The behaviour of 4D-Var for a highly nonlinear system

Halina Watson

August 21, 2006

Abstract

To forecast the future state of the atmosphere, data assimilation is required in order to provide a good estimate of the initial conditions, combining observations and model predictions. To meet current operational limitations the incremental approach is being implemented to reduce the computational cost of four-dimensional variational assimilation (4D-Var) which assumes that processes are close to linear. This dissertation will investigate the behaviour of the incremental formulation as the system becomes increasingly nonlinear. We want to conduct numerical experiments when the system becomes highly nonlinear to see whether it needs to be solved with a greater accuracy. Recent theory indicates that this may be the case in order for the incremental method to provide an adequate approximation to the nonlinear system. Indeed the results of this study verify the theory, revealing that the solution to the problem is improved for increasing accuracy. However, the effect of increasing the accuracy becomes redundant after the accuracy is increased to a certain limit.

Contents

List of Figures

4

Chapter 1

Introduction

A weather forecast is based on observations of the atmosphere and models of evolution of atmospheric flow. To produce a forecast, initial conditions, constructed from the observations and model, are needed to describe the atmosphere at an initial time of the forecast.

Data assimilation is a technique required to give a good estimate of the initial conditions by combining observational and model data to produce an optimal estimate of the state of the system. The method is implemented to fill in data voids and deal with observational error. Observations cannot be used on their own due to their irregular distribution in time and space. Data assimilation is also necessary to exploit the physical laws of the atmosphere and have the ability to use data from remote sensing techniques that cannot be used directly. An analysis is the updated forecast of the state of the atmosphere. It is produced from background information or forecast of the system from an earlier time step, along with observations made at the present time step.

The aim of this dissertation is to investigate the benefit gained by increasing the accuracy for which the nonlinear problem is solved using the incremental method. To analyze the assimilation, the Lorenz model will be used in the experiments.

1.1 The general problem

Four-dimensional variational data assimilation (4D-Var) is the assimilation method which we will be concentrating on. In this dissertation we want to conduct numerical experiments for the Lorenz system when it becomes highly nonlinear to see whether the 4D-Var incremental formultion needs to approximate the solution with greater accuracy. In the incremental 4D-Var, the tangent linear model (TLM) and adjoint model are used. The incremental method is a series of minimizations of a quadratic approximation to the full 4D-Var cost function subject to the linear constraint (TLM). Each minimizations is referred to as an inner loop. After the inner minimization an outer loop is performed which uses the solution from the inner iterations to update the nonlinear trajectory. The number of inner loops performed between each outer loop relates to how accurately the inner minimizations is being solved. The inner minimization is subject to a stopping criterion, so that the number of iterations performed stops once the inner loop problem has been solved to sufficient accuracy. It will be the effect of increasing the tolerance of this stopping criterion which shall be examined as the problem becomes more nonlinear.

1.2 Outline

The next chapter gives some background, describing 4D-Var data assimilation and the incremental method. Following this the stopping criterion used for the assimilation experiments will be considered in detail. Chapter 3 will describe the Lorenz model being used for the assimilation experiments, as well as the tangent linear model. In chapter 4 we will aim to verify the tangent linear hypothesis and give results from the tangent linear tests. The assimilation tests will begin in chapter 5 for perfect observations. The result from these tests will give us an insight into the level of accuracy needed as the nonlinearities increases to approximate the problem well using the incremental method. Chapter 6 will briefly look at the effect on the results for the assimilation experiments if error is added to the observations. The final chapter concludes the dissertation with a summary and discussion of what has been investigated and the limitations in this study.

Chapter 2

Background

2.1 Full 4D-Var

distributed over a time interval $[t_0, t_n]$. The full nonlinear 4D-Var cost function for the general state of the system is given by

$$J(\mathrm{x}) = (\mathrm{x}_0 - \mathrm{x}_0^b)^T \mathrm{B}^{-1}(\mathrm{x}_0 - \mathrm{x}_0^b) + \sum_{j=0}^{n} (\mathrm{y}_j - \mathrm{h}_j[\mathrm{x}_j])^T \mathrm{R}_j^{-1}(\mathrm{y}_j - \mathrm{h}_j[\mathrm{x}_j]) \qquad (2.1)$$

subject to the nonlinear model M

$$\mathrm{x}_{j+1} = M(t_j, \mathrm{x}_j) \qquad (2.2)$$

and for the true state of the system

$$\mathrm{y}_j = \mathrm{h}_j[\mathrm{x}_j] + {}_j \qquad (2.3)$$

The aim is to find the state vector at the initial time x_0, which minimises the variance of the analysis error whilst satisfying the model equations over the assimilation.

For a given time window, $[t_0, t_n]$, the observations are taken $n + 1$ times, the subscript j denotes the quantities at any given observation time t_j. The state vector is x_j, the observation vector at time t_j is y_j with error ${}_j$ and the background vector (or first guess) is x_0^b. For the main part of this study the observations will assumed to be perfect, so ${}_j$ is taken to be zero, see chapter 7 for further details

operational use at the Met O□ce. In the dissertation, the accuracy of the incremental method will be investigated in relation to how it alters as the model becomes more nonlinear.

2.2 Introducing incremental 4D-Var

Rather than a complete minimization of the full nonlinear cost function (2.1), the incremental method is an approximation of the full cost function by a series of minimizations of the quadratic cost functions subject to a linear model. In this study the tangent linear model (TLM) will be used. The TLM allows the cost function to be approximated by a quadratic cost function by assuming the nonlinear model is linearized. Further details will be discussed in section 3.2. Using the incremental method, inner and outer loops are carried out. The inner loops refer to the iterations which minimize each quadratic cost function. These iterations are followed by an outer loop, which use the approximations from the inner iterations to update the trajectory. The outer loops provide a better approximation of the cost function. The method be can described using an iterative algorithm [8]:

1. For the first outer iteration, $k = 0$, the background state is equal to the first iterate.

$$\mathrm{x}_0^{(0)} = \mathrm{x}_b \tag{2.6}$$

Note that the subscript refers to the time position of the state estimate, and the superscript represents the outer iteration count.

2. Run the nonlinear model M forward, keeping k fixed

$$\mathrm{x}_{j+1} = M(t_j, \mathrm{x}_j) \tag{2.7}$$

3. For the inner loop, solve the linear approximation of the cost function with respect to $\delta\mathrm{x}_0^{(k)}$. We are minimizing for all j the following cost function and then finding the optimal increment $\delta\mathrm{x}_0^{(k)}$ that gives the minimum.

$$\mathcal{J}^{(k)}[\delta\mathrm{x}_0{}^{(k)}] = \frac{1}{2}\underbrace{(\delta\mathrm{x}_0{}^{(k)} - [\mathrm{x}^b - \mathrm{x}_0{}^{(k)}])^T \mathrm{B}_0{}^{-1}(\delta\mathrm{x}_0{}^{(k)} - [\mathrm{x}^b - \mathrm{x}_0{}^{(k)}])}_{J_b}$$

$$+\frac{1}{2}\underbrace{\sum_{j=0}^{n}(\mathrm{H}_i\delta\mathrm{x}_j{}^{(k)} - \mathrm{d}_i{}^{(k)})^T \mathrm{R}_j{}^{-1}(\mathrm{H}_j\delta\mathrm{x}_j{}^{(k)} - \mathrm{d}_j{}^{(k)})}_{J_o} \tag{2.8}$$

with

$$\mathrm{d}_j^{(k)} = \mathrm{y}_j - \mathrm{h}_j[\mathrm{x}_j^{(k)}], \tag{2.9}$$

$$\delta\mathrm{x}_j^{(k)} = \mathrm{M}(t_j, \mathrm{x}_j^{(k)})\delta\mathrm{x}_0^{(k)} \tag{2.10}$$

where M is the tangent linear model, J_b the background term, J_o the observational term of the cost function and H_j is the linearization of the observation operator h_j around the state vector $\mathrm{x}_j^{(k)}$ at time t_j. The linearization of the observation operator is found using the tangent linear hypothesis;

$$\mathrm{h}(\mathrm{x}) - \mathrm{h}(\mathrm{x}_b) \approx \mathrm{H}(\mathrm{x}_b)(\mathrm{x} - \mathrm{x}_b) \tag{2.11}$$

4. The outer loop is then carried out by updating the trajectory

$$x_0^{(k+1)} = x_0^{(k)} + \delta x_0^{(k)} \tag{2.12}$$

5. Then set $k = k + 1$ and repeat the process from step 2 for the total number of iterations.

The success of the incremental procedure depends upon how well the tangent linear model approximates the nonlinear model. If the tangent linear model is a close approximation to the nonlinear model then we expect the incremental method to be an accurate estimate of the nonlinear 4D-Var problem.

The inner loop is solved using a minimization algorithm. In this study the steepest gradient method was used. The procedure works by updating the trajectory by adding a correction that is proportional to the negative value of the gradient of the cost function. It is essential to implement a stopping criterion to determine when the inner iterations have converged sufficiently. A more in depth look at this will be made in the section 2.4.

2.3 The steepest descent method

The steepest descent is an optimization algorithm, which defines the path of the minimization of the cost function. The method approaches the local minimum by trying to determine the direction for which the cost function decreases the most. It

this method is altered to give the truncated Gauss-Newton method. The algorithm goes as follows:

We consider a general nonlinear least squares problem

$$\min_x J(x) = \frac{1}{2} \| f(x) \|_2^2 = \frac{1}{2} f(x)^T f(x), \qquad (2.17)$$

with $x \in \mathbb{R}^n$, which we assume has a local minimum. We note that (2.1) can be solve for $\delta x^{(k)}$:

$$\delta x^{(k)} : (J(x^{(k)})^T J(x^{(k)}))\delta x^{(k)} = -[J(x^{(k)})^T f(x^{(k)}) + r^{(k)}], \tag{2.23}$$

Update:

$$x^{(k+1)} = x^{(k)} + \delta x^{(k)}, \tag{2.24}$$

where $r^{(k)}$ is the residual from the inner loop. In practice it may be very difficult to solve (2.23) as large systems may have to be dealt with. To avoid this computation, the solution $\delta x^{(k)}$ can be found from the inner minimization of the function

$$\mathcal{J}(\delta x^{(k)}) = \frac{1}{2}\left\| J(x^{(k)})\delta x^{(k)} + f(x^{(k)})\right\|_2^2 \tag{2.25}$$

This residual is needed to compensate for the premature termination of the minimization of the cost function as we have to take into account that the inner minimization of the cost function is not found exactly.

The following theorem gives an important result that gives conditions for the truncated Gauss-Newton (TGN) algorithm to converge [9].

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

$$\left\| r^{(k)}\right\|_2 \leq \beta_k \left\| J(x^{(k)})^T f(x^{(k)})\right\|_2, \tag{2.26}$$

where

$$\beta_k \leq \frac{\hat{\beta} - \left\|(J^T(x^{(k)}J(x^{(k)}))^{-1}Q(x^{(k)})\right\|_2}{1 + \left\|(J^T(x^{(k)}))^{-1}Q(x^{(k)})\right\|} \tag{2.27}$$

Then there exists

Chapter 3

Experimental system

3.1 The Lorenz model

The Lorenz equations were introduced by Edward Lorenz [14] to describe the chaotic nature of the atmosphere. The equations are known as a system which models the unpredictable behaviour of weather. In this study the data assimilation experiments carried out will be using the Lorenz model. This is for simplicity to demonstrate the results from the numerical experiments. Also a beneficial feature of using the Lorenz model is its sensitive dependence on the initial conditions. This describes well the chaotic system in which we live. Summing up that even if we have reasonably accurate knowledge of the initial conditions, the trajectory of the forecast diverges from the true trajectory very quickly. The set of Lorenz equations are given by the nonlinear system

$$\frac{dx}{dt} = -\sigma(x - y), \tag{3.1}$$

$$\frac{dy}{dt} = \rho x - y - xz, \tag{3.2}$$

$$\frac{dz}{dt} = xy - \beta z, \tag{3.3}$$

where $x = x(t)$, $y = y(t)$, $z = z(t)$ [14]. It should be noted that x, y and z are spectral co-ordinates. The σ, ρ, β are parameters, which have been assigned the values 10, 28 and $\frac{8}{3}$ respectively. The Lorenz model is very sensitive to the

The system is discretized using the second order Runge-Kutta method,

$$x^{k+1} = x^k - \sigma\frac{\Delta t}{2}[2(y^k - x^k) + \Delta t(\rho x^k - y^k - x^k y^k) - \sigma\Delta t(y^k - x^k)], \tag{3.4}$$

$$y^{k+1} = y^k + \frac{\Delta t}{2}[\rho x^k - y^k - x^k z^k + \rho(x^k + \sigma\Delta t(y^k - x^k)) - y^k - \Delta t(\rho x^k - y^k x^k z^k) - (x^k + \sigma\Delta t(y^k - x^k))(z^k + \Delta t(x^k y^k - \beta z^k))], \tag{3.5}$$

$$z^{k+1} = z^k + \frac{\Delta t}{2}[x^k y^k - \beta z^k + (x^k + \Delta t\sigma(y^k - x^k))(y^k + \Delta t(\rho x^k - y^k - x^k z^k)) - \beta z^k - \Delta t(x^k y^k - \beta z^k)], \tag{3.6}$$

where Δt is the time step and k is the time step index [10].

The numerical experiments are carried out with the background term from (2.8) not included and the inner loop cost function minimization is carried out using the method of steepest descent. The Runge-Kutta method is a one step method. This means that the variable at time step $n + 1$ is given in terms of the variable at time step n only. A benefit from this is that we don't have to store past history. When deciding on a value for the time step in the assimilation, we have to take into consideration that if the time step is too small then it may lead to excessive computation time. On the other hand, if we have too large a time step then we have to consider the stability of the system. Advantages of using a one-step method in comparison to the multi-step methods are that it is generally faster, due to the differences in the accuracy and computational complexity [17].

For the incremental method, the full nonlinear cost function is approximated by a series of convex minimizations using a linear model. As noted previously, in this study the TLM will be used. We now describe the TLM and verify the code for it, before discussing its role in the numerical experiements.

3.2 Tangent linear model

The nonlinear optimization problem is di□cult to solve since M is nonlinear. The

The discretization of the TLM using the Runge-Kutta method for the Lorenz model linearizes (3.4), (3.5) and 3.6) and is given as

$$\delta x^{k+1} = \delta x^k - \sigma\frac{\Delta t}{2}[2(\delta y^k - \delta x^k) + \Delta t(\rho\delta x^k - \delta y^k -$$

$$(\delta y^k x^k + \delta x^k y^k)) - \sigma\Delta t(\delta y^k - \delta x^k)], \qquad (3.8)$$

$$\delta y^{k+1} = \delta y^k + \frac{\Delta t}{2}[\rho\delta x^k - \delta y^k - \delta z^k x^k - \delta x^k z^k + \rho(\delta x^k +$$

$$\sigma\Delta t(\delta y^k - \delta x^k)) - \delta y^k - \Delta t(\rho\delta x^k - \delta y^k - \delta x^k z^k - \delta z^k x^k)], \qquad (3.9)$$

$$\delta z^{k+1} = \delta z^k + \frac{\Delta t}{2}[y^k\delta x^k + x^k\delta y^k + \Delta t(\rho x^k\delta x^k - x^k\delta y^k + x^k z^k\delta x^k$$

$$+ x^{2k}\delta z^k + y^k\delta x^k + \Delta t\rho x^\delta x_{+} \; x z^k\delta x^k \; y\sigma\Delta t(\; y^k\delta y^0$$

$$+ \; t(y^k\delta x^k - y^k\delta y^k + x y^k z^k\delta x^k + x y^k\delta z^k) + y^k\delta y^k$$

$$t(\rho x^k\delta y^k - x^k z^k\delta y^k \quad x^k\delta y^k - \Delta t(\rho x^k\delta x^k + x^k\delta$$

Firstly we find the evolved perturbation in the nonlinear model. This is defined by taking the difference of two runs of the nonlinear model. Let x_0 be the model state of the assimilation at an initial time t_0 and $\gamma\delta x_0$ a small perturbation to this state, where γ is a scalar parameter. The nonlinear model is given as M, so the model state at final time t_n is subject to the nonlinear dynamical system

$$x_n = M(t_n, x_n) \tag{3.11}$$

At time t_n the perturbation evolves as

$$\delta x_{NL} = M(t_n, x_n + \gamma\delta x_0) - M(t_n, x_n) \tag{3.12}$$

$_R$ tends to zero for

to Taylor's formula the evolved perturbation for the nonlinear

uld behave similarly for small γ, limited by machine precision.

For the TLM we write the evolved perturbation as $\delta x_L = \mathbf{M}(t_n, x_n)\gamma\delta x_0$, where $\mathbf{M}$ is the TLM. The purpose of γ is to choose the size of the nonlinearity of the Lorenz model.

From this we can calculate the relative error [17]. Firstly the error is given as

$$E = \delta x_{NL} - \delta x_L \tag{3.13}$$

Therefore the relative error defined at the final time t_n is, given as a percentage

$$E_R = 100 \quad \delta x_{NL} - \delta x_L$$

If M is exactly equal to the first order part of the nonlinear model M (of the discrete

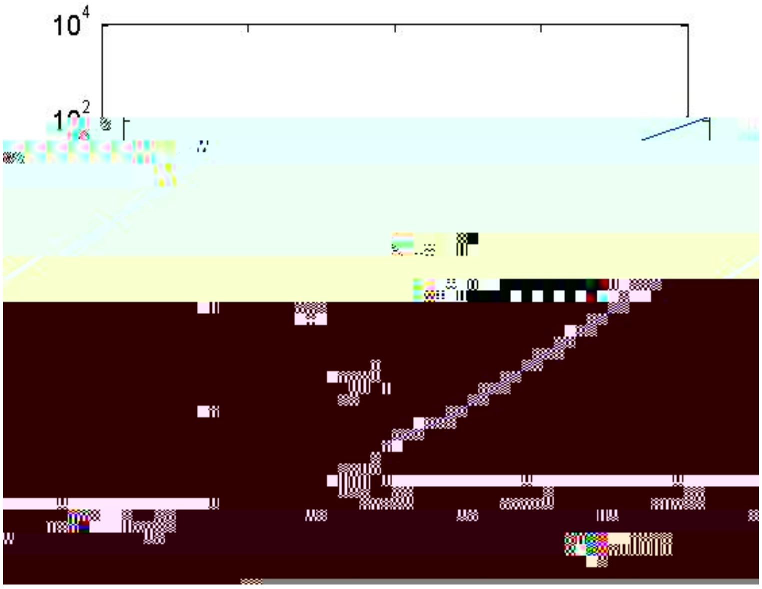


Figure 3.1: E_R plotted against γ to exam the validity of the TLM.

Chapter 4

Implementing the tangent linear model

4.1 Tangent linear test experiment

The tangent linear test measures the size of the relative error between the evolved perturbations of the nonlinear model and TLM. To enable us to investigate the effect of increasing the level of accuracy for solving the cost function minimization as the nonlinearity of the model increases, we can exploit the fact that the relative error is a measure of the nonlinearity. From this we can explore the relationship between the size of the perturbation with the nonlinearity. We expect the larger the perturbation the more nonlinear the system will be. Then for the assimilation tests perturbations as the accuracy of the inner loops are increased.

The tests are seen as a basis for the assimilation tests. The tangent linear test can be used to decide upon the values of parameters such as the range of perturbations by examining the stength of the tangent linear hypothesis. Also an appropriate time length of the assimilation has to be chosen. This shall be done using the tangent linear test by monitoring the di�erence between the nonlinear and linear model to investigate how well the tangent linear hypothesis holds over time.

The model used to test the TLM comes from the DARC website, written by Amos Lawless, 2004, [12]. Before the algorithm begins, the size of the perturbation needs to be considered by varying the value of γ. For small perturbations γ needs to be small, and vice versa for large perturbations. The unscaled perturbation for all the experiments is chosen as $(1, -1, 0.5)$, we vary this by changing γ which is a scalar multiple of the perturbation. These values will remain the same throughout all the numerical experiments. The test is as follows

1. For the initial time step $j = 0$; find the perturbation for the nonlinear model first, input x_j and $x_j + \gamma\delta x_0$ into the Runge-Kutta discretization (3.4), (3.5) and (3.6).

   Then to find the perturbation for TLM, take $\gamma\delta x_j$ and input this into the Runge-Kutta discretization (3.8), (3.9) and (3.10).

2. The first step is then repeated for n time steps.

3. The difference between $\mathbf{x}$ and $\mathbf{x} + \gamma\delta\mathbf{x}_0$ at the final time step are taken to give the evolved perturbation for the nonlinear model, giving $\delta\mathbf{x}_{NL}$. This along with the evolved perturbation from TLM at the final time step, $\delta\mathbf{x}_L$, are inputted into equation (3.14) to give the relative error.

The tangent linear test can be used to discover how good an estimate the linearization is. The smaller the relative error the better it is. Before the tests could begin we needed to consider the length of the time step. Figure 4.1 shows plots of the relative error against time for a relatively small perturbation to the nonlinear model. Each graph has a different length of time step. For the longer time step 0.05 and 0.04 the relative error was extremely large, the size of the relative error went off the scale and peaked at 100%. Generally we can conclude from the results that the larger the time length the larger the relative error. This is not ideal as we are using the relative error as a measure of the nonlinearity. To avoid a large relative error being present for all the experiments regardless of the size of the perturbation to the nonlinear model we chose the length of time step to be 0.01. A further point also worth considering is that generally as the length of time increases the relative error increases.

This is emphasized in figure 4.2 where we look at the behaviour of E_R as time increases. The relative error remains low when not many time steps have been performed. As the time reaches 1000 steps the relative error has peaked to about

150% which would not be ideal for investigating the problem. Even for 700 timesteps there is about a 50% relative error. However, we would like the length of time for the experiments to be as long as possible to get a better understanding of what we are studying. Taking this into consideration the maximum number of time steps for the future experiments were limited to 500.

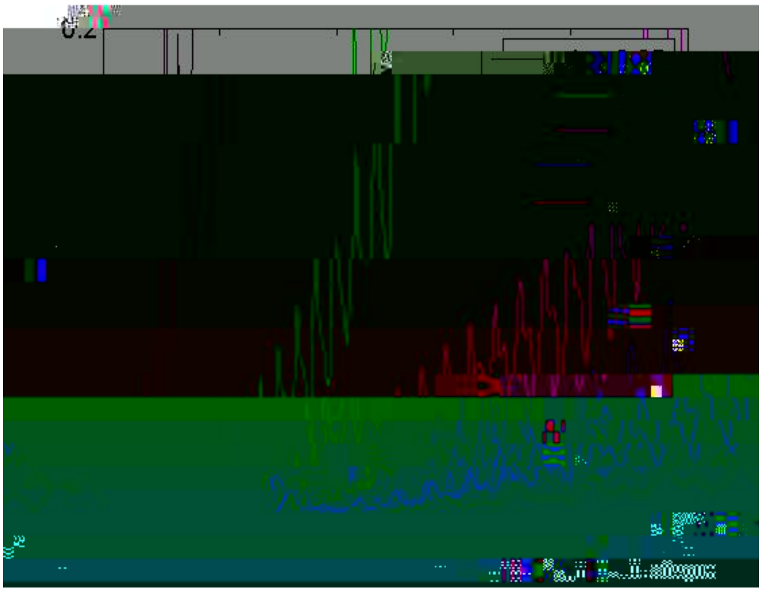


Figure 4.1: E_R plotted against timesteps of di erent length, h, with the size of the initial perturbation given as γ = 0.001 over 500 hundred time steps.

4.2 Results

A key issue with using the tangent linear hypothesis is that it neglects second order terms and higher. If the nonlinear model is weakly nonlinear then the TLM is a

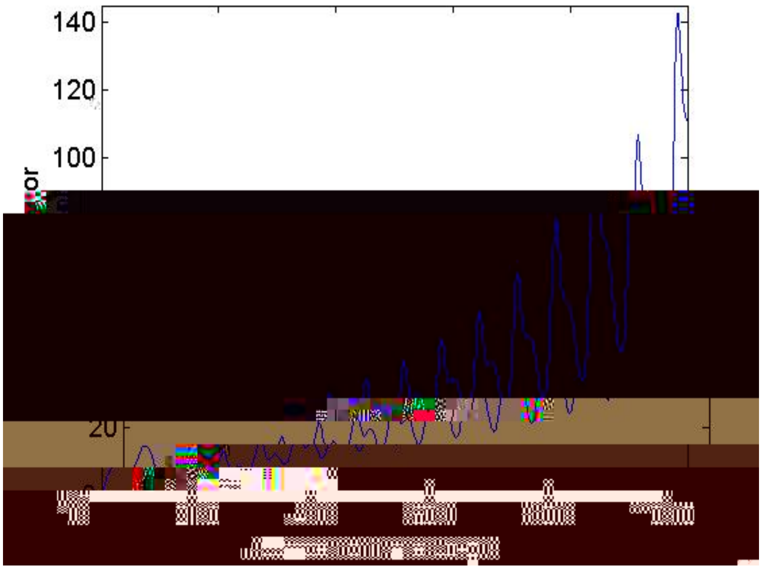


Figure 4.2: E_R

This would suggest the perturbation is evolving linearly for small values. As the perturbation becomes larger, to values of y taken to be 10 or more the relative error peaks at values over 100%. Clearly with such high values for the relative error, the linear evolution of the perturbation disagrees strongly with the evolved perturbation for the nonlinear model. Therefore the tangent linear hypothesis breaks down. The reason for this is reflected in figure 4.4 where the perturbation in x (perturbation y and z are omitted but similar) for the nonlinear and linear model are plotted against time. The two plotted on the same graph are almost identical for very small perturbations. As the perturbation increases, the phase and amplitude error gradually grow explaining the fluctations present in 4.3. For large perturbations the plots for the nonlinear and linear differ largely. The nonlinear is very flat due to the strong nonlinearities coming into play. As mentioned previously the large perturbation term $O(\delta x^2)$ will become too big to be neglected, and so as a consequence the tangent linear hypothesis will fail. The hypothesis obviously fails for y equal to 100 and 10 as the relative error grows to 100%. In figure 4.3 there is a large error at the beginning, this rapid error growth displayed in figure 4.3a) for y = 100 may be due to genuine nonlinear processes as proposed by Trémolet [18].

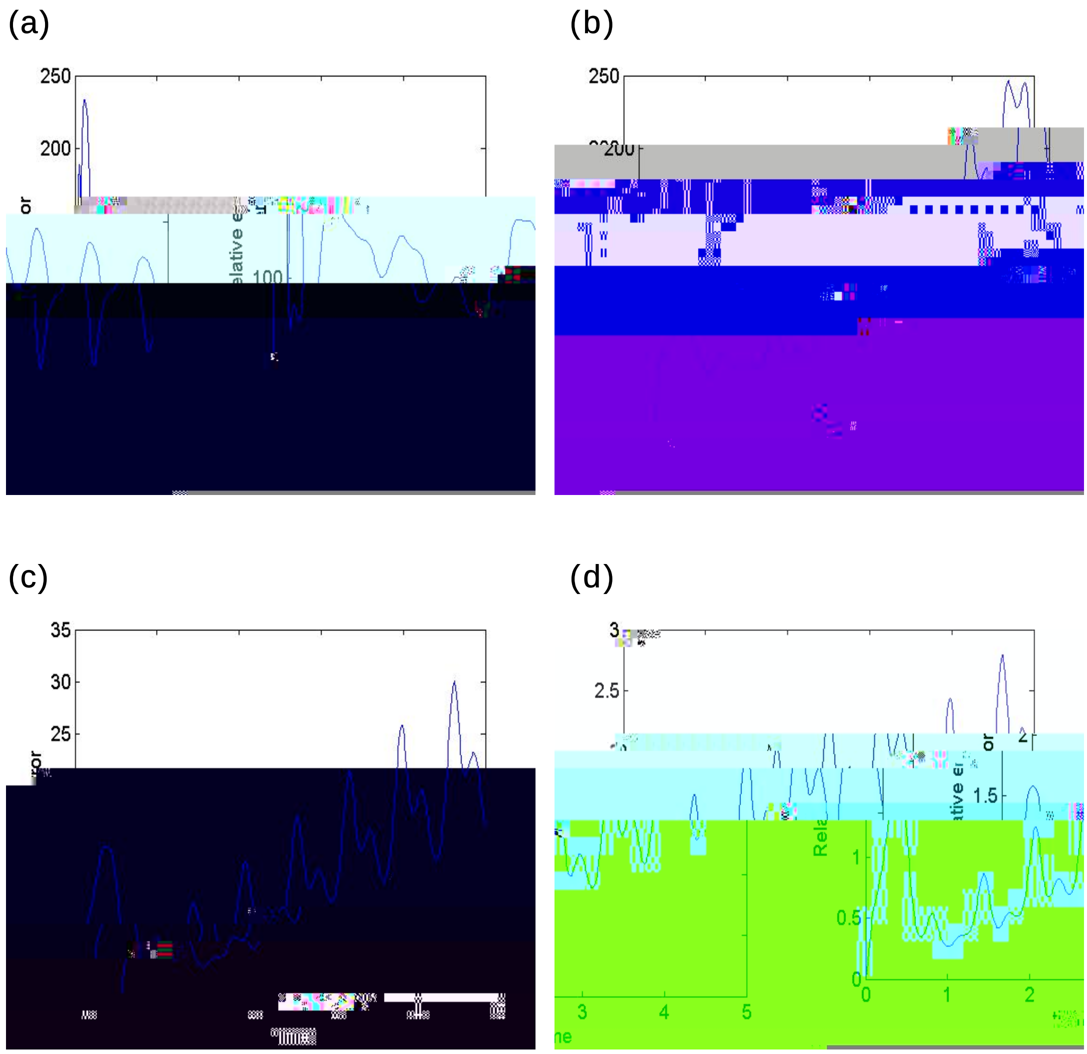


Figure 4.3: The relative error calculated using equation (3.14) against time. The time $t_n = 5$ is after 500 timesteps. The relative error is plotted over a range of different size of perturbations to the nonlinear and linear model. Graph 1(a) has the perturbation of size $\gamma = 100$, 1(b) has 10, 1(c) has 1 and 1(d) has 0.1.

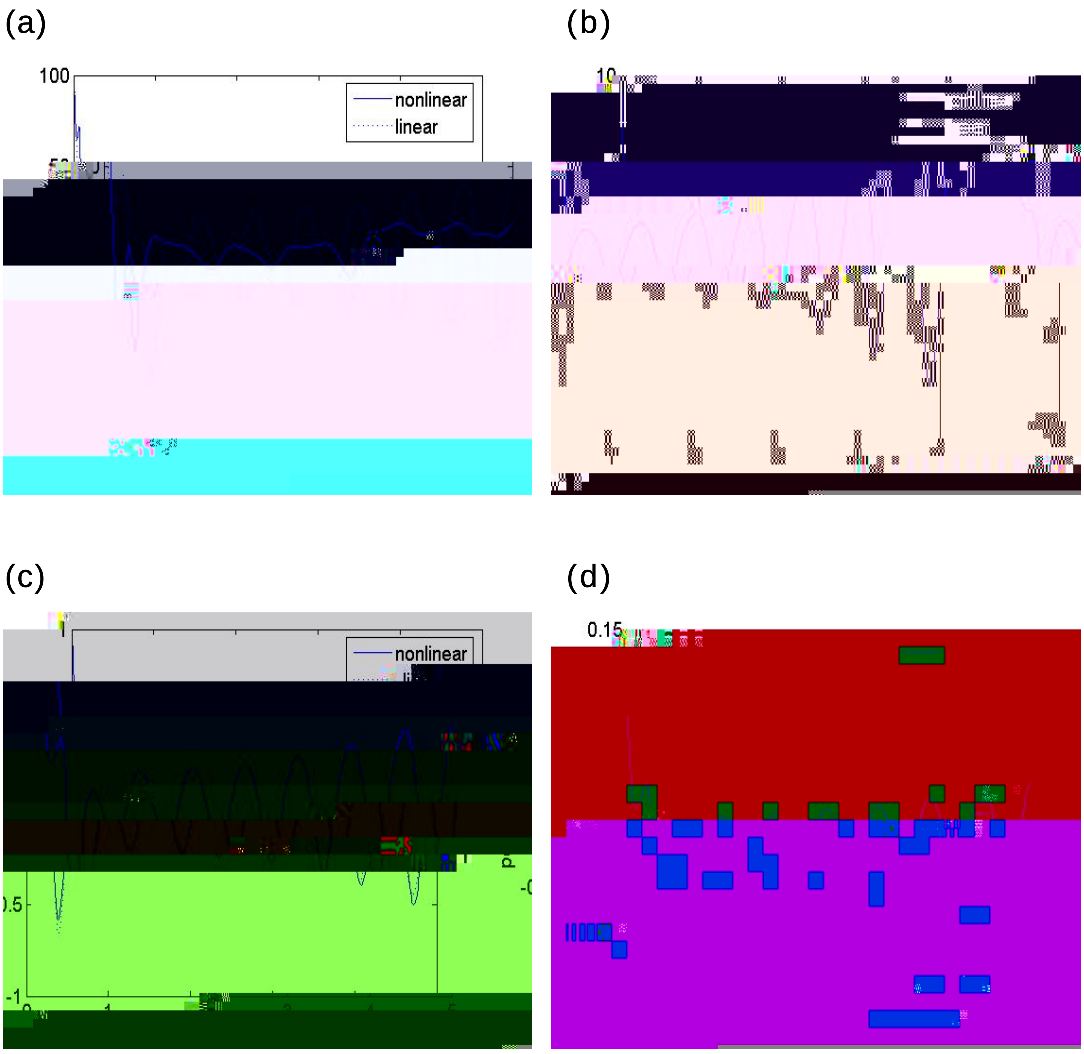


Figure 4.4: The perturbation in x plotted against time. The first graph 2(a) represents the evolution of the perturbation of size 100 for the nonlinear and tangent linear model. Figure 2(b) has size of perturbation 10, 2(c) has 1 and 2(d) has 0.1. Note that the y-axis reduces by about a factor of 10 as the size of the perturbation does.

Chapter 5

Assimilation experiments

5.1 Implementing the assimilation

For the numerical experiments using the Lorenz model, the time window is chosen to be 500 timesteps of length 0.01 to link our experiments to the tangent linear tests. The first 200 timesteps will be the assimilation, then the following 300 timesteps will be the forecast. Identical twin experiments were carried out for the nonlinear model allowing the true solution to be found, enabling us to compare the truth with the analysis of the solution retrieved from the assimilation with a perturbation added to the nonlinear model.

The incremental method is given by the algorithm in section 2.2, where the nonlinear model used is the Lorenz model. Steps 1-5 from section 2.2 are followed, although step 3 is not straightforward and must be considered in further detail.

gradient term by term, working backwards from the final time t_n to the initial time. Notice that this has to be done backwards, as M_{j-1}^T needs to be calculated before M_{j-2}^T, M_{j-3}^T etc.

After the adjoint model, the steepest descent as described in section 2.3 is implemented in step 3 to determine the path of the minimization. This is referred to as the 'inner inner' minimization which was described previously in detail in section 2.3.

5.2 Choosing the number of inner and outer iterations

When implementing the incremental method, a few factors have to be considered.

converge very quickly, if at all. A consequence may be that the system diverges from the true solution so that the final analysis may be further from the truth than the original background. On the other hand, if too many iterations are performed for the inner loop problem, then it is being solved to an unnecessary accuracy. Then after each update, the system does not benefit much from this accuracy, and so uses extra computational work at no advantage. Taking this into consideration, we experimented with the assimilation test to decide on the maximum number of inner iterations.

The model was ran with 10, 20, 30 and 40 inner iterations followed by an update to the trajectory. The cost function and its gradient at the time of the last inner iteration accompanied with the relative change in the gradient (2.33) are displayed in table 5.1. Clearly an advantage of increasing the number of inner iterations being performed is that the nonlinear problem is solved more accurately, the lowest value of the cost function occuring for 40 inner loops. However by the end of the

5.2a). For 30 and 40 inner iterations there was an evident jump in the convergence of the gradient. The jump in the gradient indicates the outer loop which redefines the inner loop cost function. Although more inner iterations means the minimization of the cost function is being solved more accurately, this is not a good approximation to the nonlinear problem resulting in the jump in convergence of the gradient when the outer loop is performed.

A compromise has to be made between the computational cost and the degree of accuracy that the cost function has to be solved. For the assimilation experiments the maximum number of inner iterations between each update will be 20. This allows the series of minimizations to be solved accurately enough to give convergence (depending on the value of the tolerance), without solving it too precisely, wasting computational effort.

Another important aspect to decide is the termination of the outer loops. If the outer loops are stopped prematurely then the cost function may still be very high and result in an inaccurate forecast. However if there are too many outer loops then this greatly increases the computational cost. In practice, the number of outer loops is normally no more than 2 or 3. This is a(o)1(r8mgr1(n)-fr8mgr1,)-279gr1comiopasl l co3F2 o of great concern. The tolerance for the outer loop is kept constant at $_{outer} = 0.01$. We don't want this to change so that we avoid getting confused with the effects of when the tolerance for the inner loop stopping criterion changes.

no. of inner iterations	cost function	gradient	relative change in gradient
10	5.7035	27.0436	0.6705
20	2.6879	8.1092	0.2480
30	2.4451	5.6929	0.1624
40	2.1901	3.5748	0.0962

Table 5.1: The number of inner iterations performed corresponding to the value of the cost function and the gradient of the cost function at the time of the last inner iteration. The table includes the relative change in gradient from equation (2.33). This compares the value of the gradient of the cost function at start of the inner iteration on the outer loop and the last inner iteration.

5.3 Details of the assimilation experiments

We begin with observations taken at every time step over the assimilation period allowing a good solution to the nonlinear problem to be found. The observations are taken of all 3 spectral coordinates of the Lorenz equations x, y and z. The tolerance of the stopping criterion for the inner minimization reflects how well the inner problem is solved. The larger the tolerance the less accurately the inner minimization is solved. When beginning the assimilation experiments we firstly want to investigate how well the nonlinear problem is solved for different strengths of nonlinearities of the model by changing the size of the perturbation to the model. For this we want to keep the tolerance constant so not to interfere with the results. Following this, we want to vary the level at which the inner minimization is solved by altering the value of the tolerance. This will enable us to investigate how well the problem can be solved as it becomes more nonlinear.

Recall that the size of the perturbation γ to the nonlinear model is related to the nonlinearity of the problem. When considering the range of γ we want to include when the tangent linear hypothesis is a good approximation to the nonlinear model (small perturbation) and when it breaks down. For comparison we include in the assimilation experiments γ in the range from 0.1 to 100. Referring to section 4.2, the tangent linear hypothesis seems to no longer approximate the problem well as γ reaches 10 and 100 as the relative error reaches 100% in figure 4.3a) and b). For $\gamma = 0.1$ the relative error is very small so that we can infer that the tangent linear hypothesis holds.

To link the results from the tangent linear test, the initial conditions for the tangent linear and Lorenz model were the same at $(1, 2, 1.5)$. The perturbation for all the experiments is chosen as $(1, -1, 0.5)$, we vary this by changing γ which is a scalar multiple of the perturbation. The first guess for the assimilation is calculated by adding $\gamma(1, -1, 0.5)$ to the true initial condition $(1, 2, 1.5)$.

5.4 Results

The solution of the Lorenz model is shown in Figure 5.2 for x and z (y is omitted as it is symmetrical to x) along with the error in Figure 5.3 between the true solution and the analysis produced from running the nonlinear with perturbation.

The first 200 timesteps represents the assimilation. The succeeding 300 timesteps gives the forecast produced from the analysis and the true solution. The first guess trajectory comes from the background state which is then run in the model, the analysis combines the observations and the model to give the trajectory for the incremental method. Recalling from section 3.2 where γ is given as the size of the perturbation to the nonlinear model, the figures 5.2 and 5.3 display plots over a range of perturbations measured using γ

figure 5.2c) seems to very close to the true solution for both x and z. Studying closely figure 5.3c) the analysis trajectory in the solution for x seems to be in phase with the true trajectory, however, there is a small amplitude error which is demonstrated in the error plot of x. In the solution for z the analysis trajectory displays a difference in the amplitude sizes as well as a phase error to the true trajectory near the end of the forecast. As the perturbation reduces to $\gamma = 1$ the analysis becomes very close to the truth. We would expect this as the nonlinearities of the Lorenz model become weaker so that the tangent linear hypothesis gains validity. This is evident from 5.3, the error between the analysis and truth is very large for large perturbation. As expected the error reduces as the size of the perturbation does, there is a significant drop in the error from the size of $\gamma = 50$ and $\gamma = 10$. When the value of γ has fallen to 0.1 the error is minor, this is reflected diagramtically in the final plot in 5.2d) case then this implies that the cost function can get only so small for each size of perturbation. So the nonlinear model with a large perturbation can not minimise the cost function to the same value compared to that of the nonlinear model with a smaller perturbation no matter how accurate the assimilation is made. However, this does not seem to be the case. Studying the values of the cost function, for example when $\gamma = 0.01$ the final value of the cost function when the tolerance is 0.05 is 0.0104 and when the tolerance is reduced to 0.01 the cost function has increased to the saturation point 0.0192. The cost function may be have reached a local minimum when the tolerance is 0.05. Therefore the saturation point is not the global minimum value of the cost function. This is an interesting result and not good news if we need to produce a forecast for a highly nonlinear model.

We must consider that this result may be a limitation of the minimization algorithm used which we commented on before as being extremely inefficient. Reaching the minimum may be incredibly slow and thus may need further inner iterations to be performed to reach the minimum. Using the steepest descent, the inner minimisation may have got 'stuck' in a local minimum and so not representing the true global minimum of the cost function. Due to the nonlinearity of the problem, the whole cost function is non-quadratic. Therefore this may result in the value of the cost function being larger (locally) than in the previous iterations when entering a 'new valley' of the cost function. This may explain why the cost function at saturation point might be larger than the cost function before this point. To investigate whether this is the case, other minimisation techniques could be used such as a conjugate or quasi-Newton methods, as mentioned in section 2.4. However, whether the inner minimization has reached a local or global minimum, the gradient of the cost function should reach zero regardless. From table 5.2 notice that the value of the gradient never reaches zero, therefore the value of the cost function is not the minimum. This is because the gradient becomes saturated after the inner minimization is solved to a level of accuracy.

The criterion used in the assimilation experiments allow the inner iteration count to stop once the relative change in gradient (2.33) has fallen below the tolerance. As the tolerance increases the inner minization is solved less accurately before the next outer loop is performed. So if the tolerance is too large then not enough inner iterations are carried out between the outer loops to allow the cost function to significantly reduce. This is demonstrated in table 5.2, when the tolerance = 0.5, the value of the cost function is larger than when is reduced. The outer loops were noted to take into account whether the outer loop stopping criterion was being satisfied. The more outer loops performed the higher level of accuracy the problem is being solved. It seems that for larger perturbation the nonlinear model the more outer loops are performed, and so demanding a greater accuracy of the outer loop to approximate the solution.

Studying further table (5.2) from a different perspective by keeping the tolerance constant we can gain an insight of how the value of the cost function changes for different values of γ. The larger the size of γ the larger the cost function and the gradient are. As a consequence more outer loops are performed for larger γ. By analyzing the table the conclusion being drawn is that the more nonlinear the model (i.e. the larger the perturbation), the higher the level of accuracy needed to solve the inner minimization problem. We can explain this by referring back to the theory in section 2.4. If the size of the perturbation gets larger then so does the nonlinearity of the Lorenz model. Therefore from (2.20), the term $Q(\mathrm{x})$ representing the second order derivatives, gets larger. As $Q(\mathrm{x})$ is included in the inequality (2.27) then the bound for β_k decreases. From (2.28), since β_k is the bound for the inner loop minimization then this is equivalent to the inner loop tolerance. If β_k decreases then the inner minimization will not be solved as accurately.

Perturbation with $\gamma = 10$		
Tolerance	Cost function	Gradient

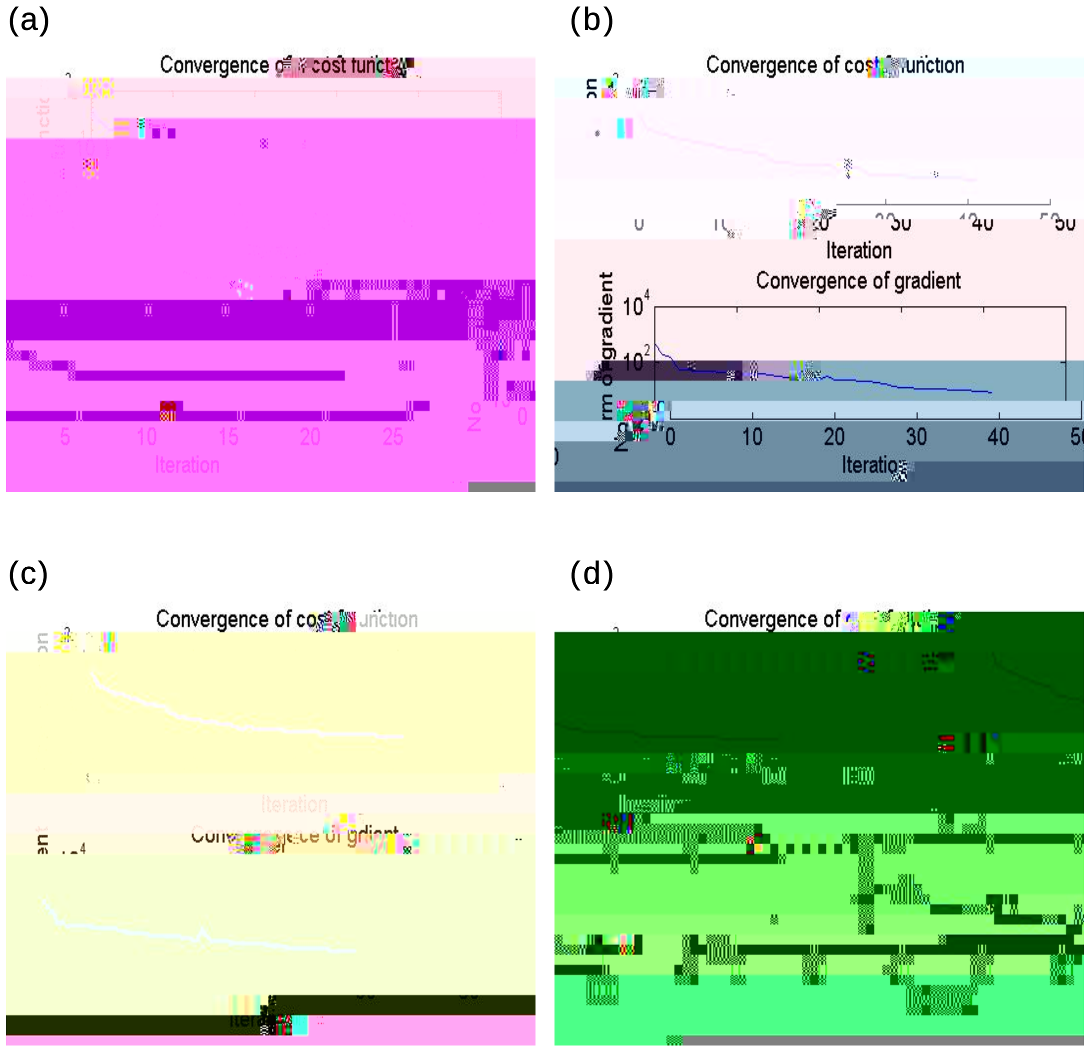


Figure 5.1: The cost function and the gradient plotted against the number of total number of inner iterations performed, 3(a) 10, 3(b) 20, 3(c) 30 and 3(d) 40. The tolerance for the inner loop was given as 0.00001, with observations at every time step. The perturbation to the nonlinear model was size $\gamma = 1$.

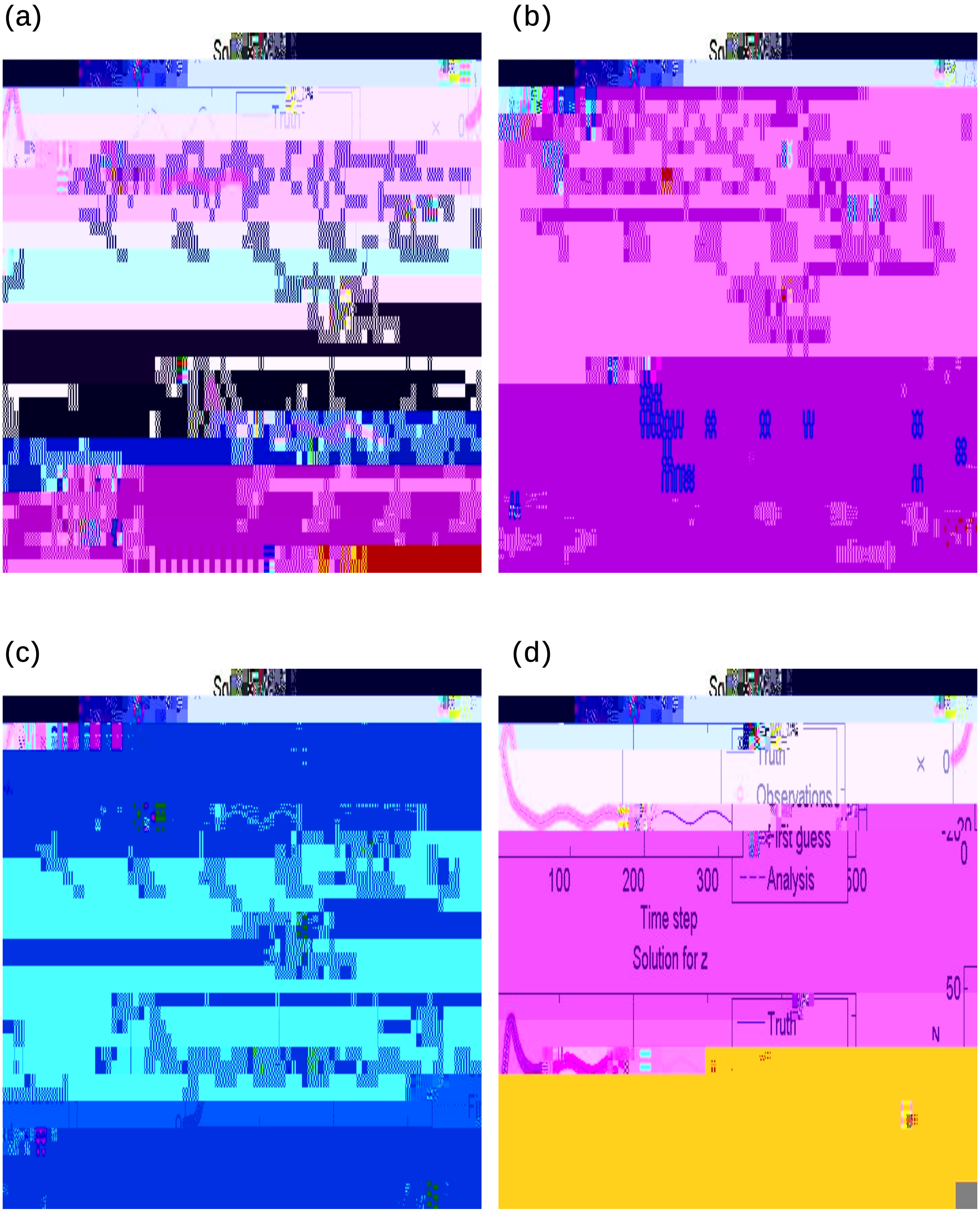


Figure 5.2: The solution for x and z of the Lorenz equations plotted against time. 4(a) shows the Lorenz model with the size of perturbation $\gamma = 100$, 4(b) $\gamma = 50$, 4(c) $\gamma = 10$ and 4(d) $\gamma = 1$. The tolerance for the inner loop stopping criterion was 10^{-5}.

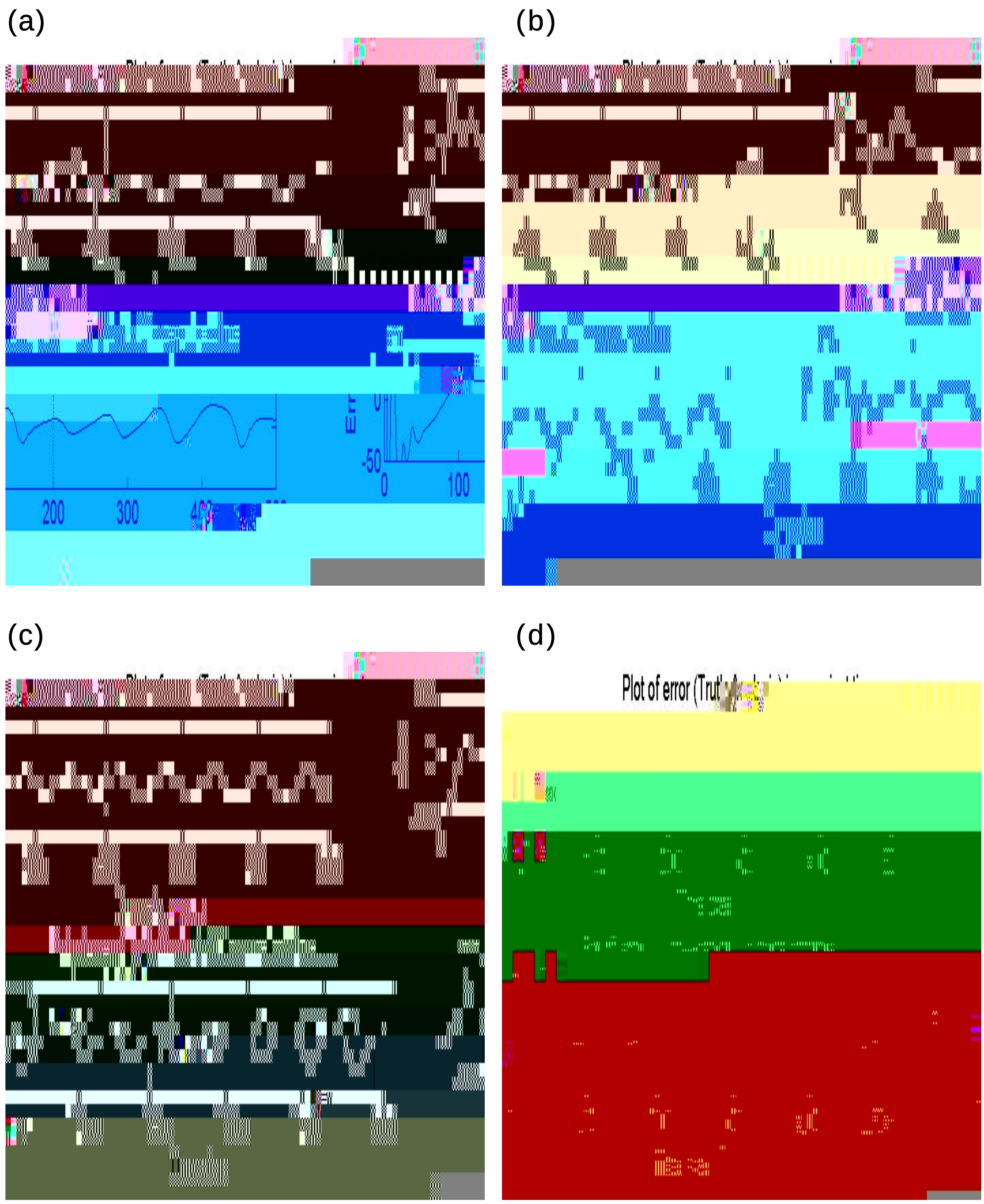


Figure 5.3: The error between the true and analysis trajectories plotted against time. Figure 5(a) γ = 100, 5(b) γ = 50, 5(c) γ = 10, and 5(d) γ = 1. The tolerance for the inner loop stooping criterion was 10^{-4}.

Chapter 6

Imperfect observations

6.1 The observational error

In practice error on all observations occur. The previous experiments were performed using perfect observations. To see whether we gather the same findings for imperfect obsevations a random unbiased error with a Gaussian distribution is added to the observations y_j with variance 0.01. Referring back to equation (2.3) the observational error added is denoted as $_j$.

For perfect observations it is possible to find a state vector x_0 such that for all j

$$\mathrm{x}_j = M(\mathrm{x}_0) = \mathrm{y}_j = \mathrm{h}_j[\mathrm{x}_j] \tag{6.1}$$

This implies that the minimization of the cost function, J, and the cost function itself are equal to zero.

However, this is not the case for imperfect observations. Recall from section 2.1 the observation at time t_j is given as

$$y_j = h_j[x_j] + \ _j \tag{6.2}$$

where $\ _j$ is the noise on the observation. If

$$x_j = M(x_0) = y_j = h_j[x_j] \tag{6.3}$$

then the cost function can be minimized so that its gradient will equal zero. However, there is no such x_j that will equal zero for all j. Comparing this to perfect observations, we must account for the fact that the minimization of the cost function may need to be solved more accurately. We expect similar results to that of the previous chapter when considering no observational error, however, the level of tolerance may be smaller.

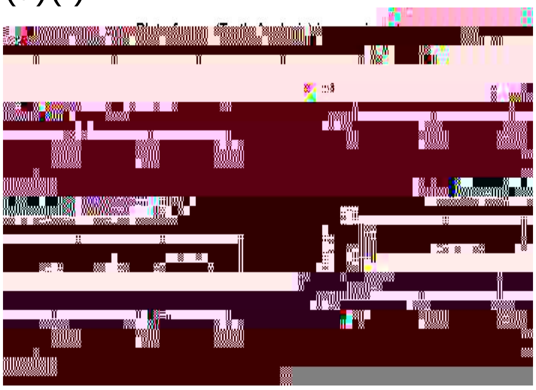
Indeed referring to [7], experiments for perfect and imperfect observations were performed using different value of tolerances for the inner loop minimization. A comparison between the two indicated that a smaller tolerance may be required when more noise is added to the observations.

6.2 The impact of observational error on the assimilation experiments

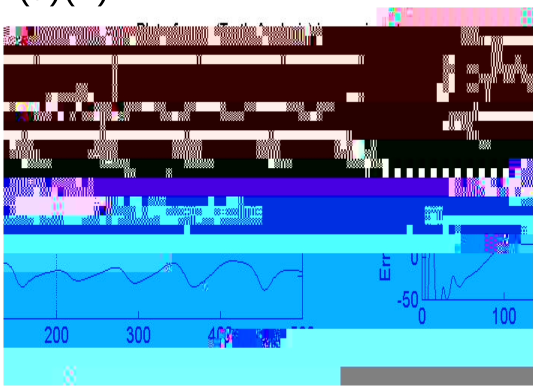
For the assimilation experiments, consistent with the experiments carried out for the perfect observations the parameters remained unchanged along with the number of inner and outer loops and their respective stopping criteria. The random error added to the observation was kept constant throughout the experiments. Figure 6.1 shows the errors between the true and analysis trajectories of the solutions to the Lorenz model. Clearly for imperfect observations, the error through the time window exhibited a more distinctive wave motion, reflecting a larger phase error than for the perfect observations. This became clearer when the scale for the vertical axis got smaller (i.e. for smaller size of perturbations). The difference between the two trajectories was larger when imperfect observations are used. Therefore we can conclude that the impact of observational error has a detrimental effect on the value of the cost function. However, similar to the perfect observations, as the size of the perturbation reduces the difference between the true and analysis trajectories

Continuing the comparison of the perfect and imperfect observations, the level of accuracy at which the inner minimization was solved was investigated. The results were presented in table 6.1 which compares the convergence of the cost function for varying tolerances and perturbations to the nonlinear model. The convergence of the value of the cost function for imperfect observations followed that of the perfect. Looking in detail however, the inner loop minimization has to be solved more accurately when noise is added to the observations to give as good a solution

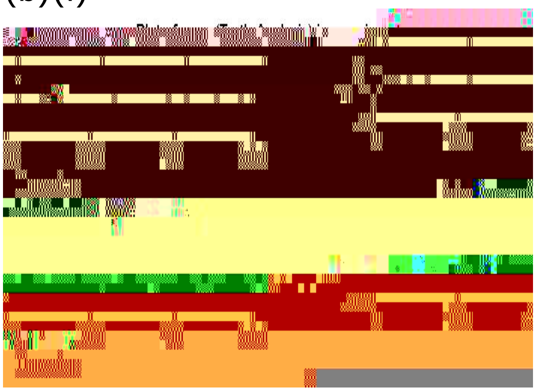
(a)(i)



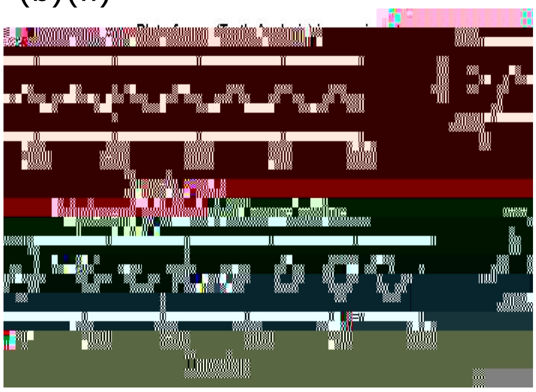
(a)(ii)



(b)(i)



(b)(ii)



(c)(i)

(c)(ii)

the cost function reaches saturation point. Generally this result agrees with our

Perturbation with γ = 10				
Tolerance	J with noise	No. outer loops with noise	J without noise	No. of outer loops without noise
0.5	206.2796	10	210.2294	10
0.1	197.5135	10	203.9113	10
0.05	178.9106	10	177.1834	10
0.01	183.9588	10	188.9216	10
0.001	183.9588	10	188.9216	10
Perturbation with γ = 1				
0.5	1.9396	10	4.9379	10
0.1	1.8283	10	4.7844	10
0.05	1.7668	10	4.7284	10
0.01	1.744	10	4.7476	10
0.001	1.744	10	4.7476	10
Perturbation with γ = 0.1				
0.5	0.0221	10	3.0072	8
0.1	0.0197	6	3.0043	6
0.05	0.0104	6	3.0062	7
0.01	0.0192	6	3.0062	7
0.001	0.0192	6	3.0062	7
Perturbation with γ = 0.01				
0.5	0.00027	6	2.0854	6
0.1	0.0002308	5	2.9853	5
0.05	0.0004477	5	2.9857	4
0.01	0.0002217	5	2.9853	5
0.001	0.0002217	5	2.9853	5

Table 6.1:

Chapter 7

Discussion

7.1 Summary and Conclusion

This dissertation has been examining the behaviour of incremental 4D-Var method for a nonlinear model. Increasing the strength of the nonlinearities of the model, the level of accuracy at which the problem must be solved was investigated.

The incremental approach was introduced by Courtier et al. (1994) [2] to reduce computational costs, making it possible to envision its operational implementation. The method replaces a direct minimization of the full 4D-Var cost function with

Firstly in this dissertation a measure of accuracy was introduced for the nonlinearity of the model. This was done using the tangent linear test which gave the comparison between the evolved perturbations of the nonlinear and linear models. The output was called the relative error. When evaluating the relative error the size of the perturbation to the nonlinear model was varied. It was found the larger the perturbation the larger the relative error. This results allowed us to use the size of perturbation as a measure of the nonlinearity of the model. This information was then implemented in the assimilation experiements using the Lorenz model. Initially the assimilations tests were carried out for perfect observations. A stopping criterion was set for the number of inner and outer loops performed. When running the assimilation tests we observed that the trajectories for the truth and the analysis agreed for small perturbations to the nonlinear model. However, as the size of the perturbation increased the error between the two trajectories also increased. This led us to consider increasing the level of accuracy that the inner minimization was solved for the incremental approach to see whether the analysis trajectory would become closer to the truth for a high nonlinear system. To a certain degree increasing accuracy did improve the analysis trajectory. However, the potential to exploit this was limited. When the level of accuracy of the inner loop minimization increased to a certain value the impact of this property became redundant.

The influence of introducing error to the observations was examined. Generally the incremental method using imperfect observations had the same qualitive be-

The observations that were used for the assimilation experiments were functions of all three spectral components of the Lorenz model measured at every time step. In reality this may not be so common. Often it is the case that the observations measured will only include one or two of the components of the model and will not be uniformly distributed in time as frequently. To allow our conclusions to be viewed in a more operational context further tests should be done with fewer observations which do not depend on terms of the three components.

For the assimilation increasing the accuracy of the incremental method was only considered for the inner loop minimization. This was prioritzed due to the vital

This implies that the nonlinearities of the model started to play an ever increasing role meaning the approximation of the TLM to the nonlinear model got worse. From our findings that the larger the nonlinearity of the Lorenz model the worse the solution to the problem was. We would expect that it may not be possible to provide a valid solution for a long time window. Indeed a more in depth look was considered by Trémolet [18] studied the aspect of extending the time window of the assimilation. The incremental approximation exhibited a potential limitation after 12 hour as the relative error came close to 100 %. The stronger the nonlinearity of the model the larger the relative error was. There have been several studies which consider the linear assumption for small perturbations such as Gilmour et al. (2001) [5]. The general conclusion seems to be that the linear approximation is valid for two to three days [16]. Highlighting the fact that the Lorenz system can produce a decent forecast for a limited time only before diverging [4], one property of the Lorenz model in particular is its sensitivity dependence on the initial conditions. This may indicate that if the time window was extended, the forecast produced from the assimilation would be inaccurate. Considering larger perturbations, from our studies we have concludedssheosdowsao

perturbati], to the nonli(lar)0414(mo)-27(del)0592(w)27(e)0592(w)27(ould)0596(exp)-27(tld)0599(b)2 minimization for the incremental method only has effect for a short time, after which there is no impact, or would benefit the whole assimilation.

The linear approximation used in this study was the tangent linear model (TLM). However, the Met Office has designed the perturbation forecast model (PFM). The TLM works by linearzing the discrete form of the nonlinear model. Alternatively the PFM linearizes the continuous equations of the nonlinear model first, then discretizes these equations. This approach as referred to as semi-continous which avoids problems that occur when linearizing complex schemes. The validity of the TLM is restricted to inifintesimal sized perturabtions to the nonlinear model. An additional advantage of the PFM is its ability to model finite perturbations of the size of uncertanties in the initial conditions accurately. It was found that the PFM can be as accurate as the TLM for finite perturbations [11]. It would be interesting to look further into how well the incremental method using the PFM approximates the solution of the nonlinear system for the size of y at which the tangent linear hypothesis breaks down, as the accuracy of the inner loop minimization was increased.

Bibliography

[1] R. Bannister, *Theory and Techniquies of Data Assimilation*, http://www.met.rdg.ac.uk/ ross/DataAssim.html, 25/07/06.

[2] P. Courtier, J. N. Thépaut, A. Hollingsworth, *A strategy for operational implementation of 4D-Var, using an incremental approach*, Q.J.R.Meterol. Soc., 120 (1994), pp. 1367–1387.

[3] ECMWF, *Training Programme 2006*, http://www.ecmwf.int/, 25/07/06.

[4] P. Gauthier, *Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model*, Tellus, 44A (1992), pp. 2-17.

[5] I. Gilmour, L. A. Smith and R. Buizza, *Linear Regime Duration: Is 24 hours a long time in synoptic weather forecasting?*, J. Atmos. Sci., 58 (2001), pp. 3525–3539.

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

[7] A. S. Lawless and N. K. Nichols, *Inner loop stopping criteria for incremental four-dimensional variational data assimilation*, Monthly Weather Review, (2005), pp. 3.

[8] A. S. Lawless, S. Gratton and N. K. Nichols, *An investigation of incremental 4D-Var using non-tangent linear models*, Quart. J. Royal Met. Soc., 131 (2005), pp. 459–476.

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

[10] A. S. Lawless, *Data assimiliation with the Lorenz equations*, http://darc.nerc.ac.uk/, 12/08/2006, simple models.

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

[12]

[18] Y. Trémolet, *Diagnostics of linear and incremental approximations in 4D-Var*, Q.J.R.Meteorol. Soc., 130, pp. 2233–2251.