The University of Reading

An investigation of incremental 4D-Var using non-tangent linear models

A.S. Lawless[1], S. Gratton[2] and N.K. Nichols[1]

NUMERICAL ANALYSIS REPORT 2/04

[1]*Department of Mathematics*
The University of Reading
Whiteknights, PO Box 220
Reading
Berkshire RG6 6AX

[2]*CERFACS*
42 Avenue Gustave Coriolis
31057 Toulouse CEDEX
France

Department of Mathematics

Abstract

We investigate the convergence of incremental four-dimensional variational data assimilation (4D-Var) when an approximation to the tangent linear model is used within the inner loop. Using a semi-implicit semi-Lagrangian model of the one-dimensional

Contents

1 Introduction

In order to make four-dimensional variational assimilation (4D-Var) operationally affordable, Courtier *et al.* (1994) proposed an incremental version of the problem whereby the minimization of the full nonlinear cost function is approximated by a series of minimizations of quadratic cost functions with linear constraints. These are derived by assuming that the evolution of small perturbations to a given base trajectory can be approximated using a linear model. Usually this linear model is taken to be the tangent linear model (TLM) of the discrete nonlinear model, but Courtier *et al.* propose that other linear models may also be used, provided that they are close to the tangent linear model.

More recently Lawless *et al.* (2003) (hereafter LNB) looked at another method of developing a linear model, which starts from the continuous linearized equations. These equations are discretized to form a perturbation forecast model (PFM). Such a model is being developed as part of the incremental 4D-Var scheme of the UK Met Office (Lorenc *et al.* 2000). The results of LNB showed that a PFM could adequately describe the evolution of a perturbation in the discrete nonlinear model, provided that the perturbation was of a reasonable size.

In the present study we consider the use of perturbation forecast models within the inner loop of an incremental 4D-Var system. We examine whether the use of a PFM within such a system will degrade the assimilation with respect to using a TLM, either in terms of the convergence rate or in terms of the final analysis. In particular, we address the difference between a TLM and a PFM for very small perturbations and assess how this affects the assimilation close to the point of convergence.

We begin in section 2 by describing the continuous and discrete models used in this study. The incremental 4D-Var assimilation system is then presented in section 3. In section 4 we perform a series of assimilation experiments to compare the performance using a TLM and a PFM, using both perfect observations and observations with error. These results are then discussed in section 5, where we interpret incremental 4D-Var as a Gauss-Newton iteration. By presenting the incremental procedure in this context, we are able to understand more fully our numerical results. Finally we draw some conclusions in section 6.

@u
Γ □0

2 Description of the models

2.1 Nonlinear model

The model we consider is the one-dimensional shallow water system for the flow of a fluid over an obstacle in the absence of rotation. The continuous problem is described by the equations

$$\frac{\mathrm{Du}}{\mathrm{Dt}} + \frac{@Á}{@x} = -g\frac{@\bar{h}}{@x}; \tag{1}$$

$$\frac{\mathrm{D}(\ln Á)}{\mathrm{Dt}} + \frac{@u}{@x} = 0; \tag{2}$$

with

$$\frac{\mathrm{D}}{\mathrm{Dt}} = \frac{@}{@t} + u\frac{@}{@x}: \tag{3}$$

In these equations $\bar{h} = \bar{h}(x)$ is the height of the bottom orography, u is the velocity of the fluid and $Á = gh$ 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 system is discretized using a two-time-level semi-implicit semi-Lagrangian integration scheme, based on the scheme of Temperton and Staniforth (1987). We use a grid staggering in the spatial domain, with points at which u is held being half a grid length from points at which $Á$ is held. In the discrete equations subscripts au and du indicate the arrival and departure points for the u variable and similarly $aÁ$ and $dÁ$ the arrival and departure points for $Á$. The time discretization is then given by

$$\frac{u^{n+1}_{au} - u^{n}_{du}}{\Delta t} + (1 - ®_1)\left(\frac{@Á}{@x} + g\frac{@\bar{h}}{@x}\right)^{n+1}_{au} - TD[f6.9816=TJ-4191T43.96TD[(n)244w1.

2.2 Linear models

We develop both the tangent linear model and a perturbation forecast model for this system, in order to compare the two within data assimilation experiments. The TLM is derived directly from the nonlinear model source code, using the normal procedure of automatic differentiation. The only exception to this rule is in the treatment of the iterative procedure used to solve the elliptic equation. For this part of the solution we solve the linearized equation within the TLM rather than differentiating the iterative procedure. Further details of the resulting numerical scheme can be found in Lawless (2001).

For the PFM we must first linearize the continous nonlinear equations (1), (2) to find the continuous linearized equations. Considering perturbations $\pm u(x;t); \pm \acute{A}(x;t)$ around a state $\bar{u}(x;t); \bar{\acute{A}}(x;t)$ which satisfies the nonlinear equations, we obtain for the linearization of the momentum equation

$$\frac{D\pm u}{Dt} + \pm u\frac{@\bar{u}}{@x} + \frac{@\pm\acute{A}}{@x} = 0 \qquad (6)$$

and for the linearization of the continuity equation

$$\frac{D}{Dt}\left(\frac{\pm\acute{A}}{\bar{\acute{A}}}\right)$$

subject to

$$\pm x^{(k}$$

section 2(c), and the implementation of both has been verified using the gradient test of Navon *et al.* (1992).

It is necessary to provide some criterion to determine when the inner loop iterations have converged sufficiently. In this study the iteration is stopped if the change in the cost function from one iteration to the next is less than a prescribed tolerance. This is defined by the test

$$\tilde{\mathcal{J}}^{(k)}_{(l+1)} - \tilde{\mathcal{J}}^{(k)}_{(l)} < \epsilon(1 + \tilde{\mathcal{J}}^{(k)}_{(l)}) \tag{15}$$

where l is the iteration count of the inner loop and ϵ is a small parameter. The reason for the addition of one on the right hand side is to ensure that when $\tilde{\mathcal{J}}$ itself is less than one, the convergence criterion does not fall to less than order ϵ (Gill *et al.* 1986, p.306).

We note that within the incremental formulation of 4D-Var it is possible to run the inner loop at a lower resolution than the outer loop. In this case the innovation vectors $\mathbf{d}_i$ are still calculated at the higher resolution, using a high resolution run of the nonlinear model. However the increment $\pm\mathbf{x}_i$ is evolved using the linear model at a lower resolution. The analysed increment at the end of each outer loop iteration must then be interpolated back to the higher resolution to perform the update step (14).

4 Numerical experiments

4.1 Experimental design

In order to investigate the behaviour of a TLM and a PFM within incremental 4D-Var we perform a series of identical twin experiments. We consider two different experimental designs, one in which the true evolution is only weakly nonlinear during the assimilation period and one in which the evolution becomes highly nonlinear. We refer to these as Case I and Case II respectively. For Case I we use a periodic domain of 1000 grid points, with a spacing $\Delta x = 0.01\ m$ between them, so that $x \in [0\ m; 10\ m]$. For Case II we use 200 grid points, also with a spacing $\Delta x = 0.01\ m$, so that $x \in [0\ m; 2\ m]$. For both cases we define an orography in the centre of the domain by

$$\bar{h}(x) = \bar{h}_c \left(1 - \frac{x^2}{a^2}\right) \quad \text{for} \quad 0 < |x| < a; \tag{16}$$

and $\bar{h}(x) = 0$ otherwise, where we choose $\bar{h}_c = 0.05\ m$ and a is taken to be $40\Delta x = 0.4\ m$. The time-weightings for the scheme are taken to be $\alpha_1 = \alpha_2 = 0.6$ and for the PFM

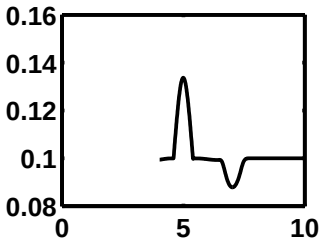


Figure 1: Initial conditions at time $t = 0$. The plots show (a) u and (b) $Á$ for Case I and (c) u and (d) $Á$ for Case II.

$®_3 = ®_4 = 0{:}6$. The gravitational constant g is set to 10 $ms^{i\,2}$ and the model time step Δt is $9{:}2 \times 10^{i\,3}$ s

The assimilation interval for Case I is taken to be 100 time steps, and for Case II it is 50 time steps. Figure 1 shows the initial conditions at time $t = 0$ for each of these cases. For the assimilation experiments we take the first guess field at time $t = 0$ to be the true solution shifted left by $0{:}5$ m, reflecting a phase error seen in a forecast background field.

We illustrate the comparative behaviour of the TLM and PFM by comparing the evolution of a perturbation in the linear models with its evolution in the nonlinear model. We define a state x_0

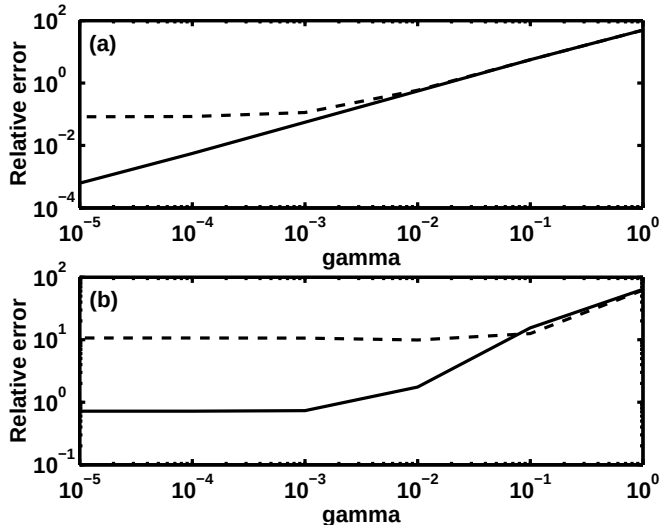


Figure 2: Plot of relative error E_R of u field against perturbation size for (a) Case I and (b) Case II. The solid line is for the tangent linear model and the dashed line for the perturbation forecast model.

shock. The error in the PFM is similar to that for the TLM for larger perturbations, but as found in LNB, the error for small perturbations is larger. We now investigate whether this difference between the TLM and the PFM for small perturbations will affect the convergence of an incremental 4D-Var scheme.

4.2 Assimilation experiments

We first perform an identical twin experiment in which there is no background term in the cost function and perfect observations are given on every time step and at every grid point. Hence the observation error covariance matrices $\mathbf{R}_i$ and the observation operators $\mathbf{H}_i$ are both equal to the identity for each time step. The inner loop is kept to be the same spatial resolution as the outer loop. Since we wish to test the effect caused by the behaviour of the difference in the linear models for very small perturbations, we run a total of 12 outer loops, thus ensuring that in later loops the perturbations being solved for are small. The iterations of the inner loop are stopped when the criterion (15) is satisfied. For this experiment the convergence parameter ^{2} is set to be 10^{-8}. The convergence of the cost function and its gradient for this experiment is shown in Figures 3 and 4 for Cases I and II respectively. We see that for both cases the convergence is almost identical whether using a TLM or a PFM.

Since we know the true solution throughout the time window by construction of the

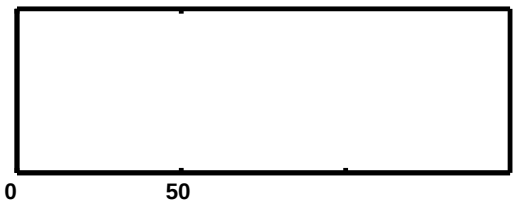


Figure 3: Case I: Convergence of (a) cost function and (b) gradient for tangent linear model (solid line) and perturbation forecast model (dashed line).

experiment, we can define the analysis error as the difference between the analysed solution and the truth. In Figure 5 we show the analysis errors for Case I for both u and $\acute{A}$ for the different assimilation runs, with the fields taken at the centre of the time window. For both the analysis using a TLM and that using a PFM, the root mean square (RMS) norm of the analysis error is of the order 10^{-8}, which is the best that we may expect for the convergence tolerance we are using, and the norm of the difference between the two analyses is of the same order. The analysis errors for Case II in the centre of the time window are shown in Figure 6. Even though the evolution for this case is highly nonlinear, with the formation of a shock, the assimilations using both linear models are able to analyse the true solution to within a high degree of accuracy and the analysis error is of order 10^{-7}. The RMS norm of the difference between the two assimilations is of the order 10^{-8} and so is within the order of the analysis error.

In order to test that the solutions around the shock remain stable as the analyses are evolved, we run a forecast of 100 time steps starting from the analysis at the start of the assimilation window. As the analysed solutions evolve we find that the errors in the forecast solutions become more confined to the region of the shock formation. In Figure 7 we show the error in the forecasts after 100 time steps in the region of the shock. At this stage almost all of the errors are around the shock position, with the maximum amplitude increasing to order 10^{-6}. However as the forecasts are continued further, the amplitude of the error decays by an order of magnitude and the system remains stable.

From these results it would appear that incremental 4D-Var is able to perform a good

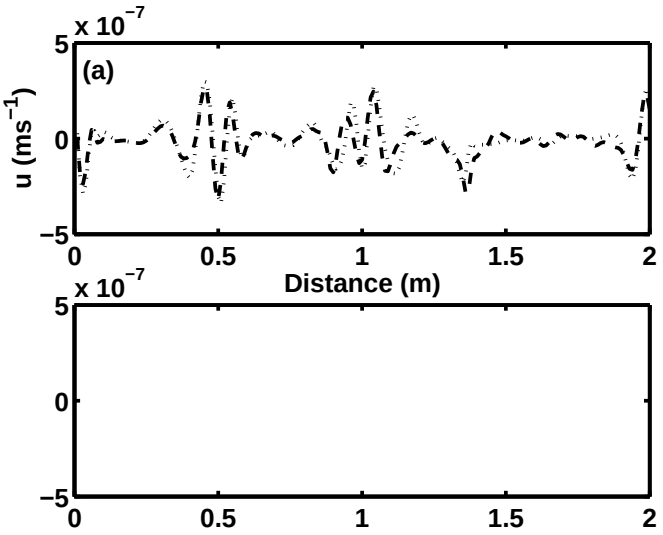


Figure 6: As for Figure 5, for Case II.

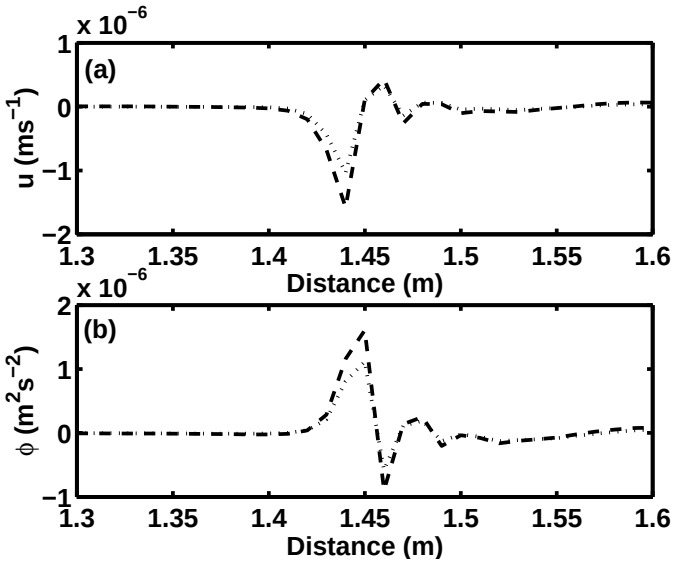
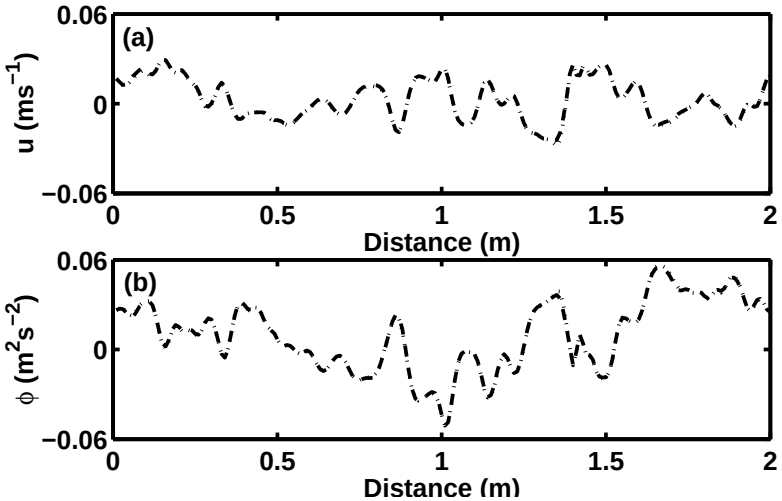


Figure 7: Case II: Error in the forecast around the shock position after 100 time steps for (a) *u* and (b) *Á*. The dotted line is for the assimilation using the TLM and the dashed line for that using the PFM.

analysis given perfect observations, using either a TLM or a PFM, even when the flow is highly nonlinear. In order to understand whether both linear models continue to be valid as the convergence is taken close to machine precision, we run again Case II using a convergence parameter $^{2} = 10^{i\,12}$ in the inner loop and running for 50 outer loops. The di□erence between the two analysed solutions is reduced by two orders of magnitude from order $10^{i\,8}$ to $10^{i\,10}$


(a)
0.06
0
−0.06
u (ms^{-1})
0
0.5
1
1.5
2
Distance (m)
(b)
0.06
0
−0.06
ϕ (m^{2}s^{-2})
0
0.5
1
1.5
2
Distance (m)

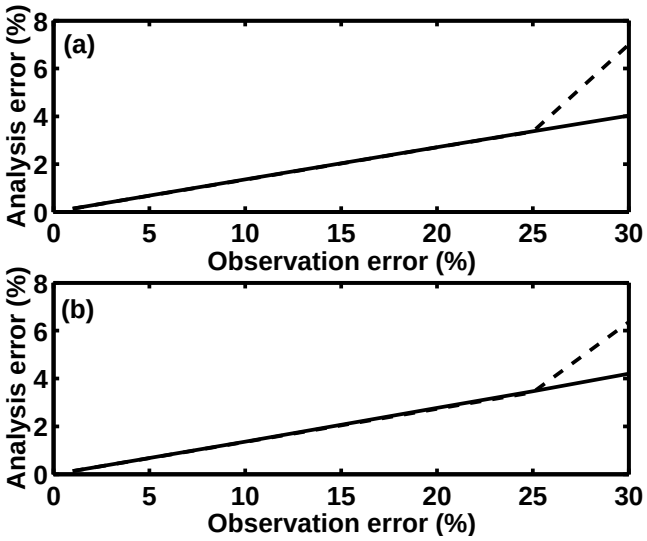


Figure 9: Case II: Plot of percentage analysis error at the centre of the time window, relative to the norm of the true solution, against level of observational noise, for (a) *u* field and (b) *Á* field. The solid line indicates the assimilation using the TLM and the dashed line that using the PFM.

increment, before it can be added on to the guess field. For both of these we use a linear interpolation. In Figure 10 we show the convergence of the cost function and its gradient for these experiments. Also shown for comparison is the convergence using the TLM at the full resolution, which corresponds to the solid curves of Figure 4. We see that the

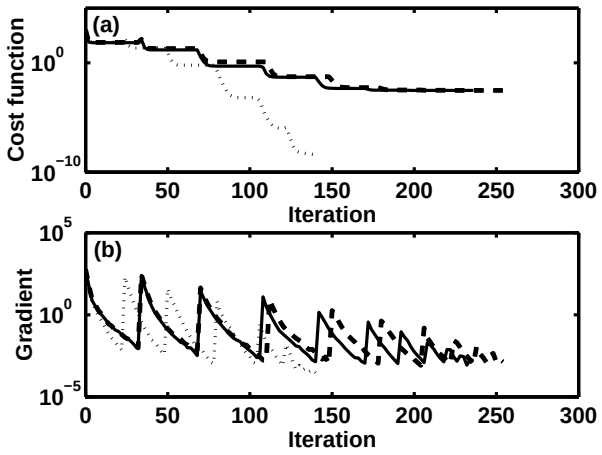


Figure 10: Case II: Convergence of (a) cost function and (b) gradient with inner loop at lower resolution. The solid line is for the tangent linear model and the dashed line for the perturbation forecast model. The dotted line shows for comparison the convergence using the tangent linear model at full resolution.

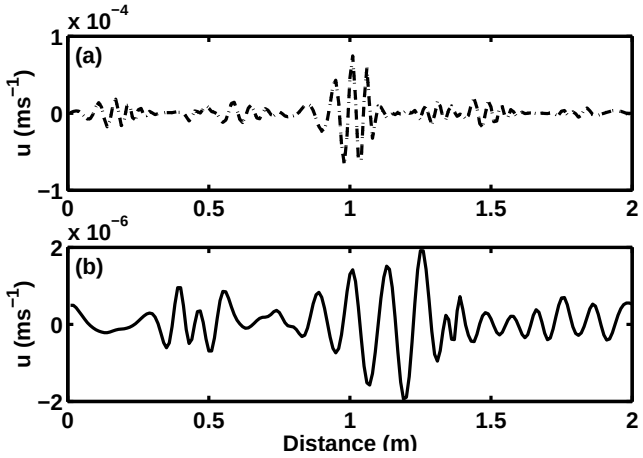


Figure 11: Case II: Analysis errors in *u* field at the centre of the time window using a low resolution inner loop. Plot (a) shows the analyis error, with the dotted line indicating the assimilation using the TLM and the dashed line using the PFM. Plot (b) shows the difference between the two analyses

Table 1: RMS norms of analysis differences using imperfect observations.

		$u(ms^{-1})$	$\acute{A}(m^2s^{-2})$
(i)	Between TLM analyses with high and low resolution inner loop	4.51×10^{-3}	6.03×10^{-3}
(ii)	Between TLM and PFM analyses with high resolution inner loop	7.75×10^{-4}	8.23×10^{-4}
(iii)	Between TLM and PFM analyses with low resolution inner loop	8.02×10^{-4}	8.75×10^{-4}

of the analysis error. For the high resolution case using a TLM the norm of the analysis error is of the order 10^{-2} for both the u and $\acute{A}$ fields. In Table 1 we detail the RMS norms of the difference in analyses caused by (i) changing from a high to low resolution inner loop using a TLM, (ii) changing from a TLM to a PFM in a high resolution inner loop

the original nonlinear cost function (10). The underlying iterative process was considered by Thépaut and Veersé (1998), who used it to derive a general form of convergence condition for incremental 4D-Var. We now show how the iteration may be interpreted as an approximation to a Gauss-Newton iteration.

The theory of a Gauss-Newton iteration for a general least squares minimization is presented in the appendix. In order to understand incremental 4D-Var from this perspective we must first write the nonlinear cost function (10) in the more general form (28). We put

$$\hat{\mathbf{d}}(\mathbf{x}_0) = -\begin{pmatrix} \mathbf{x}_0 - \mathbf{x}^b \\ H_0[\mathbf{x}_0] - \mathbf{y}_0^o \\ \vdots \\ H_n[\mathbf{x}_n] - \mathbf{y}_n^o \end{pmatrix}; \quad \mathbf{C}^{-1} = \begin{pmatrix} \mathbf{B}_0^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^{-1} \end{pmatrix}; \tag{20}$$

where $\mathbf{R}$ is the block diagonal matrix with entries $\mathbf{R}_i$ and $\mathbf{C}^{-1}$ is a symmetric positive definite matrix. Then the cost function (10) can be written

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\hat{\mathbf{d}}^{\mathrm{T}}\mathbf{C}^{-1}\hat{\mathbf{d}}. \tag{21}$$

We note that this is equivalent to the general form (28) with $\mathbf{f}(\mathbf{x}) = \mathbf{C}^{-1/2}\hat{\mathbf{d}}$. Then the Jacobian matrix of $\mathbf{f}(\mathbf{x})$ is given by

$$\mathbf{J} = \mathbf{C}^{-1/2}\hat{\mathbf{H}}; \tag{22}$$

where

$$\hat{\mathbf{H}} = -\begin{pmatrix} \mathbf{I} \\ \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{L}_1 \\ \vdots \\ \mathbf{H}_n\mathbf{L}_n \end{pmatrix} \tag{23}$$

and $\mathbf{L}_i = \mathbf{L}(t_i; t_0; \mathbf{x}^{(k)})$ is the solution operator of the exact tangent linear model.

If we were to use an exact Gauss-Newton method to minimize $\mathcal{J}(\mathbf{x})$, then from (31) and (32) we see that the this implies that for each iteration we must have

$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} - \left(\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{C}^{-1}\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^{\mathrm{T}}\mathbf{C}^{-1}\hat{\mathbf{d}}; \tag{24}$$

where $\hat{\mathbf{H}}$ and $\hat{\mathbf{d}}$ are both dependent on the current iterate $\mathbf{x}_0^{(k)}$. Expanding the variables in full shows that this is exactly the same as (19) for the case $\tilde{\mathbf{L}}(t_i; t_0; \mathbf{x}^{(k)}) = \mathbf{L}(t_i; t_0; \mathbf{x}^{(k)})$. Thus we conclude that the incremental 4D-Var iteration given by (19) is equivalent to

a Gauss-Newton iteration if an exact tangent linear model is used. If the linear model is approximated in any way, either by using a PFM instead of a TLM, or by reducing the spatial resolution of the TLM, then the incremental 4D-Var can be considered as an inexact Gauss-Newton iteration in which the Jacobian $\mathbf{J}$ is replaced by an approximation $\tilde{\mathbf{J}}$. We now make some comments on the convergence of this process.

5.2 Convergence of incremental 4D-Var

Although in practice incremental 4D-Var is only run for a few outer iterations, we can gain some useful insights by looking at what happens if we run the iterations to convergence. We assume that the iteration process has some fixed point $\mathbf{x}_0^{\alpha}$ and that sufficient conditions for convergence to this fixed point are satisfied. Then at the fixed point $\mathbf{x}_0^{\alpha}$ we have

$$\mathbf{B}_0^{-1}(\mathbf{x}_0^{\alpha} - \mathbf{x}^b) + \sum_{i=0}^{n} \tilde{\mathbf{L}}_i^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1}(H_i[\mathbf{x}_i^{\alpha}] - \mathbf{y}_i^o) = 0; \tag{25}$$

with

$$\mathbf{x}_i^{\alpha} = S(t_i; t_0; \mathbf{x}_0^{\alpha}). \tag{26}$$

We note first of all that, if $\tilde{\mathbf{L}}_i$ is equal to the exact tangent linear model $\mathbf{L}_i$, then the left hand side of (25) is equal to $\nabla \mathcal{J}[\mathbf{x}_0^{\alpha}]$. Hence in this case the fixed point of the iteration is also a stationary point of the nonlinear cost function (10).

In order to interpret the results of the experiments of section 4 we now consider the case in which no background term is present, so that at the fixed point we have

$$\sum_{i=0}^{n} \tilde{\mathbf{L}}_i^{\mathrm{T}} \mathbf{H}_i^{\mathrm{T}} \mathbf{R}_i^{-1}(H_i[\mathbf{x}_i^{\alpha}] - \mathbf{y}_i^o) = 0. \tag{27}$$

We denote the truth at time t_i by $\mathbf{x}_i^t$ and we suppose that we have perfect observations of the true state, as for the experiments of section 4(b). Then at each time t_i we have $\mathbf{y}_i^o = H_i[\mathbf{x}_i^t]$. Hence we see that $\mathbf{x}_0^{\alpha} = \mathbf{x}_0^t$ is a fixed point of the iteration, since the residual $H_i[\mathbf{x}_i^{\alpha}] - \mathbf{y}_i^o$ is zero for all times t_i and so (27) is automatically satisfied. That is, with perfect observations we have a zero-residual problem, and $\mathbf{x}_0^t$ is a fixed point of the incremental 4D-Var iteration, irrespective of the matrices $\tilde{\mathbf{L}}_i$.

We emphasize that this does not mean that an iteration with any matrices $\tilde{\mathbf{L}}_i$ will necessarily converge to this fixed point, since the convergence will depend on other conditions, including the distance of the first guess from the fixed point. However we do know that we have a fixed point equal to the true solution of the nonlinear problem. In particular, we see that by replacing a TLM with a PFM, the true solution of the nonlinear problem is

still a fixed point of the incremental 4D-Var iteration. This explains why the experiments of section 4(b) using a TLM and a PFM were able to give identical results to within the accuracy of the solution procedure, even though the two linear models behave differently for small perturbations.

We now consider what happens when the observations contain errors. In this case it will not be true in general that there exists a point $\mathbf{x}_0^a$ such that $\mathbf{y}_i^o = H_i[\mathbf{x}_i^a]$ for all times t_i. Hence the point at which (27) is satisfied will depend on the matrices $\tilde{\mathbf{L}}_i$ and we would not expect to have the same fixed point when these matrices are changed. This is reflected in the experiments of section 4(c), where the assimilations with the TLM and PFM did not have the same solution when run to complete convergence. However we did find that the solutions for the two assimilations were close. Since the fixed points must satisfy (27), where $\tilde{\mathbf{L}}_i$ are the matrices of the corresponding linear model, we may expect the analyses to be close if the matrices are not too far apart in some sense, that is if the approximations $\tilde{\mathbf{L}}_i$ are close to the true tangent linear matrices $\mathbf{L}_i$. This will be investigated further in future work.

Furthermore, since we do not have a zero-residual problem, we would also expect the observational errors to play a significant role in determining how close the fixed points are. In particular, when the error in the observations is large, then the assimilation has more freedom to fit the observations within the observational error and so the fixed points may be further apart. This is reflected in the difference in behaviour seen in Figure 9 as the observational error is increased.

6 Conclusions

This study has shown that despite the fact that a PFM may behave differently from a TLM for small perturbations, the inclusion of a PFM in an incremental 4D-Var scheme may be a valid approximation. For tests with exact observations the assimilations with a TLM and a PFM gave the same analysis to within the precision of the converged tolerance. When error was included on the observations the analyses differed, even when the incremental method was converged fully. However, the norm of the difference between the analyses using a TLM and a PFM was still found to be much smaller than the difference between either analysis and the true solution, providing that the observational noise remained below a certain level.

The difference made in replacing a TLM with a PFM was also compared with the

e□ect of using a reduced resolution TLM. For the experiments performed it was found that reducing the resolution led to a greater increase in the analysis error than the use of a PFM at either high or low resolution.

In order to understand the experimental results, the incremental 4D-Var algorithm was formulated as a Gauss-Newton iteration. This provides a clear mathematical context in which the convergence of incremental 4D-Var can be analysed. We have shown how we may expect the assimilations to converge to the same analysis in the absence of observational error, but that in general we would not expect this to occur when observational error is present. In a future paper we will address some of the more theoretical questions arising from this study, such as the convergence conditions using either a TLM or a PFM, how close the converged solutions will be for a given PFM and how quickly the iteration will converge to the solution.

Appendix: Gauss-Newton iteration

The Gauss-Newton method is an iterative method for solving a general nonlinear least squares problem of the form

$$\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2} \| \mathbf{f}(\mathbf{x}) \|_2^2 = \frac{1}{2} \mathbf{f}(\mathbf{x})^{\mathsf{T}} \mathbf{f}(\mathbf{x}); \tag{28}$$

with $\mathbf{x} \in \mathbb{R}^n$ (Dennis and Schnabel, 1996). We assume that $\mathcal{J}(\mathbf{x})$ is twice continuously di□erentiable in an open convex set $D \in \mathbb{R}^n$ and that the minimization problem (28) has a unique solution $\mathbf{x}^{\mathtt{¤}} \in D$.

The first derivative matrix of $\mathbf{f}(\mathbf{x})$ is the Jacobian matrix $\mathbf{J}$, with entries $\{\mathbf{J}\}_{ij} = @f_i(\mathbf{x})/@x_j$. Then we can write the gradient and Hessian of $\mathcal{J}(\mathbf{x})$ as

$$\nabla \mathcal{J}(\mathbf{x}) = \mathbf{J}^{\mathsf{T}} \mathbf{f}(\mathbf{x}); \tag{29}$$

$$\nabla^2 \mathcal{J}(\mathbf{x}) = \mathbf{J}^{\mathsf{T}} \mathbf{J} + Q(\mathbf{x}); \tag{30}$$

where $Q(x)$ is the second order information. We note that at the minimum point $\mathbf{x}^{\mathtt{¤}}$ we have $\nabla \mathcal{J}(\mathbf{x}^{\mathtt{¤}}) = 0$.

The Gauss-Newton iteration for solving (28) is given by

$$\pm\mathbf{x}^{(k)} = -(\mathbf{J}^{\mathsf{T}} \mathbf{J})^{i\,1} \mathbf{J}^{\mathsf{T}} \mathbf{f}(\mathbf{x}^{(k)}); \tag{31}$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \pm\mathbf{x}^{(k)}: \tag{32}$$

This is an approximation to the Newton iteration in which the second order terms of the Hessian, $Q(x)$, are neglected. It can be shown that under certain conditions, the Gauss-Newton method will converge to the minimum $x^{\ddagger}$ (Dennis and Schnabel, 1996, Wedin, 1974).

If at the minimum point we have $f(x^{\ddagger}) = 0$, then the problem (28) is referred to as a *zero-residual* problem. In this case the Gauss-Newton method is quadratically convergent. Otherwise, if the method converges, then it does so linearly (Dennis and Schnabel, 1996).

References

Chao, W.C. and Chang, L-P., (1992). Development of a four-dimensional variational analysis system using the adjoint method at GLA. Part I: Dynamics. *Mon. Weather Rev.*, 120, 1661–1673.

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.

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..

mization in meteorology. *Mon. Weather Rev.*, **115**, 1479–1502.