Toward shadowing in operational weather prediction

Kevin Judd  Carolyn Reynolds  Thomas Rosmond

January 27, 2004

Abstract

This is a report on a NICOP funded project to investigate the feasibility and benefits of computing shadowing trajectories with the Naval Operational Global Atmospheric Prediction System (NOGAPS). Previous results had established that an algorithm to compute shadowing trajectories was successful for a simple quasi-geostrophic (T21L3) model, without using any adjoint information. The first aim was to determine whether algorithms could be successfully applied to the significantly more complex NOGAPS. Experiments at various resolutions from T21L9 to T79L30, using a dry adjoint approximation, showed satisfactory convergence of an iterative, gradient descent based, algorithm. These experiments have provided useful information about the implementation and calibration of shadowing algorithms and an illustration of how shadowing methods could help identify model error and obtain better analyses and forecasts. The results encourage further investigation of shadowing techniques.

1 Introduction

This report begins with a brief introduction to the essential ideas of shadowing, followed by a detailed account of the algorithm used and results obtained. The account is intended to be detailed enough that methods and results can be independently reproduced from this report alone. However, this level of detail may be overwhelming on a first reading. To assist the reader the report adopts the following style when appropriate: each section elaborates on the statements of its first paragraph, and each paragraph elaborates on the statements of its first sentence. A glossary of terms used in this report is provided in Appendix A.6.

1.1 Shadowing: trajectories and pseudo-orbits

For the purposes of Numerical Weather Prediction (NWP), shadowing can be given the rather broad meaning that there is a solution of a forecast model that is consistent with atmospheric observations over some period of time, this solution is said to shadow the observations. There are obvious motivations for an interest in shadowing, for example, one cannot expect to make detailed 4 day forecasts, unless one can consistently shadow past observations for 4 day periods. How long, and how accurately, a forecast model can shadow observations is a useful measure of how good the model is. Systematic failure to shadow, or failure to shadow in particular circumstances, can highlight model errors. Shadowing also plays an important role in the theory of indistinguishable states [13, 14], which is a new approach to state estimation, and ensemble and probabilistic forecasting.

More precisely, we will define shadowing trajectories in the following sense. Suppose a forecast model represents the atmosphere by a d-dimensional state space $\mathbb{R}^d$, and defines the dynamics as a map $f$, of fixed time step, say 6 hours, that is, given a state $x \in \mathbb{R}^d$ at some time, then $f(x)$ is the forecast for 6 hours later, and $f(f(x))$ the forecast for 12 hours, and so on. A sequence of states $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, such that, $x_{i+1} = f(x_i)$ is called a trajectory. Shadowing asks, given an arbitrary sequence of states $y_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, can one find a trajectory that remains close to all these states, that is, $\|y_i - x_i\| < \epsilon$, for all $i = 1, \ldots, n$, for some bound $\epsilon$ on the error? Such
Figure 1: Schematic of shadowing. The red dots represent the states $y_i$ one wishes to shadow, which are usually observations, or analyses. These states are not a trajectory of the fixed time step forecast model $f$, the faint arrows and open dots represent the mappings $f(y_i) \neq y_{i+1}$. The green dots represent a shadowing trajectory of states $x_i$, these are connected by a green line to represent that they are a trajectory, so each $x_{i+1} = f(x_i)$. The large circles represent the bound $\epsilon$ on the distance $\|y_i - x_i\|$, which is usually related to the maximum supposed observation, or analysis, error.

a trajectory is said to $\epsilon$-shadow the other states, or simply, is a shadowing trajectory. Figure 1 is a schematic representation of shadowing in the sense just defined.

Suppose the sequence of states $y_i \in \mathbb{R}^d, i = 1, \ldots, n$, to be shadowed are analyses, that is, states of the forecast model estimated by assimilation of atmospheric observations by some means, say 3DVAR. Shadowing analyses is not really what we want to do, we want to shadow the observations directly. We use the analyses as a proxy for the observations. Provided the analyses are not too bad, this is not an unreasonable thing to do. At this early stage of development of the methodology it is considered an easier and more pragmatic choice.

We cannot expect a sequence of analyses to be a trajectory of the forecast model $f$, because observations of the atmosphere are both limited and inaccurate, and because the forecast model is an imperfect model of the atmosphere. If $f$ were a perfect model of the atmosphere, then there must be a sequence of states that exactly correspond to the atmospheric state at 6 hourly intervals, and this sequence of states will be a trajectory of the perfect model, and must shadow the analyses to within analysis error\(^1\). In practice, forecast models are not perfect, and as a consequence one cannot shadow analyses over arbitrary long periods, even with perfect observations. The better the model, however, the longer the period it will shadow. So typical shadowing times are a useful measure of the quality of the forecast model.

When a model is imperfect, as they always are, it is useful to relax the shadowing criteria, and look for shadowing pseudo-orbits. A pseudo-orbit is a sequence of states $x_i \in \mathbb{R}^d, i = 1, \ldots, n$, that is almost, but not quite, a trajectory. A suitable requirement is that the quality $\sum_{i=1}^n \|x_{i+1} - f(x_i)\|^2$, called the total squared indeterminism, be small. Note that for a trajectory the indeterminism is zero. The quality of an imperfect forecast model can also be measured by the minimum indeterminism of shadowing pseudo-orbits over a fixed period. This minimum might be difficult to determine, however, the algorithms developed here will at least provide an upper bound on the minimum.

1.2 Causal and non-causal analyses

Apart from identifying model error, shadowing can also be used to for obtaining analyses for making forecasts and verifying forecasts. There are two important types of analysis, causal and non-causal, and shadowing can be useful in obtaining both.

\(^1\)Strictly speaking this is not $\epsilon$-shadowing, because observational errors are not usually taken to be bounded, it is more usual that they are taken to have a Gaussian distribution. Allowing the extra freedom of just being consistent with observations is called $\iota$-shadowing [6, 19].
An analysis is a state of the forecast model obtained by assimilation of observations. If the model is perfect, then the analysis is an estimate of the state of the system, usually the best guess, or maximum likelihood estimate, taking into account the known uncertainties of the observations. When the model is imperfect, then the analysis is not a state estimate, because there is no “truth” state.

A causal analysis uses only information from the past and present to obtain the analysis. The idea of a causal analysis is illustrated in figure 2(a). Forecasts made from causal analyses are often called honest forecasts in the dynamical systems and time series literature. Honest forecasts are the basis of NWP. The theory of state estimation for linear systems shows that for perfect models the Kalman filter provides optimal state estimates, that is, these provide the best honest predictions. The theory of indistinguishable states shows that for non-linear systems shadowing trajectories can provide optimal causal state estimates for perfect models. To do this, one obtains the maximum likelihood shadowing trajectory of all observations from the past to the present, then uses the final state of this shadowing trajectory as the state estimate [13, 12]. For an imperfect model shadowing pseudo-orbits can be used to obtain useful analyses by taking model error into account [14].

Figure 2: Schematic of (a) a causal shadowing analysis and (b) a non-causal shadowing analysis. For a perfect model there is a true trajectory. The vertical line represents the present, the “bracing arrows” at top of (a) and bottom of (b) indicate the window of information used to obtain the analysis.

Non-causal analyses use information from the past and future to obtain a better analysis than a causal estimate can; the main purpose of these are for forecast verification. The idea is illustrated in figure 2(b). Forecasts from non-causal analyses are not honest forecasts, because future information was used to obtain the analysis. Non-causal analyses are particularly important for verifying forecasts of systems that have unstable modes and sensitive dependence to initial conditions. It can be shown for perfect models that shadowing trajectories provide state estimates along their entire length, where these are all non-causal, except for the final state where the trajectory comes to an end. The state estimates obtained from a finite length shadowing trajectories tend to have largest error when taken from the ends, and smaller errors when taken from the middle, although this is not guaranteed [18]. For a perfect model the end in the past has errors in the stable directions, or stable modes, where as the end in the present (or future) has errors in the unstable directions, or unstable modes. In the middle of the shadowing trajectory there is usually the smallest errors, because information from the past and future eliminate both stable and unstable errors. Under sufficiently strong restrictions (hyperbolicity, sufficiently small and bounded error), it can be shown that as the shadowing trajectory gets longer, the non-causal state estimates converge to the true states [18]. A similar behaviour seems to occur for imperfect models, but there is no strong theory. The complications to this story from non-hyperbolicity [2, 5, 16] are probably irrelevant to NWP, because near tangencies would be more prevalent when forecast models are perfect [18], but it is much more likely that model imperfection dominates both considerations [14].

As a general principle, for nonlinear systems, shadowing trajectories can provide good (perhaps optimal) causal analyses for forecasting with perfect models, and good (perhaps optimal) non-causal analyses for forecast verification. Furthermore, one expects that the causal analysis, taken from the end of a shadowing trajectory, is most likely to have errors in unstable directions, or unstable modes. The theory of indistinguishable states extends these ideas to imperfect models [13, 14, 18, 12] to obtain analyses and ensembles that can take into account model error.
In order to make the distinction between analyses obtained from shadowing and other means, and to distinguish the intended purpose of an analysis, the final state of a shadowing trajectory, or pseudo-orbit, will be called a shadowing analysis (SA) and any other state of a shadowing trajectory, or pseudo-orbit, will be called a non-causal shadowing verification (NCSV), to emphasize their purpose.

1.3 Gradient descent algorithms

A number of algorithms for finding shadowing trajectories have been around for some time. Indeed, 4D variational assimilation (4DVAR) can be viewed as a method for finding shadowing trajectories, and when coupled with weak constraints can be viewed as a method for finding shadowing pseudo-orbits.

The work reported here uses a class of algorithms called gradient descent methods [3, 4, 7, 8, 10]. These methods were originally introduced and demonstrated for simple chaotic systems, but only recently has a good theoretical understanding of their convergence been obtained [18]. New theoretical and experimental results have shown that gradient descent methods may be practical for NWP [15]. The experimental results, for example, showed that in a perfect model scenario high quality shadowing pseudo-orbits could be obtained with a T21L3 quasi-geostrophic model. These results motivated this investigation of shadowing with the Naval Operational Global Atmospheric Prediction System (NOGAPS) [1].

The aim of the experiments is to see if gradient descent shadowing algorithms can be successfully applied with NOGAPS, and if so, do they provide useful indications of model errors and better forecasts of non-causal analyses.

2 Implementation of a gradient descent shadowing algorithm

2.1 Theoretical basis of algorithm used

In the most abstract sense Gradient Descent (GD) is an optimization technique that minimizes a quantity by moving continuously in the direction of steepest descent. For the purposes of finding shadowing trajectories, the quantity to minimize is the indeterminism, which is a measure of how far a sequence of states is from being a trajectory of the model.

It is useful to adopt the moment of forecast time convention, in which time is measured relative to when a forecast is made. In this convention \( t = 0 \) refers the present (the moment of the forecast), \( t < 0 \) refers to the past, and \( t > 0 \) refers the future. At the moment of a forecast one only has information about the past and present. We will be frequently dealing with sequences of model states, for example, a sequence of analyses. It is usual to use an integer subscript \( i \) to denote the order in a sequence. In the moment of forecast time convention, the sign convention applies to the integer subscripts. For example, in a sequence of analyses, \( i = 0 \) refers to the present analysis, \( i < 0 \) past analyses, and \( i > 0 \) to analyses that will be obtained in the future.

Let \( f \) be a forecast model defined on a \( d \)-dimensional state space \( \mathbb{R}^d \), so that for \( x \in \mathbb{R}^d \), \( f(x) \) is the forecast for fixed time period, and let \( x = (x_{-w}, \ldots, x_0) \), denote an arbitrary sequence of \( w + 1 \) states \( x_i \in \mathbb{R}^d \), running from the past to the present, where \( w \) is called the window width. Note that the window width is a number, but it is often more convenient to think of it in units of time, that is, multiply \( w \) by the forecast time step of \( f \), so that window width is the time period between the first and last states in the sequence, or alternatively, how far into the past we are looking.

Define the mean squared indeterminism of \( x \) by

\[
I(x) = \frac{1}{w} \sum_{i=-w}^{-1} \|x_{i+1} - f(x_i)\|^2.
\]
Observe that $I(x) = 0$ if, and only if, $x$ is a trajectory. Furthermore, it can be shown that $I(x)$ has local minima only where $I(x) = 0$, see [13].

The gradient descent (GD) method of finding shadowing trajectories minimizes the indeterminism. To find a shadowing trajectory of a sequence of states $y = (y_{-w}, \ldots, y_0) \in \mathbb{R}^{(w+1)d}$, consider a sequence of states $x$ to be a function of a scalar $s$ and solve, in the limit of $s \to \infty$, the differential equation,

$$\frac{dx}{ds} = -\frac{\partial I}{\partial x} \quad x(0) = y,$$

where $I$ is a function of $x(s)$. That is, start at $y$ and move continuously in the steepest descent direction of $I$.

The GD method has only been proven to succeed under strict conditions [11, 18], which are almost never satisfied in practice. It is also known that in the perfect model scenario there are systems for which GD cannot succeed, because of the so-called glitch phenomenon [2, 5, 16]. Furthermore, GD certainly cannot succeed for imperfect models with large window widths. Despite the theoretical limitations of GD, it has been known since its first introduction [4, 10] that GD succeeds rather well for simple low-dimensional chaotic systems.

For NWP the theoretical limitations of GD are largely irrelevant. There are several reasons why. One reason is the observations and forecast models are far from perfect; the theoretical results only apply to perfect models and small observational error. A second reason is the local linearizations of high-dimensional NWP models have many eigenvalues that are close to zero; convergence of GD requires that all eigenvalues be non-zero, but close to zero eigenvalues imply slow convergence on these neutral, that is, slowly changing, modes [18].

In practice, the limitations of GD are not serious, because partial convergence of GD to a shadowing pseudo-orbit is often sufficient. Shadowing trajectories are not unique; once one is found, many more are easily derived. Even if a shadowing trajectory cannot be obtained, because the model is imperfect, or because convergence has not been achieved, then a shadowing pseudo-orbit $x$ that has less indeterminism than the initial $y$, will often provide useful information about model error and often provide a better initial state for forecasts.

Relaxing the requirements of shadowing allows practical application of GD with an approximate adjoint. Equation 2 implies knowledge of the adjoint of the forecast model. Explicitly,

$$\frac{\partial I}{\partial x_i} = \frac{2}{w} \times \begin{cases} -A_i(x_{i+1} - f(x_i)), & i = -w, \\ (x_i - f(x_{i-1})) - A_i(x_{i+1} - f(x_i)), & -w < i < 0, \\ (x_i - f(x_{i-1})), & i = 0, \end{cases}$$

where $A_i = df(x_i)^T$, that is, the adjoint, or transpose of the Jacobian derivative of $f$ evaluated at $x_i$. However, for high-dimensional systems, one may be able to obtain shadowing with limited, or even no, adjoint information. If $A_i$ is substituted by an approximation of the derivative, then it can be shown that convergence to a shadowing trajectory can still be achieved [15]. The approximation can be extreme, for example, setting $A_i = \lambda I$ where $I$ is the identity matrix and $\lambda$ and scalar, the so-called $\lambda I$ approximation. For a perfect model experiment with a T21L3 quasi-geostrophic model, quality shadowing was achieved with $\lambda = 1/2$ [15].

For NOGAPS, an adjoint of the dry dynamics is available and was used as an approximation of the full adjoint. Early NOGAPS experiments used the $\lambda I$ approximation with $\lambda = 1/2$, but since NOGAPS can compute the product of the dry adjoint and a vector for about the same cost as computing a forecast, it was considered worth taking advantage of the faster convergence, and probably better quality solutions, that the dry adjoint would provide.

Equation 2 can be reduced to an iterative method. When a model is imperfect, or when only a shadowing pseudo-orbit is required, there is no need to solve equation 2 accurately, or drive it to the
limit. Solving the integration by a fixed step Euler method reduces GD to the iteration
\[
x_i \mapsto x_i - \frac{2\Delta}{w} \times \begin{cases} 
-A(x_i)(x_{i+1} - f(x_i)), & i = -w, \\
(x_i - f(x_{i-1})) - A(x_i)(x_{i+1} - f(x_i)), & -w < i < 0, \\
(x_i - f(x_{i-1})), & i = 0,
\end{cases}
\]
where \(\Delta\) is the step size, and \(A(x_i)\) is a suitable approximation of the adjoint \(df(x_i)^T\).

2.2 The meaning of indeterminism

The quality \(I(x)\) we call the mean squared indeterminism in equation 1 is a mathematical tool, it is not a quantity in the physical sense. For example, the quantity does not have units, unless all the state variables are expressed in the same units, perhaps as energy. Alternative definitions of \(I(x)\) could be used in GD, but change of variables alone will not affect the results of GD; see Appendix A.1 for discussion of both points.

It does not make much sense to try to interpret the value of \(I(x)\), other than as measure of the average mismatch between states and forecasts. It is dangerous to talk of, or think of, this quantity as a measure of average forecast error. It is only valid to do so for a sequence of analyses, because analyses are causally related. States obtained from shadowing algorithms are not causally related, except for the very end state.

2.3 A robust algorithm versus better algorithms

At the outset of this project a number of decisions had to be made about the main objectives and how to achieve these. The main aim was to assess the feasibility of shadowing methods and get some broad understanding of what might be learnt and gained from using them. It was not considered necessary to devise the most efficient algorithm, rather implement a simple robust algorithm that could be used in a range of different tests.

It is a common feature of iterative optimization algorithms that there is a trade-off between the amount of computation and the accuracy of results, there is also often a trade-off between the rate of convergence and the stability of the algorithm. Early tests, discussed in Appendix A.2, showed that there were potential instabilities in shadowing algorithms, that would result in poor quality solutions, or worse, failure of convergence. It was already known from experiments with simple low-dimensional chaotic systems and a T21L3 quasi-geostrophic model, that algorithms could fail under certain circumstances. Tests at T21L9, T21L18, and T47L24 showed that NOGAPS could display similar instabilities, and also new instabilities that are discussed in Appendix A.2.

It was decided that at this early stage a robust implementation of a shadowing algorithm was more desirable than a fast, or efficient, implementation. This would allow using the same algorithm in a number of scenarios, with a range of windows, model resolutions, data initializations. Using always the same robust algorithm would allow comparison of results from different scenarios without fear of results being compromised by failure of the algorithm. The cost of reliability was more computation.

Probably the single most important decision was to use a fixed step-size iteration exactly as equation 4. Ideally the step size \(\Delta\) in the iteration 4 should be adjusted adaptively. See Appendix A.2 for detailed discussion of the reasons for this choice.

There are better iterative methods to obtain shadowing trajectories. For example, by using adaptive step sizes. Also, since \(I(x)\) is a quadratic form, there is a good chance that conjugate gradient (CG) methods [17] can be used. CG is often used to solve linear equations, and is very efficient at doing so. Minimizing \(I(x)\) is not so easily done, because \(I(x)\) is fundamentally nonlinear, and a Hessian is not available. However, the Hessian products required by CG can be approximated by two adjoint computations. There seems to be no reason why this cannot be done, but at the time of commencing this study there was doubt whether useful shadowing trajectories could be obtained
using any method, so it was felt to be more expedient not to invest time in a sophisticated iterative method before it had been established that the effort was worthwhile. There are also other more sophisticated techniques for finding shadowing trajectories that have been demonstrated in simple low-dimensional chaotic systems, however, at the present time it seems that these would require matrix decompositions impossible to compute at the scale of NWP.

2.4 Implemented algorithm

Based on preliminary computations it was decided to use the algorithm exactly as stated in equation 4, using the spectral coordinates, and implemented in Octave\(^2\).

The algorithm adopted used a step size \( \Delta = 0.1w/2 \), a 6 hour forecast model, and the dry adjoint approximation. Larger step sizes could have been used, but since instabilities were known to be a potential problem (See Appendix A.2) and it could not be predicted how the algorithm would perform with all envisaged experiments, this conservative value was used. Reasonable convergence was achieved in 30–50 iterations, which was acceptable for T47L24 experiments, but tedious for T79L30\(^3\). The forecast period of 6 hours was used because analyses were provided at this time-step and it seemed reasonable given there are real diurnal cycles. The T21L3 quasi-geostrophic model had shown that convergence rates in its three layers varied; its upper layer, for example, converged better with longer forecast periods. No investigation of different forecast steps was made with NOGAPS. Preliminary computations used the \( \lambda_I \) approximation [15] for the adjoint, and convergence was achieved. However, because of concerns about instabilities, and because of the dry adjoint computation costs about the same as a forecast, it was considered better to use the dry adjoint. The rate of convergence was faster with the dry adjoint, which moderates the extra cost. No comparison of the differences of shadowing with dry adjoint and \( \lambda_I \) approximation were made.

The algorithm is very easily coded. Since NOGAPS keeps all state information in files, several utility functions are needed to load and save states from files, compute the forecasts, and compute adjoint-vector products. Figure 3 shows an implementation of the algorithm in Matlab/Octave, essentially as it was used with NOGAPS. Note that the order of the two \( \text{if} \) statements is important, because \( dx \) is calculated in the previous pass of the inner loop. Note also the indeterminism can be computed from a cumulative sum of the \( dx \) squared within the inner loop.

In all experiments the state is in terms of the spectral variables. This might seem surprising, because the prognostic variables (vorticity, divergence, potential temperature, specific humidity, and surface pressure) have typical ranges that differ by several orders of magnitude between variables, and between model levels. The algorithm is quite robust to these differences, so all calculations were done without rescaling; see also Appendix A.1. Experiments with a T21L3 quasi-geostrophic model showed that this GD worked equally well when the state was expressed in spectral coordinates or grid-point coordinates; this experiment was not done for NOGAPS, because it was more convenient to work with spectral coordinates.

All experiments used a fixed number of iterations, usually \texttt{max\_iteration} was either 30, 50 or 100 iterations. Often optimization algorithms are given either absolute or relative accuracy goals, rather than a fixed number of iterations. For these preliminary experiments there was little idea of what convergence accuracy could be reasonably expected, so a fixed number of iterations were used. A fixed number of iterations also allowed comparison of convergence for different model resolutions.

\(^2\)Octave is free software, essentially equivalent and largely compatible with Matlab, available for Linux systems. Octave was used because KJ has extensive experience with it and little experience with Fortran.  
\(^3\)About 24 hours for 30 iterations on a 7 day window, 6 hour forecast interval, T47L24 using dual 2GHz processors. About 4 days for T79L30.
% GD for on-disk states
% dtg variable to specify datetime of 0-th state
% load_state(dtg,i) loads the i-th state relative to dtg
% save_state(dtg,i,x) saves x in the i-th state file
% forecast_state(dtg,i) computes forecast of i-th state and loads it
% compute_adjoint_product(dtg,i,v) computes product $A(x_i)'*v$ for i-th state

step_size = 0.1;
for iteration = 1:max_iteration
        y = load_state(dtg,-w);
        for i = -w:0
                x = y;
                if i>-w
                        x = x - step_size * dx;
                endif
                if i<0
                        y = load_state(dtg,i+1);
                        dx= y - forecast_state(dtg,i);
                        x = x + step_size * compute_adjoint_product(dtg,i,dx);
                endif
        save_state(dtg,i,x);
        end
end

Figure 3: An implementation of the shadowing algorithm in Matlab/Octave. Note that the order of the two if statements is important, because $dx$ is calculated in the previous pass of the inner loop.

3 Analysis data and forecast error

We begin by describing the data sets we use as analyses, then describe some simple geometrical experiments that reveal inconsistencies between the model dynamics and the analyses.

3.1 Data sets used to obtain analyses

The analyses first used were obtained by interpolating down to the model resolution used from the one degree analysis fields obtained from the operational NOGAPS T239 runs. Later spectral histories obtained from a NAVDAS data assimilation run of a T79L30 NOGAPS were used.

Three data sets were used. Two data sets were the one degree analysis fields, at 6 hour intervals obtained from the operational NOGAPS T239; for one period from 00Z01Mar2003 to 18Z31Mar2003, and a second period from 00Z01Oct2003 to 18Z31Oct2003. The third data set was spectral histories of a NAVDAS data assimilation into NOGAPS T79L30 for the period 00Z01Oct2003 to 18Z31Oct2003 at 6 hour intervals.

In all subsequent discussion and computations an analysis refers to a state in unweighted spectral coordinates. For the purposes of displaying graphs and tables the prognostic variables are sometimes scaled by a power of ten, and this will be indicated when done.

For the one degree field data the analyses were obtained by using NOGAPS to interpolate the one degree fields into spectral histories of the model’s resolution.

Unless otherwise stated computations are for the 7 day window 00Z01Mar2003 to 00Z8Mar2003 at 6 hour intervals, making 29 states.

3.2 Geometrical analysis of analyses

A simple geometrical analysis shows that there is a significant systematic difference between a forecast of an analysis and the verifying analysis.
Figure 4 shows the geometric relationship between an analysis $A_0$, its forecast $fA_0$, and the verifying analysis $A_1$, obtained by computing the separation of the three points in state space. Actually we have computed the separation component for each prognostic variable to avoid issues of their relative scale. The three separation distances define a unique triangle in state space. Having computed these triangles for each triplet of states in the window, the triangles are oriented so they lie in the same plane, with $A_0A_1$ edges aligned; see inset diagram of figure 4.

We should not be happy about plots like those in figure 4, because the length $b$ is large relative to $a$ and $c$, that is, the forecast error is large compared to the change in state over a 6 hour period. The triangles shown in figure 4 are real, in the sense that the three states are really related by such triangles under the five orthogonal projections shown. Ideally we want $fA_0$ close to $A_1$, relative to the change in state, that is, long thin triangles with length $b$ small relative to $a$ and $c$. For some variables $b > a$, which implies that the forecast is not better than persistence at forecasting those variables.

One might argue that although the length $b$ is not relatively small, the forecast error might be unavoidable random errors, but this does not seem to be the case. Consider the vector $A_0fA_0$ as being equal to the vector $A_0A_1$ plus a random vector, and suppose this random vector is a mean vector $v$ plus an random vector that has mean zero, (isotropic) variance $\sigma^2$, a correlation coefficient $C$ with $A_0A_1$, and uncorrelated with $v$. If the mean value of the components of $v$ is $\mu$, and $v$ makes an angle $\theta$ with the vector $A_0A_1$, then it can be shown (see Appendix A.3) that the expected triangles in figure 4 have

$$b \approx \sqrt{d(\mu^2 + \sigma^2)},$$

$$\cos \phi \approx -\left(\frac{\mu \cos \theta + \sigma C}{\sqrt{\mu^2 + \sigma^2}}\right),$$

where $d$ is the dimension of the state space, assumed large for the approximation to hold. The first expression shows how the mean and variance of the random errors determine $b$. It follows from the second expression, that if the forecast error has mean zero and is uncorrelated with $A_0A_1$ ($\mu = 0$ and $C = 0$), then triangles are approximately right angled. This is also true if the mean forecast error vector $v$ is perpendicular to $A_0A_1$. It is not possible to determine the three quantities $\mu$, $\sigma$ and $\theta$. 

\[9\]
from the known information, however, we can conclude that either, the mean error is not zero, or the random errors are correlated with the vector $A_0, A_1$, or both. Note that if the forecast errors were Gaussian, then $b$ would have a Chi-squared distribution with $d$ degrees of freedom. Since $d > 1000$, we do not expect to see much variation in the length of $b$, even if $\mu$ and $\sigma$ are large relative to $a$.

We conclude from the plots of figure 4 that there are systematic differences between the analyses derived from interpolation of one degree fields and the dynamics of the T47L24 model. This could be caused by a combination effects including:

1. The interpolation of the high resolution analysis fields to a lower resolution spectral state space, which results in biasing the state by unbalancing physical processes or over smoothing spatial and temporal variations;
2. The low resolution model is inadequate, even if a high resolution model was almost perfect, because it does not adequately represent small scale processes;
3. The data assimilation methods have intrinsic errors, perhaps because of over-smoothing observations, or because the analyses are not dynamically balanced atmospheric states, or because the analyses are not on the model’s attractor;
4. The model has intrinsic errors, perhaps because the parametrization is not appropriate for the model resolution, or because the model has more fundamental problems, perhaps in the model physics.

Possibilities 1 and 2 are no doubt a factor, but the evidence gathered so far does not suggest this is the whole story. Possibility 1 can be addressed by using states obtain from an assimilation cycle at the model’s resolution. Appendix A.5 compares results from a T79L30 model using interpolation and direct NAVDAS data assimilation. These results show that although there is a significant reduction in mismatch when data assimilation is used, there is still significant mismatch, so possibility 1 cannot be the whole story. Possibility 2 can be addressed by investigating the forecast error for increasing model resolutions. The results in appendix A.5 also show that increasing model resolution to T79L30 does not change the mismatch relative to the change in state, that is, the aspect ratio of the triangles are similar. Higher resolution experiments would be needed to better assess the importance of possibility 2.

Possibilities 3 and 4 are the most problematic from the point of view of data assimilation and forecasting, because even if we determine that this is the case, we need to determine what the systematic errors are. This is where shadowing plays a useful role. Later shadowing results suggest that possibilities 3 and 4 both make a significant contributions to mismatch.

### 3.3 Where the mismatch is greatest in the analyses

It is helpful to investigate the spatial characteristics of the forecast errors before moving onto shadowing. Define mismatch to be the forecast error vector $A_1 f A_0$ in figure 4 inset; mismatch is synonymous with the indeterminism between just two states. Figure 5 shows temporal and zonal averages of fields and their mismatch. We note that if a model were perfect, but observation, and so analyses, were imperfect, then one would expect averaged mismatch should be nearly zero. Non-zero averaged mismatch is indicative of either systematic analysis bias, or model bias. Where the averaged mismatch is nearly zero, the variance of the mismatch is still an indication systematic analysis error, or model error.

**Vorticity:** The averaging of the vorticity (left graph) shows the strength of the jet streams, and shows that the greatest variation of the vorticity is associated with the jet streams, which is not surprising. The averaging of the mismatch, shows no significant bias. However, there is significant variance at the northern upper boundary, which is not surprising because this is an artificial boundary, and so one should expect model error in the region. There is also significant variance associated with
Figure 5: From the analyses we obtain these zonal and temporal average of (on the left) the stated analysis field, and (on the right) its mismatch, using T47L24 model. The contour lines display the zonal and temporal average of the quantity and the shading displays the standard deviation about the mean. Note we use temperature, rather than the prognostic variable potential temperature, because it is more revealing.
the jet streams. This is not surprising, because the rapid dynamics here imply that small analysis
errors are amplified into large mismatches.

**Divergence:** Like vorticity, divergence shows significant mismatch problems with the upper
boundary. There is no significant bias, but also like vorticity, there is significant variability of
mismatch where the divergence is large or variable. Once again we attribute this to rapid unstable
dynamics having larger errors and amplifying small errors.

**Temperature:** Temperature is displayed here rather than potential temperature because otherwise
most of the variability is masked by height dependence. Here we see an overwhelming mismatch
problem over the surface of the Northern pole. Such are large bias and variation is mostly like due
to poor analysis or model error associated with the extremes of the late northern winter. We again
observe problems at the upper boundary. However, we also observe significant mismatch bias of low
level topical temperatures, with mean mismatch of one degree with a standard deviation of 1–1.5.
There is also much variance in the lower tropic and subtropic temperature mismatch. Unlike vorticity
and divergence, the variance of mismatch is not strongly correlated with variance of the temperature;
this could be an indication of model error or analysis bias.

**Specific humidity:** Like temperature, we see here significant mismatch bias and variance in the
lower level tropic and subtropic. It is not surprising that humidity and temperature have similar bias
and variance. Unlike vorticity and divergence, the variance of mismatch is not strongly correlated
with variance of the specific humidity. Like temperature this is a possible indication of model error.

4 Experiment results : Shadowing

Most of the experiments were performed first with the T21L9 NOGAPS, then T47L24. Some initial
calibration experiments were done with T21L9, T21L18, T21L24, and T21L9 with equally spaced
vertical levels. Some subsequent and on-going experiments are with T79L30. For consistency of
narrative, the results displayed here are for T47L24, because this has been most extensively studied,
but some results for other resolutions are given in Appendices A.4 and A.5.

4.1 Variables, units, indeterminism, and mismatch

As stated previously the algorithm calculations were done in unscaled spectral coordinates, because
the algorithm is robust to the significant difference in the range of these variables (see Appendix A.1),
and these were more convenient to use. However, for the purposes of display and interpretation, it
is convenient to rescale.

Many of the following graphs display a quantity called **indeterminism**, which has not been strictly
defined, and should be interpreted in graphs as follows. Equation 1 defines \( I(x) \) as a sum of squares
of euclidean distances, which are themselves sums of squares of variables. It is useful to separate
these into partial sums, for example, into a sum for each prognostic variable, or separate these further
into partial sums for each layer. Sometimes it is useful to consider the **mismatch** at each state of the
pseudo-orbit. In the following graphs, indeterminism refers to the square root of a partial sum, that
is, it is the component of the total Euclidean distance. It should be clear from the context which
component of the total indeterminism is being considered. The partial sums always include the 1/w
factor.

4.2 Convergence of fixed step-size algorithm

Figure 6 shows the typical convergence of the algorithm when initialized with analyses. There is an
initial rapid convergence in the first 10–20 iterations, followed by slower convergence. The convergence
can clearly differ for different fields and levels.

The change in the rate of convergence can be understood using a simple linear model of conver-
genence. If GD is close to its optimal solution \( x^* \), then one could linearize the descent about \( x^* \);
Figure 6: Typical convergence of algorithm for T47L24 model, 6 hour forecast step, initialized from 1 degree field of NOGAPS operational analyses, for period 00Z13Mar2003 to 00Z20Mar2003, that is, a window of 7 days or $w = 28$. The indeterminism here is computed and displayed for each prognostic field and vertical level separately.
for example, by appeal to the generalized Hartman-Grobman theorem [9]. Since GD involves a sum of squares, we can assume the linearization is a symmetric matrix. Hence, there exist orthogonal coordinates \( \delta \in \mathbb{R}^{(w+1)d} \), so that solving equation 2 gives \( \delta_k(s) = a_k e^{-\lambda_k s} \), and so,

\[
\text{Indeterminism}(s) \approx \left( \frac{1}{w} \sum_{k=1}^{m} a_k^2 e^{-2\lambda_k s} \right)^{1/2},
\]

for some \( a_k > 0 \) and \( \lambda_k > 0 \). The \( a_k \) represent the initial error at \( s = 0 \), and the \( \lambda_k \) eigenvalues of the linearization. Clearly, indeterminism will converge exponentially only if all the \( \lambda_k \) are identical. In order to obtain the convergence observed in figure 6, it can be shown, with a little experimentation, that one requires, in almost all cases, a few terms with large \( a_k \) and \( \lambda_k \), and many terms with small \( a_k \) and \( \lambda_k \). This gives the effect of a rapid initial decrease almost to zero, followed by slow subsequent decrease to zero. If the terms with small \( \lambda_k \) have relatively large \( a_k \), or there is a very large number of them, then one sees a different result, most notably seen in surface pressure graph of figure 6, where indeterminism decreases to a value considerably above zero and then only slowly decreases.

This simple linear model of convergence does not explain the entire picture. Such a model can only give monotonic decrease in the rate of convergence, but from figure 6, it is clear that the rate of convergence can vary non-monotonically, for example, specific humidity, level 19, red line. This is evidence of nonlinear effects on convergence.

This picture of convergence is consistent with there being a few modes with rapid growth or decay, and many modes with more neutral growth. It is known that GD shadowing algorithms should easily repair mismatches in rapid modes, and have most trouble with more neutral modes [18].

Convergence can be affected by model error. With an imperfect model, GD should initially converge to a shadowing pseudo-orbit that is the best representation of the analyses. Further convergence will be slow, and tend to move the shadow away from the analyses. There is no evidence of this occurring, because as we will see later the shadowing pseudo-orbits obtained are consistent with the analyses.

For NOGAPS T47L24, it should be expected to see convergence affected by neutral modes, perhaps made worse by systematic errors in the analyses, and affected by model error. The details cannot necessarily be isolated, but some of the details of the effects will be revealed in subsequent discussion.

### 4.3 Reduction of indeterminism

From figure 6 it can be seen that the final indeterminism in each field and layer is reduced to one quarter or less of its initial value. Figure 7 shows initial and final total indeterminism as a function of field and level. It can be seen that the reduction of in indeterminism varies with level.

In the initial analysis states, indeterminism tends to be relatively larger at several places: vorticity, divergence and temperature at the upper-most levels; vorticity around level 10, that is, around 200-250 mb; potential temperature and specific humidity at the lowest levels.

The final shadowing pseudo-orbit states have lower overall indeterminism, generally similar trends, but significant differences: the indeterminism in vorticity is more pronounced just below level 10, but reduced just above it; there is relatively less indeterminism in potential temperature and specific humidity at the lowest levels; the vorticity, divergence and temperature at the upper-most levels are relatively reduced.

It will be argued later (section 4.8) that the relatively larger indeterminism in the shadow near level 10 is probably related to neutral modes in the jet-stream, where as the relatively smaller indeterminism in that shadow at the lowest and highest levels can be attributed to correction of model error effects.
4.4 Variation of mismatch along a pseudo-orbit

Figure 8 shows how the mismatch varies along the initial and final pseudo-orbits; recall that mismatch is the contribution to indeterminism at each time step.

The initial analyses have almost constant indeterminism with time, except for some diurnal cycles. The final shadowing pseudo-orbit does not have such noticeable diurnal cycles, but shows significantly reduced indeterminism at the beginning, and also a smaller decrease at the end. Both of the these effects are well known [18], and are an artifact of the GD algorithm. This occurs because in the middle of the pseudo-orbit the algorithm propagates corrections in a forward and backward direction, where as, at the ends, the corrections are propagated in only one direction, see equation 3. The result is slightly faster convergence at the ends.

There is another way of interpreting faster convergence at the ends. With a perfect model it can be shown that shadowing trajectories tend to diverge from the true trajectory at the beginning and at the end. This divergence is in the stable directions at the beginning, and in the unstable directions at the end [13, 18]. Similar results appear to hold for imperfect models too. The rate at which the mismatch increases to its typical value, for example, the slope of mismatch curves over the first 48
hours in figure 8, can be related to the rate of convergence and divergence in the dynamics in the stable and unstable directions. This is also apparent in the discussion of the next section.

4.5 Effect of window width on shadow

Figure 9 shows the effect of window width on convergence of the shadowing algorithm. It can be seen that window width of at least 48 hours is necessary to obtain convergence to numerical accuracy for this T47L24 model.

![Figure 9: (a) Comparison of the distance between shadowing pseudo-orbits of T47L24 states after 30 iterations of algorithm for different window widths. The window widths are stated in the legend in hours. Distance is the separation from a reference shadowing pseudo-orbit obtained with window width of 168 hours. (b) Shows the base 10 logarithm of the separation distance, which reveals convergence to numerical accuracy for windows greater than 48 hours.](image)

The minimum window width is closely related to the faster convergence at the beginning and end of shadowing trajectories, as discussed in the last section, which is in turn is related to the rate of convergence and divergence of the model dynamics. In order to obtain a useful shadowing trajectory, the window needs to be sufficiently long that errors in the stable directions at the beginning have been eliminated, because these can adversely effect state estimates [12].

This minimum window width has interesting implications for techniques like 4DVAR; it implies that for this model 4DVAR would need to be applied to a window of at least 48 hours. This is a much larger window than 4DVAR is typically applied, and implies that 4DVAR is probably not eliminating certain dynamical errors. On the other hand, 4DVAR is too unstable to apply on such wide windows, and hence suggests the simple shadowing algorithm might be useful for improving state estimates.

It should also be noted from figure 9(b) that when the numerical accuracy limit is reached, seen as by the leveling-off of the separation distance, the shadowing pseudo-orbits obtained for different window widths are not necessarily the same, but are all comparable shadowing pseudo-orbits. The graph shows the separation distance between shadow pseudo-orbits and the reference pseudo-orbit obtain from a 168 hour window. Observe that for window of width 48 to 96 the shadowing pseudo-orbits appear to converge to the same trajectory, because all the curves level-off at the same distance from the reference. However, this distance is not zero. The 120 and 144 hour windows level-off at different distances. The fact that the distances do level-off implies that all the pseudo-orbits are more or less equivalent, at least once this leveling-off has occurred, that is, these pseudo-orbits all closely shadow each other.

All subsequent experiments have used a 168 hour, or 28 step, window. This was used because it is appears to be sufficiently long to allow convergence without end-effects for around two-thirds of the states in the shadowing pseudo-orbit. This allowed reliable determination of both causal and non-causal analyses.
4.6 Causes of residual indeterminism

It is an important question whether the residual indeterminism of the shadowing pseudo-orbit obtained from analyses is an indication of the shadowing algorithm not having achieved convergence, an indication of systematic errors in the analyses, or an indication of model error. This can be examined by applying artificial noise to states and seeing what the algorithm does in response to it. Two experiments were tried: adding gaussian white-noise to a previously obtained shadowing pseudo-orbit, so that it has similar indeterminism to the original analyses, and adding noise in the same manner to a trajectory of the model, obtained by running the model from a single analysis state. The results of these experiments are shown in figures 10 and 11. In the first experiment one expects to recover the shadowing trajectory and in the second experiment to recover the model trajectory.

![Figure 10](image1.png)  
Figure 10: The initial (a) and final (b) total indeterminism of each level, for a renoising of a shadow of analyses. Fields are scaled by stated amount to allow each to appear on the same graph. Note the two graphs have different indeterminism scales; the right graph fits inside the lowest tic-mark of the left graph.

![Figure 11](image2.png)  
Figure 11: The initial (a) and final (b) total indeterminism of each level, a noising of a true trajectory of model. Fields are scaled by stated amount to allow each to appear on the same graph. Note the two graphs have different indeterminism scales; the right graph fits inside the lowest tic-mark of the left graph.

The level of noise added to the shadowing pseudo-orbit and model trajectory was determined from the differences between analyses and shadow. We computed independently for each spectral variable the root mean squared difference between the analysis and shadow pseudo-orbits, that is, the mean is taken over the 28 states of the pseudo-orbit. The noise applied independently to each
spectral variable was Gaussian with mean zero and standard deviation equal to the computed root mean square difference.

The white noise produces different distributions indeterminism in the levels; most notable, the relative indeterminism of the specific humidity is considerably more for the original analyses, but relatively less for the potential tempeprature. This is almost certainly because the potential tempeprature and specific humidity differences are correlated. The initial indeterminism of the renoised shadowing pseudo-orbit is considerably more than that of the noised model trajectory, this happens because before adding the noise, the shadowing pseudo-orbit has its own residual indeterminism, whereas the model trajectory has none. The initial indeterminism of the noised model trajectory has relatively more indeterminism in vorticity and divergence around level 20, it is not clear what causes this effect. More runs will be require to determine if it is a random effect.

In all three cases, original analyses, renoised shadow, and noised model trajectory, the final indeterminism of all variables at all levels is reduced by approximately one quarter of its initial values; only the specific humidity of the analyses, and large indeterminism at the upper and lower boundaries, were reduced significantly more, almost an order of magnitude reduction.

This suggests that GD is performing more or less the same in the perfect model case (noised model trajectory) as it is in the imperfect model case (original analyses). This is encouraging, because GD has been demonstrated to obtain good quality shadowing trajectories in other perfect model scenarios, that is, low dimensional systems [13, 18] and a T21L3 quasi-geostrophic system [20], and low dimensional imperfect model scenarios [14].

However, it is not entirely clear what GD is doing, in particular, what is the relationship between the adjustments made and the residual indeterminism. The more detailed studies in the following sections provide more insight into why indeterminism varies the way that it does.

### 4.7 Geometrical analysis of shadow

Figure 12 shows the mismatch as a triangle diagram similar to figure 4; it is seen that mismatch is greatly improved. Not only is the mismatch $S_1 f S_0$ greatly reduced, the triangles are nearly right angles as would be predicted if mismatch had only random errors about the $S_1$ that were uncorrelated with $S_0 S_1$.

![Diagram](image)

Figure 12: Separation between a shadowing state $S_0$, its forecast $f S_0$, and the next shadowing state $S_1$. The lengths in plots are as illustrated in diagram. Compare with figure 4.

Following the analysis of section 3.2 we would conclude that shadowing pseudo-orbit obtained has only small random errors and relatively small systematic bias. It is true that the triangles are
closer to being right angled, simply because mismatched is reduced, but only the surface pressure shows significant systematic bias, with specific humidity and divergence to a lesser extent.

4.8 Where the mismatch is greatest in the shadow

By comparing the spatial characteristics of analyses and shadow fields and mismatches, we can obtain a better idea of what the shadowing algorithm has done. Figure 13 shows the temporal and zonal average of the shadow fields and their mismatch, and should be compared with figure 5. One should keep in mind that the shadow has mismatch of at least one quarter of that of the analyses, and the variance of mismatch is often even smaller, for example, for all fields the peak mismatch variance of the shadow is within, or below, the lowest colour band of the analyses’ mismatch variance.

Vorticity: Comparing the corresponding graphs of figures 5 and 13, one sees that the average vorticity of the shadow shows very little difference from the analysis, although there is slightly less variance in the shadow. The mismatch, however, shows significant difference. Not only is the magnitude less, but the mismatch problems at the upper boundary has been almost completely removed, that is, the shadowing algorithm has found states that although not significantly different from the analysis, they are consistent with the boundary constraints. More importantly, we observe that the variance of the mismatch of the shadow is much more concentrated in the jet stream region, with what appears to be a sharp transition at level 10. The concentration of the variance suggests that the shadowing algorithm has most difficulty in correcting indeterminism in the jet stream region. This could be a consequence of instabilities of the dynamics there, or because the dynamics there is governed by wave motions that have neutral growth, and hence will converge slowly in gradient descent. This needs more investigation.

Divergence: The divergence in the shadow has a number of differences from the analyses. Although the peak variance in the divergence field (left graphs) are the same, there is an overall reduction and there is significant reduction near the upper boundary and in the jet stream. Generally divergence is a more variable field, but this overall reduction in variance suggests that the model does not require this variance. This might be an indication of model error. The mismatch of the analyses at the upper levels are completely removed in the shadow. The peak mismatch variance is very narrowly concentrated in the jet stream region. Otherwise the variance of the mismatch is correlated with the variance of the divergence.

Temperature: The averaged shadow temperature field appears to completely remove the low level tropical mismatch bias seen in the analyses. This has resulted in an increase in tropical temperatures. Otherwise the averaged temperature fields are very similar, except for some slight changes in the location of the variance in the field in the mid-latitude northern hemisphere. The mismatch of the shadow still shows bias over the north pole, although the variance in much less, it appears more as a significant bias. This suggests there is some very delicate behaviour associated with this, or that the analyses are very much at odds with what the model wants to do here. We also note that the other locations of variance in the mismatch correlated strongly with the mismatch of divergence; presumably there is some dynamical link.

Specific humidity: As might be expected, the increased lower level tropical temperatures of the shadow are accompanied by increased moisture content. This change has removed the mismatch bias seen in the analyses; the shadow mismatch has no significant bias. The peak mismatch variance occurs around level 18. The location of the peak variance is curious, because for all other fields this is a location of minimum variance.

4.9 Differences of analyses and shadow

The previous subsection suggests that the main action of GD has been to produce a shadow of the analyses that is consistent with the model’s dynamics, and the only differences between the shadow and analyses are to accommodate these dynamical corrections. This proposition can be tested by
Figure 13: From the shadowing pseudo-orbit we obtain these zonal and temporal average and variance of (on the left) the stated field, and (on the right) its mismatch. The contour lines display the zonal and temporal average of the quantity and the shading displays the standard deviation about the mean. Note we use temperature, rather the prognostic variable potential temperature, because it is more revealing.
comparing the differences between the shadow and analyses, with the mismatch of the analyses.

Figure 14 shows the geometrical relationship of analysis mismatch $A_1 fA_0$ and shadow mismatch $S_1 fS_0$. This figure combines information contained in the triangle plots of figures 4 and 12, and includes new information showing the position of the shadow states relative to the analysis states.

![Figure 14: Separation between analyses $A_1$, shadowing states $S_1$ and forecasts of these, $fA_0$, and $fS_0$. This figure shows that the shadowing states are generally no further from analyses, then the mismatch of analysis forecasts. The exception is the surface pressure.](image)

Observe that shadowing states generally lie no further from the analysis than an analysis forecast lies from its verifying analysis. It is implicit in GD that it tries to find the shadowing trajectory closest to the analyses. Ideally one would hope the shadow lies approximately midway between $A_1$ and $fA_0$, but since section 3.2 implies there is significant inconsistencies between the analyses and model dynamics it is not surprising to see the shadowing states displaced away from both $A_1$ and $fA_0$, but slightly closer to $fA_0$. The only exception is the surface pressure, where $A_1$ and $fA_0$ are closer to each other than $S_1$ and $fS_0$ are to them. It is not clear why this happens.

Figure 15 shows the temporal and zonal averaged differences of the shadow and analyses. The left column of graphs should be compared with the right column of figure 5, that is, compare the differences with the analysis mismatch. Observe that there is good correlation between both sets of graphs in both arrangement and magnitude, and hence conclude that GD has obtained a shadow that is close to the best sequence of states of the model that are consistent with both analyses and the model dynamics.

It is important to ask whether the differences between shadow and analyses result from random observational errors, or systematic differences between the model and analyses; we can investigate this by looking at the global time averaged differences shown in the right column of figure 15. Here we show only the results for level 24, the lowest level of the atmosphere.

For vorticity the main average differences appear random for most of the globe, although may be a correlation with the more mountainous and polar regions. For example, Himalayas, west coast of South America, Antarctica, Greenland. The divergence corrections shows some association with mountainous regions, and a strong equatorial stripe. Temperature and specific humidity have significant systematic errors. We seen systematic temperature increases over the tropical oceans in the shadow. There are two or three very localized increases, west coast of South America, North Pole; the former is certainly a model error caused by spectral truncation. The moisture also shows increases that correlate strongly with temperature increases.

At other levels there are differing amounts of corrections, although at higher levels the averaged differences are less systematic. At level 11, where the mismatch correction in vorticity and divergence was largest, and where the largest residual is the shadow remains, it is seen (graphs not shown), that
Figure 15: On the left is shown for the shadow minus the analysis the zonal and temporal average and variance, and on the right is shown the temporal average of this difference for the stated fields and levels.
the differences are either mainly random errors, or move with the flow in such a way as to be averaged out over the window of one week used here. There is no obvious structure, except possibly an increased divergence over the Tibetan plateau.

4.10 Spatial-temporal structure of shadowing

In this section we investigate how shadow and analysis differ in more detail. It will be see that shadow states have more spatial-temporal structure than analyses. This is not entirely surprising, because analyses do not aim to be consistent with the model dynamics, whereas the shadow does. The greater coherent structure of the shadow does say something about the nature of the model’s attractor, which appears to play an important role when we later investigate the nature of forecast errors and shadowing.

Figure 16: Snapshot of analysis and shadow states. The top two panels compare the vorticity and the lower two panels compare the divergence. These are a snapshot of a day in the middle of the shadowing window, and for the lowest level of the model.

Figure 16 compare analysis and shadow states. On the large and synoptic scale there are no obvious difference, as the mean fields show in figure 15 would suggest. All the difference are in relatively small scale features, but even here the mean differences seen in figure 15 are not readily apparent. The most obvious differences are in the structures of strong frontal features, but there is obvious systematic difference; for example, in whether a front is a continuous feature or broken into a string of more locally concentrate features.

Figure 17 compares the spatial-temporal structure of the vorticity of analyses and shadow over the North-eastern Pacific. In this closer inspection two interesting distinctions are noticable. First it can be see that the shadow appears to have more low amplitude, small scale structure than the analyses, for example, compare central regions of corresponding pairs of frames. Secondly, the shadow appears to have more temporal coherence in the low amplitude, small scale structure. For example,
examining the tropical region, the lower two rows of grid boxes, it can be see that analyses has a stronger random component from frame to frames, whereas the shadow has a more coherent structure of horizontal bands. It is not clear whether this low amplitude, small scale structure is important or relevant characterization of the atmosphere. It probably is relevant characterization of the model’s attractor, that is, model trajectories evolve towards states having these structures, which probably has important implications to forecasting as we will see later. Whether these structures are relevant to the atmosphere will require comparison with atmospheric observations, which we have not done.

5 Experiment results : Forecasting

We now consider using the shadowing analysis for forecasting future analyses. From the preceeding discussion we conclude that the analyses are not dynamically consistent with the model, because there is a systematic difference between analyses and shadow, however, it cannot be determined whether this is due to analysis bias or model error. It seems from the following forecasting experiments that there are significant amount of both sysmetatic analysis error and model error. It is argued that the attractor of the model plays a significant role in forecast error.

So far we have applied shadowing as though the model was perfect, but the results so far might suggest otherwise. There are a number of ways to take model error into account, the most sophisticated methods require modifying the shadowing algorithm. Here we will first continue to act as though the model were perfect, which through its failure will add further evidence of model error, then try some crude error corrections.

5.1 Forecasting future analyses

Figure 18 shows the error in forecasting future analyses from various states. The figure shows that the shadowing analysis forecasts furture analyses as well as, sometimes slightly better than, the current analysis. It also shows that simple modifications to a shadowing analysis, to account for the known bias in analysis states, provide consistly better forecasts, although the improvement is marginal.

The important reference is the red line, which is the error of the forecast started from the current analysis. This line starts at zero at time zero, because it is the analysis.

Next note the green line, which is the forecast from the shadowing analysis; it is sometimes completely obscured by the blue line. This is an honest forecast, because it uses only the analyses upto that time. This forecast starts above zero at time zero, because it is different from the analysis.
Figure 18: Analysis forecast errors of the T47L24 model. Forecasts are started at 00Z08Mar2003, using the analysis, shadow state, the 48 hour non-causal shadow state. Plotted are unweighted distance between forecast and target states, except for specific humidity which the simple distance. We also plot errors with entire shadowing forecast is shifted by a constant vector at all time steps:
At the 6 hour forecast most of this difference is removed. The following details may not be significant but are included because they are difficult to read from graphs. For vorticity the forecast error is less than the analysis forecast, except for 0–12 and 24–48. For divergence the forecast is better beyond 6 hours. For potential temperature the forecast is better only beyond 72 hours, with comparatively worse forecasts before then. For specific humidity the forecast is better only for 48–72 hours and beyond 144 hours; for the first 48 hours the forecasts are as bad as for potential temperature. For surface pressure the forecast is worse up to 48 hours, but significantly better beyond 72 hours.

The important point to make is although the shadowing analysis is different from the analysis, beyond 6 hours it gives forecasts that are of comparable quality, that is, the differences are insignificant given the size of the errors. This is important because we have shown there exist states of the model that are consistent with both the analyses and dynamics of the model, and provide essentially equivalent forecasts.

For comparison figure 18 also shows, as the blue line, the forecast error for a non-causal shadowing verification (NCSV) at time zero. The NCSV used was from a 168 hour window shadowing pseudo-orbit, shifted 48 hours into the future. This forecast from the NCSV is not an honest forecast, because the NCSV uses information derived from analyses up to 48 hours into the future. However, if the model were perfect, it should be a better analysis, and hence give better forecasts. These forecast errors are almost indistinguishable from those of the shadowing analysis, although we will see in figure 19 that these forecasts are not the same trajectory.

5.2 Taking known analysis bias into account

The forecasts from the shadowing analysis and the NCSV shown in figure 18 are at something of a disadvantage, since we know there are systematic differences between the analyses and shadow. It is legitimate to suggest that these known biases should be added back to the shadow forecasts, and we find that doing so improves these forecasts.

We consider two simple ways to provide bias correction: adding the mean difference between analysis and shadow computed over the shadowing window; and adding the difference between the analysis and shadow analysis at $t = 0$. In both cases we add the correction to each forecast, this just shifts the forecast trajectory in state space a fixed amount.

Adding the mean difference should correct climatological differences between analyses and model, whereas adding the time-zero difference will correct some flow dependent differences. We would expect the latter to give additional advantages in the short term, which degrade as the flow changes.

The magenta line shows the forecast errors when the shadow analysis forecasts are corrected by the mean difference and the cyan line shows the forecasts errors when corrected by the time zero difference. Both forecasts are always better than the forecast from the shadowing analysis, and are generally better than the analysis forecasts. The time-zero difference correction is generally the better of the two, except for specific humidity, where the mean difference gives better forecasts beyond 24 hours.

We conclude that the time-zero difference corrects for most of the climatological difference represented by the mean differences, except for moisture related differences. This should be investigated further.

5.3 Forecasting non-causal verification states

It is common practice to assess forecast models by their ability to forecast future analyses. Shadowing provides a new kind of analysis, so it is appropriate to investigate how well the model forecasts these analyses. A discussion of model verification is beyond the scope of this report, but If one wants to verify shadowing analyses, then one should use the non-causal analyses, because these would be better estimates of the true model state if the model were perfect. Here we consider verifying against the
NCSV, the non-causal shadowing states. Our experiments give interesting, even surprising, results, which provide evidence of significant analysis and model error.

Figure 19 shows the forecast errors when forecasting future NCSV. This graph includes the error of the analysis forecasting future analyses for reference, upper red line. First note the blue line that shows the forecast error when the NCSV at time zero is used to forecast future NCSV. If the model were perfect, and the NCSV were equal to the true states, then the forecast error would be zero. If the model were perfect, but the NCSV were only close to the true states, then the forecast error should increase at an exponential rate, as small errors in the unstable
Figure 20: Schematic of the effects of model error on forecast error. (a) Shows how model error can result in linear growth of forecast error. With an imperfect model the non-causal states are not aligned with model trajectories as they would with a perfect model. (b) Shows illustrates how forecasts of non-causal states from an analysis can result in a forecast error that decreases then increases, and has less error than a forecast from a non-causal state. In this figure the states toward the top of the diagram represent states that are attractor-like, that is, they are on the model’s attractor and are balanced according to model dynamics. States toward the bottom of the diagram are observation-like, that is, they need not be on the attractor or balanced according to the model. Because the model is imperfect, shadowing states cannot be found near the attractor, they are always pulled toward the model attractor. When forecasts are made from each state at $t = 0$, the trajectories move toward the model attractor.

Note that the blue line has almost linear growth for at least 48 hours in all variables; this is evidence of model error. Figure 20(a) illustrates how model error can result in linear forecast error growth. To obtain linear error growth the forecast must be moving in a direction different from the direction from one NCSV to the next, and growth of unstable modes must be relatively small. The growth of unstable modes does seem to be relatively small. The lower red line in figure 19 shows the separation between the trajectories of the NCSV at time zero and the shadowing analysis. These two states are close, and should differ mainly in their unstable modes. The separation between the states does appear to grow at an exponential rate, but the growth is much less than the linear separation of the non-causal states.

The green line of figure 19 shows the forecast error of the shadowing analysis. This error is non-zero at time zero because it differs from the NCSV at time zero. The forecast errors grow very much the same as for the NCSV at time zero. Note that the separation of these two trajectories is not apparent in their forecast errors until about 144 hours.

The magenta line of figure 19 shows the forecast errors for the analysis when forecasting the NCSV: this graph is surprising for two reasons. Firstly, the forecast errors decrease for the first 24 hours before increasing, and, secondly, the forecast errors become less than those of the NCSV at time zero and remain so. One can understand how this might happen by considering the schematic of figure 20(b). As we have previously observed, the analyses have systematic differences from the shadowing states, and these differences are not consistent with the model dynamics. We can say that in some sense the analyses have characteristics that are observation-like, where as the shadowing states are more attractor-like, that is, closer to the attractor (or climatology) of the model. When forecasts are made from any of these states, the trajectories move towards the model attractor. It can be seen that for this to happen the trajectory from the analysis can first move toward the NCSV, pass them by, then move away. Having done so it might also remain closer to the NCSV, then the trajectory that begun from the NCSV at time zero.

A reasonable hypothesis would account for the behaviour of forecasts described here, is that the model has a very definite attractor, and that all forecasts first move onto this attractor and then move along attractor. This idea is illustrated in figure 21. Any dissipative dynamical system,
analyses
NCSA
model attractor
flow

Figure 21: Schematic of relationship between analyses, NCSV and model attractor. The analyses away from the model’s attractor. The NCSV are effectively a projection of the analyses on the model attractor. The flow on the model attractor is in a different direction from the general progression of the analyses.

like an atmospheric model, must have an attractor, although for our purposes we need only be concerned with a positively invariant set that is attracting [9]. A feature of such an attractor is that any dynamically unbalanced\(^4\) model state will move onto this attractor. We can think of the analyses as projections of “reality”, as represented by the observations, into the model space. These analyses are not consistent with the model dynamics, and so are not expected to lie on the model attractor. In figure 21 this is represented by having the analyses “above” the plane of the attractor. The shadowing states, however, by their very construction, lie in, or much closer to, the attractor; this expectation is supported by earlier evidence that the shadowing states are more dynamically consistent, for example, figure 12. We can view the NCSV as projections of the analyses onto the model attractor. Forecast trajectories from a NCSV will move in the attractor, whereas forecast trajectories from an analysis will down onto the attractor then along the attractor. Figure 20(b) can be understood as a projection of figure 21, whereas figure 20(a) represents just the attractor plan.

It can be argued that figure 21 is misleading; having a flat 2-dimensional attractor in a 3-dimensional space, when the model space has a very high dimension and the attractor may have a complex structure, especially on small scales. On the other hand, we argue that evidence so far, and some to come, suggests this is not too simplistic. Furthermore, it has been shown that the analyses and shadow have systematic differences, that is, the differences are not just random, for example, figure 15. Consequently, it not entirely misleading to represent the analyses as all lying more-or-less a fixed distance above NCSV on the attractor surface, rather than, random positioned relative to these.

Figure 22 presents a geometrical analysis of analyses, NCSV, and forecasts of these, that tests, and appears to confirm, our hypothesis about the nature of forecasts as explained by the existence of a model attractor and motions toward and along the model attractor.

Figure 22 uses the geometric fact that four points in a high-dimensional state space form a tetrahedral, and the shape of this tetrahedral is determined entirely by the distances between the four points. To create this figure we track four states: the forecast \(f^tA_0\), at time \(t\), of the analysis \(A_0\), the verifying analysis \(A_t\), the forecast \(f^tN_0\), at time \(t\), of the NCSV \(N_0\), and the verifying NCSV \(N_t\). At each time \(t\) there is a tetrahedra \(T_t\) defined by these four points. For each tetrahedra the vertex \(N_t\) of \(T_t\) is taken as a reference point, the edge \(N_tA_t\) as a reference line, and the plane through the reference line and the midpoint of \(f^tA_0\) and \(f^tN_0\) as a reference plane. We can then imagine transporting the tetrahedra so that they all share the same reference point, line and plane. Once this is done the two vertices \(f^tA_0\) and \(f^tN_0\) are projected onto the reference plane. Figure 22 shows the points paths of the \(f^tA_0\) and \(f^tN_0\) in the reference plane, where the vertical axis is the \(N_tA_t\) axis. Note that the vertices \(N_t\) are all at the origin of the graph, and the points \(A_t\) are shown on the vertical axis. Much of the information in this plot comes from previous, more conventional, plots, in particular, figures 14, 18 and 19.

Figure 22 provides evidence in support of our hypothesis, as represent in figure 21, by suggesting

\(^4\)Dynamically unbalanced can mean more than an unbalanced state in the sense of buoyancy. In this context it is synonymous with not being on the attractor.
Figure 22: Computed distances between an analysis $A_t$ and non-causal state $N_t$, at some forecast time $t$, and the forecasts of the initial analysis $F^t A_0$ and non-causal state $f^t N_0$ valid at that time. The plots are formed by computing the tetrahedral of the four points at each $t$, then aligning tetrahedra on the $N_t$ points, the $A_t-N_t$ axes, and the plane through this axis and the mid-point of the opposite edge, then projecting onto this plane.
there is a model attractor, more-or-less perpendicular to the $A_t N_t$ axis, that the $f^t N_0$ move in this attractor away from $N_t$, whereas the $f^t A_0$ move onto it. Under the hypothesis the model error will cause the forecast $f^t N_0$ to move in the attractor away from its verifying NCSV at a more-or-less constant rate. In figure 22 we observe the path of the $f^t N_0$, lower blue line, moves away from the reference axis more-or-less perpendicular to it and at a constant rate. This also suggests that the $A_t N_t$ axis is more-or-less perpendicular to the model attractor. Also note that the points $A_t$ are clustered on the reference axis, suggesting the analyses are a fairly constant distance from the model attractor. The fact the this curve of the $f^t N_0$ moves downward at first suggests the NCSV are not quite on the attractor, which is not unexpected given the shadowing algorithm was never converged to a trajectory. The most important observation is the path of the $f^t A_0$ moves, at first, quickly away from the $A_t$, and down towards where the path of $f^t N_0$ lies. This strongly suggests the $f^t A_0$ are moving onto a model attractor.

The important fact, regardless of whether our hypothesis is well-supported or not, is that the forecasts from analyses are more like shadowing states than they are like future analyses. This fact was already clearly demonstrated by figure 19.

5.4 Energy norm and synoptic errors

We finally consider forecast errors using the energy norm. So far we have generally used the Euclidean norm in spectral space. The energy norm is a spectral space norm too, but puts almost all its weight on the large and synoptic scale components. The euclidean norm, by weighting all components equally, puts more weight on the small scale, especial as the resolution of the model increases. By comparing the forecast error in these two norms it is clear that shadowing gain most of its advantage in forecasting small scale features and the tropics, where synoptic scale features are less pronounced.

Figure 23: Forecast errors in energy norms for all atmosphere excluding the tropical band 20°S to 20°N and below 500mb.

Figure 23 shows forecast errors in the energy norm, and its various components, excluding the tropics. This is figure should be compared with figure 19. The most significant difference in using
the energy norm is that the errors of the analysis forecasting the future analyses is less than the errors of the shadowing state forecasting the future NCSV. The sub-panels show that this due large contribution of rotational kinetic energy, and to a less extent the potential energy. This is not true of the divergence kinetic energy. The forecast errors of the shadowing states are still better for the first 24 hours.

Figure 24: Forecast errors in energy norms for just the tropical band 20°S to 20°N and below 500mb.

Figure 24 shows forecast errors in the energy norm, and its various components, for the tropics, 20°S to 20°N and below 500mb. We see here that the errors of the analysis forecasting the future analyses is greater than the errors of the shadowing state forecasting the future NCSV, as they were in figure 19.

We conclude from this comparison of figures 19, 23 and 24 that the better forecast performance of the analysis forecasting the future analyses in the energy norm is derived entirely from synoptic scale extra-tropical features. We have not investigated further the why this occurs. In any case, it should be observed that the analysis still forecasts NCSV better than it forecasts analyses in any metric.

6 Conclusions

This study of the application of gradient descent shadowing algorithms to NOGAPS appears to largely successful. It has demonstrated that the methodology can be practically applied, although it will need to be refined if it were to be applied to operational models. The methods appear to reveal that current analyses are not consistent with the dynamics of the forecast models, and that there appear to be systematic analysis errors that could be removed using shadowing methods. The shadowing methods also appear to reveal new information about the nature of the model attractor, how this attractor affects forecast errors, and possibly also reveal something about the nature of model error. There are a number of future research directions opened by this research, and a number of directions not fully explored by the experiments discussed here.
Acknowledgements

KJ thanks ONR for funding this project through the NICOP grant N00014-03-10389, also NRL, Monterey, and UWA, for their support. In particular, KJ thanks, in alphabetical order, Craig Bishop, Simon Chang, Les Jennings, Carolyn Reynolds, Thomas Rosmond, Leonard Smith, Steven Tracton, Joao Teixeira, and the many friends, colleagues, and assistants who helped in many ways.

References


Appendices

A.1 Change of variables and different $I(x)$

Gradient descent of indeterminism implies solving the differential equation

$$\frac{dx}{ds} = -\frac{\partial I(x)}{\partial x}, \quad (A-1)$$

$$I(x) = \sum_i (x_{i+1} - f(x_i))^T (x_{i+1} - f(x_i)). \quad (A-2)$$

Making the linear change of coordinates $x_i = C y_i$, it is straightforward to show that the above differential equation is equivalent to, that is, has the same solutions as,

$$\frac{dy}{ds} = -\frac{\partial J(y)}{\partial y}, \quad (A-3)$$

$$J(y) = \sum_i (y_{i+1} - g(y_i))^T C^T C (y_{i+1} - g(y_i)). \quad (A-4)$$

where $g = C^{-1} f C$ is the map $f$ expressed in the variable $y$. This implies it is not necessary to scale variables, or express them all in the same units.

Linear change of coordinates is not the same as changing the norm. Consider,

$$\frac{dx}{ds} = -\frac{\partial K(x)}{\partial x}, \quad (A-5)$$

$$K(x) = \sum_i (x_{i+1} - f(x_i))^T P^T P (x_{i+1} - f(x_i)). \quad (A-6)$$

It can be seen that

$$\frac{dx}{ds} = -\frac{\partial K(x)}{\partial x} \iff \frac{dx}{ds} = -P^T P \frac{\partial I(x)}{\partial x}, \quad (A-7)$$

where $P$ is a block diagonal matrix with repeated $P$ blocks. This significantly changes the solution.

One might be tempted to introduce change of norm $P^T P$ to reflect model uncertainty or stochastic dynamics as one does in various data assimilation techniques, but this should not be done with shadowing without very good reason. Experiments have shown that it is better to not to change the norm, than it is to try to estimate $P^T P$. 

34
A.2  Adaptive step size and instabilities

The gradient descent problem, equation 2, is inherently stiff, because the linearizations of forecast models have modes with small eigenvalues that may later become unstable. Stiff differential equation solves must employ adaptive step sizes, although we have not investigated whether stiff integration might give insight into iterative algorithms for finding shadowing trajectories.

Ideally the step size $\Delta$ in the iteration 4 should be adjusted adaptively. Since $I(x)$ is a quadratic form, it well known that the rate of convergence will slow as the solution is approached. The simplest method is to scale the step size by the gradient vector, that is, divide $\Delta$ by $\|\partial I / \partial x\|$. The results in a more uniform rate of convergence.

However, it was found that the above simple adaptive step-size scheme excited instabilities in the jet-stream. Although the scheme gave an initial convergence fast, almost linear convergence, after 10 iterations the convergence slowed rapidly. By 18 iterations the vorticity and divergence in the layers around 200–250mb gained a minimum and then began to increase.

A.3  Triangle diagrams

Let $x \in \mathbb{R}^n$ be a fixed vector and $y = x + v + \delta$, where $v$ is a fixed vector and $\delta$ is a random vector with independent components with mean zero, variance $\sigma^2$, and uncorrelated with either $x$ or $v$. Let $a = \|x\|$, $b = \|y - x\|$, $c = \|y\|$, $\|v\| = \mu \sqrt{n}$, and $x \cdot v = a \mu \sqrt{n} \cos \theta$. Then

$$b^2 = \sum_i (v_i + \delta_i)^2 = n\mu^2 + 2v \cdot \delta, \quad (A-8)$$

$$c^2 = \sum_i (x_i + v_i + \delta_i)^2 = a^2 + n\mu^2 + 2a\mu \sqrt{n} \cos \theta + 2(x + v) \cdot \delta. \quad (A-9)$$

If $\phi$ is the angle between $x$ and $(y - x)$, then by the cosine rule, $c^2 = a^2 + b^2 - 2ab \cos \phi$, and so

$$\cos \phi = -\left( \frac{\mu \cos \theta + x \cdot \delta / (a \sqrt{n})}{\sqrt{\mu^2 + \sigma^2 + 2v \cdot \delta / n}} \right) \quad (A-11)$$

Since $\delta$ is uncorrelated with either $x$ or $v$, then for large $n$, $x \cdot \delta / \sqrt{n} \approx 0$ and $v \cdot \delta / n \approx 0$.

If $\phi'$ the angle between $y$ and $(x - y)$, then

$$\cos \phi' = -\left( \frac{\mu \cos(\theta + \phi + \phi') - y \cdot \delta / (c \sqrt{n})}{\sqrt{\mu^2 + \sigma^2 + 2v \cdot \delta / n}} \right). \quad (A-12)$$

When $\mu = 0$, that is, there is no error bias, then we find $\phi$ and $\phi'$ are related by

$$\cos \phi = \frac{(x \cdot \delta) / a}{\sin \phi} = \frac{(y \cdot \delta) / c}{\sin \phi'} \quad (A-13)$$

Using the sine rule

$$\frac{\sin \phi}{c} = \frac{\sin \phi'}{b} = \frac{\sin \alpha}{b} \quad (A-14)$$

we have

$$\frac{(x \cdot \delta)}{(y \cdot \delta)} = -\frac{\cos \phi / c}{\cos \phi' / a} = -\frac{\sqrt{1 - (\sin \phi / c)^2}}{\sqrt{1 - (\sin \phi' / a)^2}} = -1, \quad (A-15)$$

and

$$\cos \phi = \frac{c}{a} = -\sqrt{1 + \frac{n \sigma^2}{a^2} + \frac{2x \cdot \delta}{a^2}}. \quad (A-16)$$
A.4 T21L9 results

Here are included for comparison some results for T21L9 NOGAPS, which was always run first as a test case for any new computation.

Figure A-1 is the equivalent of figure 4.
Figure A-2 is the equivalent of figure 18.
Figure A-3 is the equivalent of figure 19.

Figure A-1: T21L9 – Separation between an analysis $A_0$, its forecast $fA_0$, and the verifying analysis $A_1$.
The lengths in plots are as illustrated by the inset diagram, which shows how the triangle of states have been reoriented. One triangle is plotted for each consecutive pair of analyses in the window 00Z01Mar2003 to 00Z8Mar2003, 28 pairs in all.

A.5 T79L30 results

Here are included for comparison some results for T79L30 NOGAPS.

Figure A-4 is the equivalent of figure 4.
Figure A-5 is the equivalent of figure A-4, but using analyses from a NAVDAS assimilation into the T79L30 model. These analyses should not be affected by interpolation from operational T239 resolution down to T79 resolution. We note, however, that the triangles still show significant inconsistency between analyses and model dynamics.

Figure A-6 is the equivalent of figure 12, except that shadowing algorithm was initialized with NAVDAS analyses, and so should be compared with figure A-5. The algorithm was run for only 10 iterations, but already should significant movement toward a shadowing pseudo-orbit with smaller and random mismatch. The shadowing algorithm has not been properly calibrated for T79L30 in this computation, in particular the adjoint parameters are not optimal. It is clear that an instability in vorcity and divergence has been excited in the top two levels, which requires either decreasing the integration step, or increasing the dissipation in the these levels.

Figure A-7 is the equivalent of figure A-6 using NAVDAS data, it should be compared with Figure A-5.
Figure A-8 is the equivalent of figure 5.
Figure A-9 is the equivalent of figure A-8, but using analyses from a NAVDAS assimilation into the T79L30 model. One can see that these analyses still show significant mismatch bias and variability.
Figure A-10 is the equivalent of figure 13, except that shadowing algorithm was initialized with NAVDAS analyses, and so should be compared with figure A-9.
Figure A-2: T21L9 – Analysis forecast errors. Forecasts are started at 00Z08Mar2003, using the analysis, shadow state, the 48 hour non-causal shadow state. Plotted are unweighted distance between forecast and target states, except for specific humidity which the simple distance. We also plot errors with entire shadowing forecast is shifted by a constant vector at all time steps:
Figure A-3: T21L9 – Analysis forecast errors. Forecasts are started at 00Z08Mar2003, using the analysis, shadow state, the 48 hour non-causal shadow state. Plotted are unweighted distance between forecast and target states, except for specific humidity which the simple distance. We also plot errors with entire shadowing forecast is shifted by a constant vector at all time steps:
Figure A-12 is the equivalent of figure 15, this is, it is the difference between the NAVDAS analyses and shadow states for the T79L30 model.

Figure A-4: T79L30 – Separation between an analysis $A_0$, its forecast $fA_0$, and the verifying analysis $A_1$. The lengths in plots are as illustrated by the inset diagram, which shows how the triangle of states have been reoriented. One triangle is plotted for each consecutive pair of analyses in the window 00Z01Oct2003 to 00Z8Oct2003, 28 pairs in all.

Figure A-5: T79L30 – Separation between a NAVDAS analysis $A_0$, its forecast $fA_0$, and the verifying NAVDAS analysis $A_1$. The lengths in plots are as illustrated by the inset diagram, which shows how the triangle of states have been reoriented. One triangle is plotted for each consecutive pair of analyses in the window 00Z01Oct2003 to 00Z8Oct2003, 28 pairs in all.
Figure A-6: T79L30 – Separation between a shadowing state $S_0$, its forecast $fS_0$, and the next shadowing state $S_1$. The lengths in plots are as illustrated by the inset diagram, which shows how the triangle of states have been reoriented. One triangle is plotted for each consecutive pair of analyses in the window 00Z01Oct2003 to 00Z8Oct2003, 28 pairs in all.

Figure A-7: T79L30 – Using the NAVDAS data the separation between a shadowing state $S_0$, its forecast $fS_0$, and the next shadowing state $S_1$. The lengths in plots are as illustrated by the inset diagram, which shows how the triangle of states have been reoriented. One triangle is plotted for each consecutive pair of analyses in the window 00Z01Oct2003 to 00Z8Oct2003, 28 pairs in all.
Figure A-8: T79L30 – From the analyses we obtain these zonal and temporal average of (on the left) the stated analysis field, and (on the right) its mismatch, using T79L30 model. The contour lines display the zonal and temporal average of the quantity and the shading displays the standard deviation about the mean. Note we use temperature, rather the prognostic variable potential temperature, because it is more revealing.
Figure A-9: T79L30 – From the NAVDAS analyses we obtain these zonal and temporal average of (on the left) the stated analysis field, and (on the right) its mismatch, using T79L30 model. The contour lines display the zonal and temporal average of the quantity and the shading displays the standard deviation about the mean. Note we use temperature, rather than the prognostic variable potential temperature, because it is more revealing.
Figure A-10: T79L30 — From the shadow states we obtain these zonal and temporal average of (on the left) the stated analysis field, and (on the right) its mismatch, using T79L30 model. The contour lines display the zonal and temporal average of the quantity and the shading displays the standard deviation about the mean. Note we use temperature, rather the prognostic variable potential temperature, because it is more revealing.
Figure A-11: T79L30 – From the NAVDAS initialized shadow states we obtain these zonal and temporal average of (on the left) the stated analysis field, and (on the right) its mismatch, using T79L30 model. The contour lines display the zonal and temporal average of the quantity and the shading displays the standard deviation about the mean. Note we use temperature, rather the prognostic variable potential temperature, because it is more revealing.
Figure A-12: T79L30 – On the left is shown for the shadow minus the analysis the zonal and temporal average and variance, and on the right is shown the temporal average of this difference for the stated fields and levels.
Figure A-13: T79L30 – Using NAVDAS data. On the left is shown for the shadow minus the analysis the zonal and temporal average and variance, and on the right is shown the temporal average of this difference for the stated fields and levels.
Figure A-14: T79L30 – Analysis forecast errors. Forecasts are started at 00Z 8 Oct 2003, using the analysis, shadow state, the 48 hour non-causal shadow state. Plotted are unweighted distance between forecast and target states, except for specific humidity which the simple distance. We also plot errors with entire shadowing forecast is shifted by a constant vector at all time steps.
Figure A-15: T79L30 – Analysis forecast errors. Forecasts are started at 00Z 8-Oct-2003, using the analysis, shadow state, the 48 hour non-causal shadow state. Plotted are unweighted distance between forecast and target states, except for specific humidity which the simple distance. We also plot errors with entire shadowing forecast is shifted by a constant vector at all time steps.
Figure A-16: T79L30 NAVDAS – Analysis forecast errors. Forecasts are started at 00Z 08Oct2003, using the analysis, shadow state, the 48 hour non-causal shadow state. Plotted are unweighted distance between forecast and target states, except for specific humidity which the simple distance. We also plot errors with entire shadowing forecast is shifted by a constant vector at all time steps.
Figure A-17: T79L30 NAVDAS – Analysis forecast errors. Forecasts are started at 00Z 8 Oct 2003, using the analysis, shadow state, the 48 hour non-causal shadow state. Plotted are unweighted distance between forecast and target states, except for specific humidity which the simple distance. We also plot errors with entire shadowing forecast is shifted by a constant vector at all time steps.
A.6 Glossary

The following are definitions of terms as used in this report. This definitions may differ from common usage, particularly when common usage assumes a perfect model scenario, or fails to make the distinction between a system and a model.

- **System** A dynamical system completely defined by a state and an evolution operator, which takes the state at some time to the state at a future time. The system can be identified as the thing under study or being forecast. The details of the system are unknown, except in a perfect model scenario. The system can be thought of as reality. It is not necessary that the system can be expressed in a closed mathematical form.

- **Model** A dynamical system that is a mathematical representation, or approximation, of the system under study. In practice the model is always imperfect, and an approximation of the system. The model is usually implemented as a continuous flow, or a discrete time map of state space. The report only considers discrete time maps.

- **Perfect model scenario** The unrealizable situation where the model and system are identical as dynamical systems. In this scenario there is a truth, which when known, the model provides perfect forecast of all future states of the system.

- **Truth** A fictitious concept that only makes sense in a perfect model scenario. Truth is a state of the model that is equivalent to the system’s state at a given time. It allows making perfect predictions of the entire future.

- **Analysis** A state of the model obtained by assimilation of observational data into the model. In the perfect model scenario it is an estimate of truth, otherwise, it is just a state that is hoped will provide good forecasts. (In the report an analysis refers to a spectral history.)

- **Assimilation** A process that obtains a state of a model from observations of the system. The observations may be incomplete or inaccurate.

- **Temporal sequence** A sequence of states such that each state is associated with a distinct moment in time; the states are ordered by this associated time.

- **Trajectory** A temporal sequence of states of the model that evolve one into next, usually, either continuously in time, or as states equally spaced in time.

- **Pseudo-orbit** A discrete temporal sequence of states that is almost, but not quite, a trajectory, that is, each state almost evolves, or maps, into the next state.

- **Shadowing trajectory** A temporal sequence of states that are a trajectory and remain close to some target sequence of states, usually a sequence of analyses.

- **Shadowing pseudo-orbit** A temporal sequence of states that are a pseudo-orbit and remain close to some target sequence of states, usually a sequence of analyses.

- **Indistinguishable states** States of a perfect model that, given past observations, are indistinguishable from the truth; they are necessarily the final state of a shadowing trajectory. Also, states of an imperfect model that can be arrived at by a shadowing pseudo-orbit; usually one focuses on such states of high probability.

- **Shadowing analysis (SA)** The final state of a shadowing trajectory, or pseudo-orbit, that shadows the observations up to and including a target time. In practice, shadowing of analyses up to and including the analysis at the target time might be used.
- **Non-causal shadowing verification (NCSV)** A state at a particular time that lies within a shadowing trajectory, or pseudo-orbit, that shadows some window of past, present, and future observations.

- **Mismatch** The difference in a temporal sequence of states between a state and the forecast of this state from the preceding state under the model. Mismatch also refers to the magnitude of this difference. Mismatch may also refer to magnitude, or difference, for a restricted set of variables, for example, a particular prognostic variable or for a particular layer of atmosphere model. (Distinctions should be clear from context or labelled.) Mismatch is synonymous with the indeterminism between one state another in a temporal sequence.

- **Indeterminism** The root mean squared mismatch of a temporal sequence of states. May also refer to root mean squared mismatch of a restriction of variables, as with a mismatch.

- **Window** The view of a system, or model, over a particular period of time. Also the length of this time period, or the number of states in a temporal sequence that exactly spans this time period, where the states are usually equally spaced in time.