usually run to complete convergence, so it is important to understand the effect of the different inner loop stopping criteria where the outer loops are stopped prematurely. For example, if limited computing resources are available, then it may be necessary to stop the assimilation after a given amount of computation. If the inner and outer iterations are of similar cost, this corresponds to stopping after a fixed number of iterations. From Figures 2 and 3 we see that after the first few iterations the assimilation using stopping criterion (C3) always gives a lower value of the cost function than the assimilations using the other criteria. It is possible in such a case that the inner-loop gradient norm will be lowest in the assimilation using stopping criterion (C1), for example if the experiment with imperfect observations is stopped after 10 iterations. However, when the solution to the inner problem is added onto the linearization state to give a new outer iterate, the gradient of the nonlinear cost function may be very different. In fact, we see that if we choose to stop the assimilation after a fixed number of outer loops (as is often done in practice), then not only is the cost function highest for the experiment using criterion (C1), but the gradient norm is also highest for this experiment.

5.4 Summary

From the experiments in this section the new stopping criterion (C3) is seen to be the most useful of the three convergence criteria proposed. It avoids the performance of excessive inner iterations, while providing faster convergence of the outer loop cost function and gradient. This conclusion seems to hold whether the outer loops are converged or whether they are stopped after a fixed amount of computational work. Hence the numerical experiments support the conclusion from the Gauss-Newton theory that a relative gradient stopping criterion is a more natural way of stopping the inner loop in incremental 4D-Var. However, the usefulness of this criterion depends on a good choice of the tolerance. We consider this question in the next section.

6 Choosing the value for the tolerance

From Theorem 1 we have a clear way of defining the tolerance for stopping criterion (C3), given by the right hand side of (16). However, it is unlikely that this bound could ever be calculated, except for a very simple model, since the second order terms $Q(\mathbf{x})$ require the second derivative of the nonlinear model with respect to the state variables. In this section we consider whether a suitable tolerance can be found from a sample of a few experiments.

6.1 Perfect observations

We begin by running again the experiment with perfect observations from Section 5, where we use stopping criterion (C3) with different tolerances ϵ. In Figure 4 we show the convergence of the cost function and its gradient for values of inner loop tolerance $\epsilon = 0.01$, 0.1, 0.5 and 0.9. The number of outer loops completed for each of these experiments is 5, 5, 6 and 12. We find that the best overall convergence is given when ϵ is chosen to be 0.1 or 0.5. If the tolerance is chosen to be too small, then excessive inner iterations are performed and the convergence rate is worse. This can be seen from the convergence when $\epsilon = 0.01$. In this case, although the number of outer loops is the same as for the experiment with $\epsilon = 0.1$, the total number of iterations performed is much higher. On the other hand, if the tolerance is chosen to be too large, then the inner loop may not be solved accurately enough. This can lead to too many outer loops being performed, with the possibility that they do not converge. An example of this is seen in the experiment with $\epsilon = 0.9$, which shows a very slow convergence rate, even though 12 outer loops are performed. However, it is clear that from a few experiments a tolerance can be chosen which gives a good rate of convergence. For this system the experiments performed indicate that the optimal value of the tolerance is of the order 0.1–0.5. We now investigate whether, having found a tolerance for a given system, this value will also give good convergence with different data.

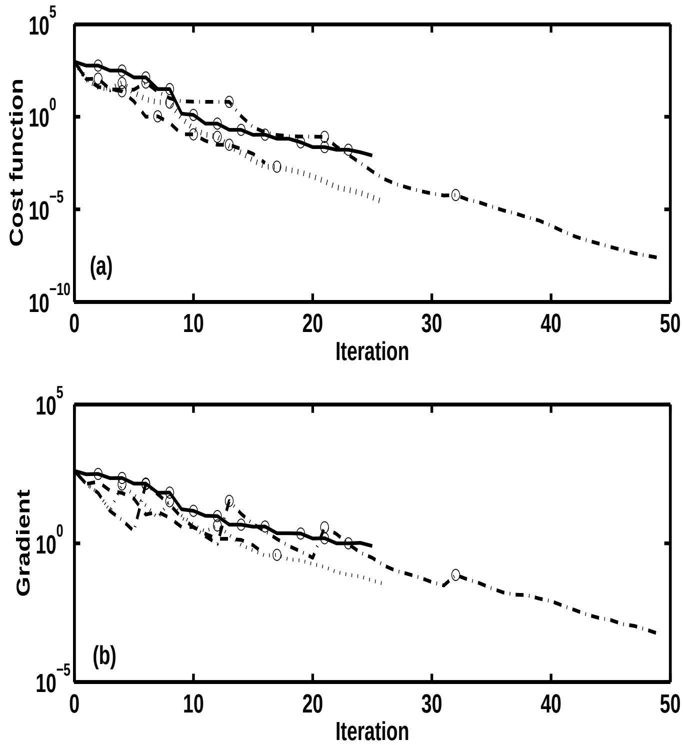

Figure 4: Convergence of (a) cost function and (b) gradient for experiment with perfect observations using criterion (C3). The different curves are for $\epsilon = 0.01$ (dot-dashed), 0.1 (dotted), 0.5 (dashed) and 0.9 (solid). The circles indicate a new outer loop iteration.

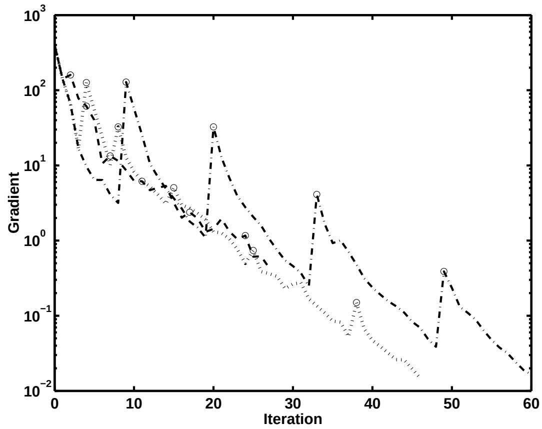

Figure 5: Convergence of gradient for experiment with imperfect observations with 5% standard deviation. The different curves are for $\epsilon = 0.01$ (dot-dashed), 0.1 (dotted) and 0.5 (dashed). The circles indicate a new outer loop iteration.

6.2 Imperfect observations

It is important to know if a tolerance chosen from a few experiments can provide good convergence when we have different levels of noise on the observations. To understand this we run the assimilation again with random, Gaussian noise added to the observations, firstly with a standard deviation of 5% of the mean values of the fields, then with a standard deviation of 10%. For these cases we find that the initial convergence is very much slowed down for $\epsilon = 0.9$ and so we compare only values of $\epsilon = 0.01, 0.1$ and 0.5. The plots for the convergence of the gradients in these two cases are shown in Figures 5 and 6. For both experiments we again find that when $\epsilon = 0.01$ the gradient is slow to converge and an excessive number of inner iterations are performed. In the figures only the first 60 iterations are shown for clarity, but the total number reaches 85 for the 5% error case and 198 for the 10% error case. The curves for values of $\epsilon = 0.1$ or 0.5 show good convergence.

If we compare these results with the perfect observation case, we see that choosing a very low value of ϵ leads to too many iterations, whether or not there is error on the observations. For higher values of ϵ we can achieve fast overall convergence, provided that the tolerance is not chosen to be too high. A comparison of the different experiments indicates that a smaller value of the tolerance may be needed when there is more noise on the observations. In the experiments performed larger

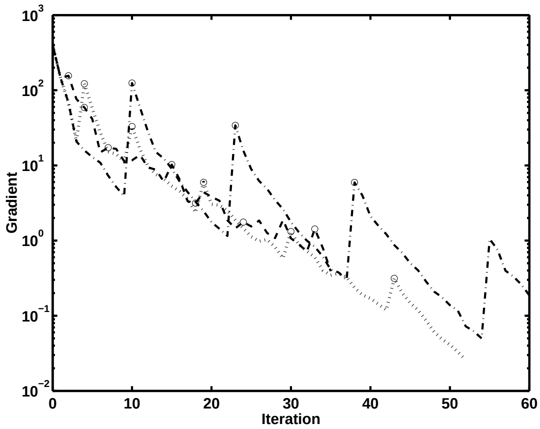

Figure 6: As Figure 5, but for imperfect observations with 10% standard deviation.

tolerances were permitted with perfect observations than for imperfect observations. Nevertheless, the results do indicate that by performing a few experiments for a given system, a suitable value of the tolerance may be chosen which reduces the total number of iterations while still ensuring convergence. Provided that the lower value of ϵ which gives good convergence is chosen, this value may be used when the noise on the observations changes.

7 Low resolution inner loop

In practical data assimilation the inner loop of incremental 4D-Var is often run at a lower spatial resolution than the nonlinear model used to calculate the innovation vectors $\mathbf{d}_i$. Although Theorem 1 does not then apply directly, other theorems we have proved for the Gauss-Newton method with inexact Jacobians (Gratton *et al.* 2004) indicate that the relative gradient criterion (C3) is a more natural stopping criterion for this case also. In order to investigate this, we run an assimilation experiment as before with a 50 time step assimilation window and perfect observations of all variables at each time step, but with the inner loop at half the spatial resolution of the outer loop. In Figure 7 we show the convergence of the cost function using the different stopping criteria with a tolerance of $\epsilon = 0.1$. As for the case with a full resolution inner loop, we find that using criterion (C3) gives a much faster overall convergence compared to using criterion (C1). Again we find that the convergence using criterion (C2) follows that of (C3) initially, but then slows down. The relative

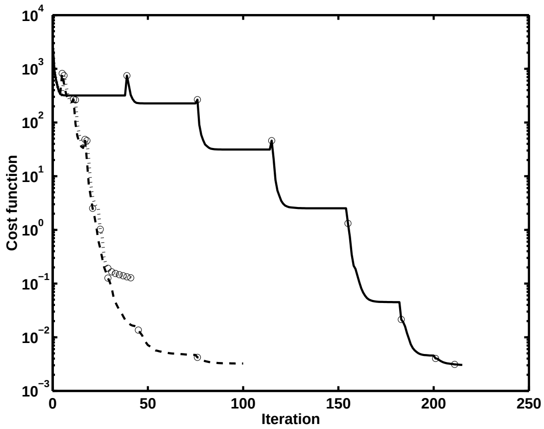

Figure 7: Convergence of cost function for a low resolution inner loop. The different curves are for stopping criteria (C1) (solid line), (C2) (dotted line) and (C3) (dashed line). The circles indicate a new outer loop iteration.

function criterion (C2) stops the inner loop too soon, which leads to the maximum number of outer loops being performed.

We repeat the experiment with imperfect observations, using random Gaussian noise with a standard deviation of 5% of the mean value of the variables. The convergence of the cost function and its gradient is shown in Figure 8. We see that the cost function behaves in a similar way to the case for a full resolution inner loop, shown in Figure 3, with an initial large decrease followed by further small reductions. The convergence of the gradient again provides a clearer comparison. We see the same pattern of behaviour as with the other experiments in this study, with criterion (C3) providing a much faster and smoother convergence of the gradient compared to the other two criteria. Further testing of criterion (C3) with different levels of tolerance also shows the same qualitative behaviour as with a full resolution inner loop, with a tolerance of $\epsilon = 0.1$ providing a good rate of convergence. Thus the new stopping criterion we have proposed appears to be suitable even when the inner loop resolution is reduced, whether perfect or imperfect observations are used.

8 Conclusion

The choice of stopping criterion for the inner loop of an incremental 4D-Var system can greatly influence its overall performance. If the inner loop is stopped too early

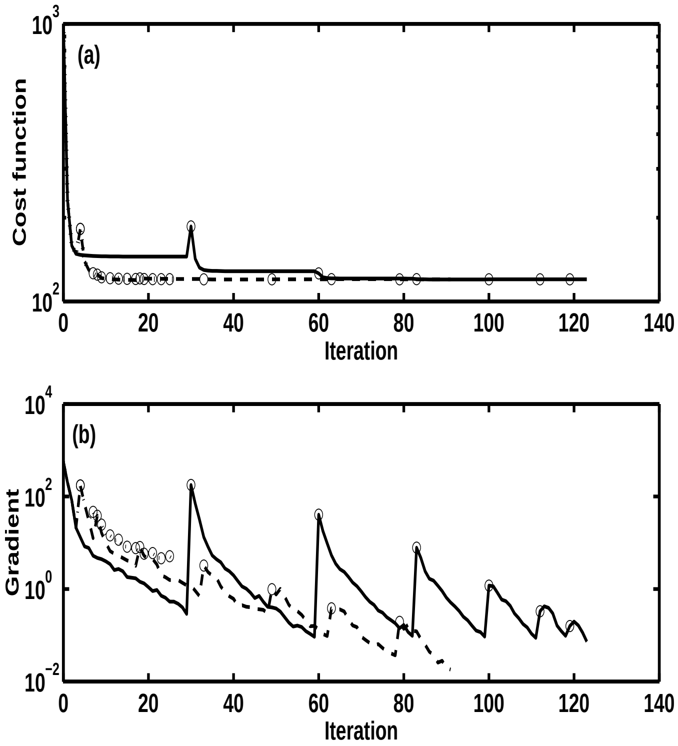

Figure 8: Convergence of (a) cost function and (b) gradient for a low resolution inner loop with imperfect observations. The different curves are for stopping criteria (C1) (solid line), (C2) (dotted line) and (C3) (dashed line). The circles indicate a new outer loop iteration.

then the final solution may not be close to the minimum of the nonlinear cost function. However, if the inner minimization is solved to too great an accuracy, much computational effort is wasted for very little gain. In this study we have proposed a new criterion for stopping the inner loop, which uses the relative change in the gradient of the inner loop cost function. This criterion is founded on our earlier work which analysed theoretically the incremental 4D-Var algorithm (Gratton *et al.* 2004). It is designed to guarantee convergence of the outer loops, with an appropriately chosen tolerance, while avoiding excessive iterations within the inner loop minimizations.

In data assimilation experiments using an idealised model we have shown how this new criterion provides a faster and smoother convergence than other criteria which are commonly used. Furthermore, we have demonstrated how the tolerance may be tuned with a few experiments to give good convergence in different situations, such as having different levels of noise on the observations. The criterion has also been shown to give good results when the inner loop is at a lower spatial resolution than the outer loop. We therefore conclude that a relative gradient stopping criterion is a very appropriate way to stop the inner loop of an incremental 4D-Var system.

References

Courtier, P. and Talagrand, O., (1987). Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Q. J. R. Meteorol. Soc.*, **113,** 1329–1347.

Courtier, P., Thépaut, 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.

Dennis, J.E. and Schnabel, R.B., (1996). *Numerical methods for unconstrained optimization and nonlinear equations*, Society for Industrial and Applied Mathematics.

Gill, P.E., Murray, W. and Wright, H.R., (1986). *Practical optimization*, Academic Press.

Gratton, S., Lawless, A.S. and Nichols, N.K., (2004). Approximate Gauss-Newton methods for nonlinear least squares problems, Numerical analysis report 9/04, Department of Mathematics, The University of Reading. *Submitted for publication.*

Laroche, S. and Gauthier, P., (1998). A validation of the incremental formulation of 4D variational data assimilation in nonlinear barotropic flow. *Tellus*, **50A,** 557–572.

Lawless, A.S., Nichols, N.K. and Ballard, S.P., (2003). A comparison of two methods for developing the linearization of a shallow-water model. *Q. J. R. Meteorol. Soc.*, **129,** 1237–1254.

Lawless, A.S., Gratton, S. and Nichols, N.K., (2005a). An investigation of the convergence of incremental 4D-Var using non-tangent linear models. *Q. J. R. Meteorol. Soc.*, **131,** 459–476.

Lawless, A.S., Gratton, S. and Nichols, N.K., (2005b). Approximate iterative methods for variational data assimilation. *Int. J. Numer. Methods in Fluids*, **47,** 1129–1135.

Lawless, A.S., Gratton, S. and Nichols, N.K., (2005c). An investigation of the convergence of incremental 4D-Var. *Proceedings of the fourth WMO international symposium on assimilation of observations in meteorology and oceanography*, Prague, April 2005, p. 150.

Lorenc, A.C, Ballard, S.P., Bell, R.S., Ingleby, N.B., Andrews, P.L.F., Barker, D.M., Bray, J.R., Clayton, A.M., Dalby, T., Li, D., Payne, T.J. and Saunders, F.W., (2000). The Met. Office global 3-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.*, **126,** 2991-3012.

Mahfouf, J.-F. and Rabier, F., (2000). The ECMWF operational implementation of four dimensional variational assimilation II: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.*, **126,** 1171–1190.

Rabier, F. and Courtier, P., (1992). Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, **118,** 649–672.

Rabier, F., Thépaut, J.-N. and Courtier, P., (1998). Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Q. J. R. Meteorol. Soc.*, **124,** 1861–1887.

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A., (2000). The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126,** 1143–1170.

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

Thépaut, J.N., Hoffman, R.N. and Courtier, P., (1993). Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.*, **121,** 3393–3414.

Weaver, A.T., Vialard, J. and Anderson, D.L.T., (2003). Three and four-dimensional variational assimilation with a general circulation model of the Tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks. *Mon. Wea. Rev.*, **131,** 1360–1378.