THE UNIVERSITY OF READING

A comparison of two methods for developing the linearization of a shallow water model

by

A.S. Lawless, N. K. Nichols and S.P. Ballard

Numerical Analysis Report 4/01

DEPARTMENT OF MATHEMATICS

A comparsion of two methods for developing the linearization of a shallow water model

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

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

[2] *Met Office*
London Road
Bracknell
Berkshire RG12 2SZ.

Numerical Analysis Report 4/01

Abstract

We develop the linearization of a semi-implicit semi-Lagrangian model of the one-dimensional shallow water equations using two different methods. The usual tangent linear model, formed by linearizing the discrete nonlinear model, is compared to a model formed by first linearizing the continuous nonlinear equations and then discretizing. Both models are shown to perform equally well for finite perturbations. However, the asymptotic behaviour of the two models differs as the perturbation size is reduced. This leads to difficulties in showing that the models are correctly coded using the standard tests. To overcome this difficulty we propose a new method for testing linear models, which we demonstrate both theoretically and numerically.

Acknowledgements

The work described in this paper was undertaken as part of a PhD project with The University of Reading while the first author was working at the Met Office. The first author gratefully acknowledges support of this project by the Met Office and by the Mathematics Department of The University of Reading.

Dr Mike Cullen is acknowledged for his original design of the Met Office's perturbation forecast model, on which the model in this paper is based. Thanks are also due to Dr Andrew Staniforth for his guidance in setting up the shallow water model used.

Contents

1 Introduction

A feature of many modern data assimilation systems is the need for a linearization of the nonlinear forecasting model. Such a linear model is used directly within incremental four-dimensional variational data assimilation (4D-Var) systems to model the evolution of a perturbation (Courtier *et al.* 1994). It is also appears within the extended Kalman filter as an explicit model of the background error covariances (Ghil and Malanotte-Rizzoli 1991, Thepaut and Courtier 1991) and is used within 4D-Var in its full, non-incremental form as a means of obtaining the adjoint model (for example Chao and Chang 1992, Li *et al.* 1994).

The usual method of obtaining the linear model is to linearize the discrete form of the nonlinear model. This is known as the *discrete* method and the discrete linear model formed in this way is called the *tangent linear model* (TLM). The adjoint model is then found by a transposition of the matrix representation of the discrete TLM. This method has the advantage that the TLM can be found by directly linearizing the nonlinear model source code. The adjoint model can then be found by transposing the TLM source code. This process is known as *automatic differentiation* and can be performed either by hand or using one of the automatic differentiation compiler tools which are available, such as TAMC or Odyssée (Chao and Chang 1992, Rostaing *et al.* 1993, Giering and Kaminski 1998, Bartholomew-Biggs *et al.* 2000).

For the incremental 4D-Var scheme being designed at the Met Office, an alternative approach to obtaining the linear model has been proposed. The continuous equations of the nonlinear model are first linearized to form a set of linear equations. These continuous linear equations are then discretized using a suitable numerical scheme to form the discrete linear model, called a *perturbation forecast model* (PFM). The adjoint model can then be obtained from the perturbation forecast model by a transposition of the PFM source code, so that the incremental 4D-Var scheme still contains the exact adjoint of the discrete linear model (Lawless 1996, Lorenc *et al.* 2000). We refer to this approach as the *semi-continuous* method. [1]

There are two principal advantages to the semi-continuous method. The first is based on the premise that although the tangent linear model is valid for infinitesimal perturbations, what we actually want to model are finite perturbations of the size of uncertainties

[1] The term semi-continuous is used to distinguish from the *continuous* method of obtaining an adjoint model, whereby the model is coded from the continuous adjoint equations.

in the initial conditions (Errico and Raeder 1999). A PFM can be designed with this aim in mind. Thus it can be based on physical principles and can make some small approximations to the true TLM. This is consistent with the original proposal of the incremental formulation of 4D-Var (Courtier *et al.* 1994) which proposed that the exact TLM could be replaced by a model close to it, provided that the error in this approximation is smaller than the approximation error made in neglecting the model error terms. The semi-continuous approach to developing a linear model allows such approximations to be made both in the equations of the linear model, before any discretization, by appealing to scale analysis and in the actual implementation of the numerical scheme. In this way savings can be made in the execution costs of the linear model and its adjoint. As numerical models become more complex such approximations are likely to become more important, since a direct linearization of the discrete nonlinear model is likely to lead to many small terms which are costly to evaluate, but add little information to the data assimilation process.

A second advantage to the semi-continuous approach is that it is possible to avoid some of the problems which occur when linearizing complex schemes. For example, it has been shown how problems may arise in the direct linearization of interpolation with semi-Lagrangian advection schemes or the direct linearization of iterative solution procedures (Polavarapu *et al.* 1996, Tanguay *et al.* 1997, Polavarapu and Tanguay 1998). By forming the linear model from the continuous linear equations, many of these difficulties are avoided.

In this study we examine whether a PFM formed by the semi-continuous approach can provide a linear model as accurate as a TLM. The model we use is a semi-implicit semi-Lagrangian model of the one-dimensional shallow water equations. The numerical scheme used contains both semi-Lagrangian advection and an iterative solution of an elliptic equation and so has some of the complexities of a full weather forecasting model. In the next section we set out the continuous equations for the nonlinear model and show their linearization for use in the PFM. Section 3 then describes the numerical schemes. We describe the scheme for the nonlinear model and show how we obtain the schemes for the TLM and the PFM. Section 4 discusses how we should measure the accuracy of the linear models. A weakness in applying present methods to test a PFM is identified and we derive a new method designed to test such a model. The two linear models are then compared numerically in Section 5. Further numerical experiments in Section 6 are used

to verify the new method for testing a PFM introduced in Section 4. Finally Section 7 summarizes the conclusions of this study.

2 Analytical models

2.1 Nonlinear equations

The model we consider is the one-dimensional shallow water system describing the flow of a single-layer fluid over an obstacle in the absence of rotation. The governing equations can be written

$$\frac{\mathrm{Du}}{\mathrm{Dt}} + \frac{\partial \phi}{\partial x} = -g\frac{\partial H}{\partial x}, \tag{1}$$

$$\frac{\mathrm{D}(\ln \phi)}{\mathrm{Dt}} + \frac{\partial u}{\partial x} = 0, \tag{2}$$

where

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

is the material derivative. In these equations $H = H(x)$ is the height of the bottom orography, u is the velocity of the fluid and $\phi = gh$ is the geopotential, where g is the gravitational constant and $h > 0$ the depth of the fluid above the orography. The problem is defined on the domain $x \in [0, L]$ and we let $t \in [0, T]$. The spatial boundary conditions are taken to be periodic, such that at any time t we have

$$u(0,t) = u(L,t), \quad \phi(0,t) = \phi(L,t), \quad H(0) = H(L). \tag{4}$$

The values of u and ϕ are specified everywhere at the initial time, with

$$u(x,0) = u_0(x), \quad \phi(x,0) = \phi_0(x). \tag{5}$$

The system (1), (2) has a steady state solution which obeys the relationships

$$\frac{u^2}{2} + \phi + gH = K_1 \tag{6}$$

and

$$u\phi = K_2, \tag{7}$$

where $u = u(x)$, $\phi = \phi(x)$ and K_1, K_2 are constants of integration which are independent of x and t.

2.2 Linearized equations

In order to form the set of linear equations we consider the fields u, ϕ as perturbations δu, $\delta\phi$ about a spatially and temporally varying basic state $\bar{u}$, $\bar{\phi}$ which satisfies the non-linear equations. Thus we have

$$u(x,t) = \bar{u}(x,t) + \delta u(x,t), \tag{8}$$

$$\phi(x,t) = \bar{\phi}(x,t) + \delta\phi(x,t). \tag{9}$$

These expressions are substituted into the nonlinear equations (1) and (2) and products of perturbations are neglected to give the linear equations. We obtain, for the linearization of the momentum equation (1)

$$\frac{\mathrm{D}\delta\mathrm{u}}{\mathrm{Dt}} + \delta u\frac{\partial\bar{u}}{\partial x} + \frac{\partial\delta\phi}{\partial x} = 0 \tag{10}$$

and for the linearization of the continuity equation (2)

$$\frac{\mathrm{D}}{\mathrm{Dt}}\left(\frac{\delta\phi}{\bar{\phi}}\right) + \delta u\frac{\partial(\ln\bar{\phi})}{\partial x} + \frac{\partial(\delta u)}{\partial x} = 0, \tag{11}$$

where the material derivative D/Dt is defined as in (3), but using the linearization state wind $\bar{u}$. The linearized system also has a steady state solution, which obeys relationships equal to the linearization of (6) and (7). Thus we have

$$\bar{u}\delta u + \delta\phi = \delta K_1 \tag{12}$$

and

$$\bar{u}\delta\phi + \delta u\bar{\phi} = \delta K_2, \tag{13}$$

with δK_1 and δK_2 constant. These solutions will be used to interpret the results of the linear models in Section 5.

3 Numerical models

The scheme for the numerical models is chosen to match as closely as possible the new integration scheme being developed for the Unified Model at the Met Office (Cullen *et al.* 1997, Davies *et al.* 1998). We employ a two-time-level semi-implicit semi-Lagrangian scheme, based on the scheme of Temperton and Staniforth (1987), but with an off-centred time averaging of the forcing terms along the trajectory, as in Rivest *et al.* (1994). In the scheme of Cullen *et al.* the mass continuity equation is treated with an Eulerian scheme,

but for this study we also treat this equation with a semi-Lagrangian discretization. This is similar to the scheme being used operationally in the GEM model of the Canadian Meteorological Centre (Côté *et al.* 1998). We use a staggered grid, with points at which the wind u is held being half a grid length from points at which ϕ is held. This reflects the Arakawa C grid being used in the Met Office scheme.

3.1 Nonlinear model

We indicate the arrival and departure points for the u variable by the subscripts au and du respectively and the arrival and departure points for ϕ by $a\phi$ and $d\phi$. The time discretization for the nonlinear model is then

$$\frac{u_{au}^{n+1} - u_{du}^{n}}{\Delta t} + (1-\alpha_1)\left(\frac{\partial \phi}{\partial x} + g\frac{\partial H}{\partial x}\right)_{du}^{n} + \alpha_1\left(\frac{\partial \phi}{\partial x} + g\frac{\partial H}{\partial x}\right)_{au}^{n+1} = 0, \tag{14}$$

$$\frac{(\ln\phi)_{a\phi}^{n+1} - (\ln\phi)_{d\phi}^{n}}{\Delta t} + (1-\alpha_2)\left.\frac{\partial u}{\partial x}\right|_{d\phi}^{n} + \alpha_2\left.\frac{\partial u}{\partial x}\right|_{a\phi}^{n+1} = 0, \tag{15}$$

where superscripts indicate the time level and the coefficients α_1, α_2 are time-weighting parameters chosen to lie in the interval $[0.5, 1]$. The numerical solution consists of the following steps:

1. First we calculate the position of the departure points x_{du} and $x_{d\phi}$ for each grid point. For this step we follow the method of Temperton and Staniforth (1987) and use a time extrapolation of the wind from the time level n and $n-1$ values.

2. The next step is to calculate the known terms of (14) and (15) at the departure points. The terms are first calculated at the grid points and then interpolated to the departure point using a cubic Lagrangian interpolation.

3. Having calculated the known terms we are left with a system of coupled equations for u and ϕ at time level $n+1$ at all spatial points. We eliminate u at each grid point to obtain an equation at each point x_i of the form

$$-C\phi_{i+1}^{n+1} + 2C\phi_i^{n+1} - C\phi_{i-1}^{n+1} + (\ln\phi)_i^{n+1} = R_i^n, \tag{16}$$

where

$$C = \frac{\alpha_1\alpha_2\Delta t^2}{\Delta x^2} \tag{17}$$

and R_i^n is a known right hand side calculated from time level n values of the fields. This is a discretization of an elliptic equation on the domain and is weakly nonlinear through the presence of the $\ln\phi$ term on the left hand side.

4. In order to solve (16) we must define a suitable iterative procedure. We first put

$$\phi_i^{n+1} = \Phi_{ref} + \phi_i', \tag{18}$$

where $\Phi_{ref} > 0$ is a constant reference value. Then rewriting (16) and adding a term ϕ_i'/Φ_{ref} to both sides of the equation, we obtain

$$-C\phi_{i+1}' + \Big(2C + \frac{1}{\Phi_{ref}}\Big)\phi_i' - C\phi_{i-1}' = R_i^n - \ln(\Phi_{ref} + \phi_i') + \frac{\phi_i'}{\Phi_{ref}}. \tag{19}$$

This equation is mathematically equivalent to the original equation (16). However the left hand side now consists of a strictly diagonally dominant tridiagonal matrix and is therefore invertible. The solution of the system can then be found iteratively by successively solving the series of equations

$$\begin{aligned} -C{\phi_{i+1}'}^{(m+1)} \;+\; & \Big(2C + \frac{1}{\Phi_{ref}}\Big)\phi_i'^{(m+1)} - C{\phi_{i-1}'}^{(m+1)} \\ = \; & R_i^n - \ln(\Phi_{ref} + \phi_i'^{(m)}) + \frac{\phi_i'^{(m)}}{\Phi_{ref}}, \end{aligned} \tag{20}$$

where m is the iteration count and ${\phi_i'}^{(0)} = 0$ for all i. This is a fixed point iteration which is guaranteed to converge for Φ_{ref} less than any value of ϕ_i at time level $n+1$ (Lawless 2001). Assuming that convergence occurs after M iterations we obtain updated values of ϕ_i from

$$\phi_i^{n+1} = \Phi_{ref} + \phi_i'^{(M)}. \tag{21}$$

5. As a final step we use these new values of ϕ_i to calculate values of u at each grid point at the new time level from (14), thus completing one time step of the scheme.

3.2 Tangent linear model

In order to form the tangent linear model we follow the normal procedure and differentiate the nonlinear model source code. The non-differentiable procedures within the semi-Lagrangian part of the scheme are treated by assuming that the perturbations do not move the departure point outside of the grid interval defined by the linearization state. For the cubic Lagrangian interpolation we are using this may lead to some errors in the linearization, but the error made by this assumption will be small compared to that made by assuming linearity (Polavarapu *et al.* 1996).

When we come to the linearization of the solution of the elliptic equation, a different method is used to derive the TLM. Instead of linearizing the iterative solution procedure

we linearize the discrete equation of the nonlinear model (16) and then solve this linear equation, thus following the normal procedure of automatic differentiation (Bartholomew-Biggs *et al.* 2000). The linear equation can be solved by using the same direct solver as is used in the nonlinear model.

3.3 Perturbation forecast model

To develop the perturbation forecast model we begin by taking the continuous linear equations (10) and (11) and seeking some suitable discretization. A comparison of these equations with the nonlinear model equations (1) and (2) reveals that they have the same structure, but with the linear equations having an extra term in which the perturbation wind multiplies the gradient of the linearization state. The most natural method of treating these terms in a semi-implicit semi-Lagrangian context is to discretize them as off-centred averages along the trajectory. The other terms in (10) and (11) can be discretized in the same way as the corresponding terms in the nonlinear model. Thus we obtain the scheme

$$\begin{aligned}\frac{1}{\Delta t}\left(\delta u_{au}^{n+1}-\delta u_{du}^{n}\right) \quad &+ \quad (1-\alpha_1)\frac{\partial \delta\phi}{\partial x}\bigg|_{du}^{n}+\alpha_1\frac{\partial \delta\phi}{\partial x}\bigg|_{au}^{n+1} \\ &+ \quad (1-\alpha_3)\left(\delta u\frac{\partial \bar{u}}{\partial x}\right)_{du}^{n}+\alpha_3\left(\delta u\frac{\partial \bar{u}}{\partial x}\right)_{au}^{n+1}=0, \qquad (22)\end{aligned}$$

$$\begin{aligned}&\frac{1}{\Delta t}\left(\left(\frac{\delta\phi}{\bar{\phi}}\right)_{a\phi}^{n+1}-\left(\frac{\delta\phi}{\bar{\phi}}\right)_{d\phi}^{n}\right)+(1-\alpha_2)\frac{\partial \delta u}{\partial x}\bigg|_{d\phi}^{n}+\alpha_2\frac{\partial \delta u}{\partial x}\bigg|_{a\phi}^{n+1} \\ +\quad &(1-\alpha_4)\left(\delta u\frac{\partial(\ln\bar{\phi})}{\partial x}\right)_{d\phi}^{n}+\alpha_4\left(\delta u\frac{\partial(\ln\bar{\phi})}{\partial x}\right)_{a\phi}^{n+1}=0, \qquad (23)\end{aligned}$$

where α_i are time-weighting coefficients for $i = 1, \ldots, 4$. The numerical scheme used is very similar to that used in the nonlinear model and proceeds as follows:

1. We first calculate the departure points x_{du} and $x_{d\phi}$. Since the perturbation fields are advected only by the linearization state wind, these departure points and hence their derivation is exactly the same as in the nonlinear model.

2. The next step is to calculate the known terms of (22) and (23) at the departure points. As for the nonlinear model, we calculate these terms at the grid points and then interpolate to the departure point with a cubic Lagrangian interpolation.

3. Having calculated the known terms we are left with a system of coupled equations for δu and $\delta\phi$ at time level $n+1$ at all spatial points. We find that we can eliminate δu to form a system of equations for $\delta\phi$ provided that $\Delta t < (\alpha_3|\partial\bar{u}_j/\partial x|)^{-1}$ for each

spatial point x_j (further details of this are given in Lawless 2001). This may seem to be an extra restriction which is not present in the TLM. However the iterative procedure by which we calculate the departure point itself imposes the restriction $\Delta t \max |\partial \bar{u}_j / \partial x| < 1$ (Pudykiewicz *et al.* 1995). Hence for values of α_3 in the range $[0, 1]$, this extra condition on the time step must hold in any case.

4. The resulting equations for $\delta\phi$ at time level $n+1$ form a diagonally dominant tridiagonal system, which can be solved using the same numerical solver as is used in the nonlinear and tangent linear models. The result can then be used to calculate updated values of δu at the new time level using (22).

4 Measuring the accuracy of the linear models

4.1 Standard measures of accuracy

Before trying to compare numerically the two different linear models, we must decide what measure of accuracy we will use to assess the models. The usual method of testing a TLM is to compare the evolution of a perturbation in the linear model to the evolution defined by the difference of two runs of the nonlinear model. We define $\mathbf{x}_0$ to be our model state vector at initial time t_0 and $\gamma\delta\mathbf{x}_0$ to be a small perturbation to this model state, where γ is a scalar parameter. We let S be the solution operator of our nonlinear model, such that at time t_n the model state $\mathbf{x}_n$ is given by

$$\mathbf{x}_n = \mathrm{S}(t_n, t_0, \mathbf{x}_0). \tag{24}$$

Then the perturbation evolved in the nonlinear model at time t_n is

$$\mathrm{N}_n[\gamma\delta\mathbf{x}_0] = \mathrm{S}(t_n, t_0, \mathbf{x}_0 + \gamma\delta\mathbf{x}_0) - \mathrm{S}(t_n, t_0, \mathbf{x}_0). \tag{25}$$

We compare this with the perturbation evolved in the linear model, which we write $\mathbf{L}(t_n, t_0)\gamma\delta\mathbf{x}_0$, where $\mathbf{L}$ represents the solution operator of either the TLM or the PFM.

In order to quantify the error we define the linearization error $\mathbf{E}_n$ of the linear model at time level n by

$$\mathbf{E}_n = \mathrm{N}_n[\gamma\delta\mathbf{x}_0] - \mathbf{L}(t_n, t_0)\gamma\delta\mathbf{x}_0. \tag{26}$$

The size of this error can then be compared to the size of the linear or nonlinear perturbation. If we compare to the linear perturbation, we have the relative error E_R of the linear

model, defined by

$$E_R = 100 \frac{\| \mathbf{E}_n \|}{\| \mathbf{L}(t_n, t_0)\gamma\delta\mathbf{x}_0 \|}, \tag{27}$$

for some given norm $\| \quad \|$ (Rabier and Courtier 1992). A comparison with the nonlinear perturbation gives the solution error E_S,

$$E_S = 100 \frac{\| \mathbf{E}_n \|}{\| \mathrm{N}_n[\gamma\delta\mathbf{x}_0] \|} \tag{28}$$

(Vukićević and Bao 1998).

A standard method of proving that a TLM is correctly coded is to show that the relative error E_R tends to zero as the scalar parameter γ is reduced. For a correct tangent linear this will be true, since if $\mathbf{L}$ is exactly equal to the first order part of the discrete nonlinear model, we have

$$\mathrm{N}_n[\gamma\delta\mathbf{x}_0] = \mathbf{L}(t_n, t_0)\gamma\delta\mathbf{x}_0 + O(\gamma^2) \tag{29}$$

and $\mathbf{E}_n$ is second order in γ. However, for a PFM, in general the linear model operator differs from the nonlinear model by a first order residual term, which we write $\mathbf{R}(t_n, t_0)\gamma\delta\mathbf{x}_0$, so that

$$\mathrm{N}_n[\gamma\delta\mathbf{x}_0] = \mathbf{L}(t_n, t_0)\gamma\delta\mathbf{x}_0 + \mathbf{R}(t_n, t_0)\gamma\delta\mathbf{x}_0 + O(\gamma^2). \tag{30}$$

Hence we find that as γ tends to zero, the relative error E_R tends to a non-zero constant which is determined by $\| \mathbf{R}(t_n, t_0)\delta\mathbf{x}_0 \| / \| \mathbf{L}(t_n, t_0)\delta\mathbf{x}_0 \|$ and so is dependent on the linearization state. Thus for the PFM we do not have the same objective test as we have for a TLM. This is not a problem for our present study, since we can compare with the true TLM. However, in a realistic 4D-Var system, such as that being developed at the Met Office, the PFM will be developed as an alternative to the TLM and so this comparison will not exist. Some other method must therefore be devised to test a PFM. We now propose a new method of measuring the accuracy of such a model, which does not require the use of a TLM. We use the nonlinear model to estimate the linearization error we would obtain if we did have an exact TLM. This can then be compared to the actual linearization error obtained using a PFM.

4.2 Estimating the tangent linear model error

We consider our discrete nonlinear model to be of the form (24) and expand the nonlinear model operator using a Taylor series. To do this we follow Rabier and Courtier (1992).

We assume that the vector $\mathbf{x}_0$ is represented by p components x^l and that the model S is represented by p scalar components S^l with $l = 1, \ldots, p$, so that

$$x_n^l = S^l(t_n, t_0, \mathbf{x}_0). \tag{31}$$

Then for a perturbation $\delta\mathbf{x}_0$ to $\mathbf{x}_0$ a Taylor expansion gives

$$\begin{aligned} S^l(t_n, t_0, \mathbf{x}_0 + \delta\mathbf{x}_0) &= S^l(t_n, t_0, \mathbf{x}_0) + \sum_{i=1}^{p} \frac{\partial S^l}{\partial x^i}(t_n, t_0, \mathbf{x}_0)\delta x_0^i \\ &+ \frac{1}{2!}\sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial^2 S^l}{\partial x^i \partial x^j}(t_n, t_0, \mathbf{x}_0)\delta x_0^i \delta x_0^j \\ &+ \frac{1}{3!}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \frac{\partial^3 S^l}{\partial x^i \partial x^j \partial x^k}(t_n, t_0, \mathbf{x}_0)\delta x_0^i \delta x_0^j \delta x_0^k \\ &+ h.o.t. \end{aligned} \tag{32}$$

Hence the linearization error $\mathbf{E}_n$ of a true TLM at time level n consists of the p components E_n^l given by

$$\begin{aligned} E_n^l &= \frac{1}{2!}\sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial^2 S^l}{\partial x^i \partial x^j}(t_n, t_0, \mathbf{x}_0)\delta x_0^i \delta x_0^j \\ &+ \frac{1}{3!}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \frac{\partial^3 S^l}{\partial x^i \partial x^j \partial x^k}(t_n, t_0, \mathbf{x}_0)\delta x_0^i \delta x_0^j \delta x_0^k \\ &+ h.o.t. \end{aligned} \tag{33}$$

It is this quantity that we now seek to estimate. The method proceeds as follows. We first run the nonlinear model from three different initial conditions, $\mathbf{x}_0, \mathbf{x}_0 + \delta\mathbf{x}_0, \mathbf{x}_0 + \beta\delta\mathbf{x}_0$, where β is a small scalar parameter. From these runs we can calculate the two nonlinear perturbations

$$\mathrm{N}_n[\delta\mathbf{x}_0] = \mathrm{S}(t_n, t_0, \mathbf{x}_0 + \delta\mathbf{x}_0) - \mathrm{S}(t_n, t_0, \mathbf{x}_0), \tag{34}$$

$$\mathrm{N}_n[\beta\delta\mathbf{x}_0] = \mathrm{S}(t_n, t_0, \mathbf{x}_0 + \beta\delta\mathbf{x}_0) - \mathrm{S}(t_n, t_0, \mathbf{x}_0), \tag{35}$$

consisting of the components $N_n^l[\delta\mathbf{x}_0], N_n^l[\beta\delta\mathbf{x}_0]$ respectively, with $l = 1, \ldots, p$. We now define a function $\boldsymbol{\mathcal{E}}_n$ by

$$\boldsymbol{\mathcal{E}}_n = \frac{\mathrm{N}_n[\beta\delta\mathbf{x}_0] - \beta\mathrm{N}_n[\delta\mathbf{x}_0]}{\beta^2 - \beta} \tag{36}$$

with components $\mathcal{E}_n^l$. Then using the Taylor expansion (32) we find that

$$\begin{aligned} \mathcal{E}_n^l &= \frac{1}{2!}\sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial^2 S^l}{\partial x^i \partial x^j}(t_n, t_0, \mathbf{x}_0)\delta x_0^i \delta x_0^j \\ &+ (1+\beta)\frac{1}{3!}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \frac{\partial^3 S^l}{\partial x^i \partial x^j \partial x^k}(t_n, t_0, \mathbf{x}_0)\delta x_0^i \delta x_0^j \delta x_0^k \\ &+ h.o.t. \end{aligned} \tag{37}$$

A comparison of this expression with (33) shows that for small values of β and small perturbations we have for each of the p components $\mathcal{E}_n^l \approx E_n^l$ and so $\boldsymbol{\mathcal{E}}_n \approx \mathbf{E}_n$. Since the expression (36) is calculated by using only the nonlinear model, it provides an estimate of the expected linearization error of a TLM without the need for such a model. This estimate can then be compared with the linearization error of a PFM to assess its accuracy.

We can quantify the size of the expected linearization error with respect to the perturbations by adapting the formula for the solution error of a linear model (28). Since we know the size of the nonlinear perturbation and have an estimate for the linearization error of a TLM, we can define the estimated TLM solution error $\hat{E}_S$ using

$$\hat{E}_S = 100 \frac{\| \boldsymbol{\mathcal{E}}_n \|}{\| \mathrm{N}_n[\delta \mathbf{x}_0] \|}. \tag{38}$$

This can then be compared with the actual solution error (28) calculated with $\gamma = 1$ for a PFM.

Although these formulae are analytically robust, it is necessary to test them numerically to see how well they hold in practice. Since for the model being used in this study we have the TLM available, we can verify these formulae experimentally. We do this in Section 6. First we compare how the two linear models behave.

5 Numerical experiments

The experiment we use to test the models is based on that described by Houghton and Kasahara (1968). For time $t < 0$ the fluid is at rest and the geopotential ϕ is equal to $\phi_0 - gH(x)$, with ϕ_0 constant. At $t = 0$ the fluid is impulsively set in motion with a constant velocity u_0 for all x. From this impulse a wave motion develops and moves away from the obstacle in both directions, while the solution in the vicinity of the obstacle converges to a steady state solution. This problem is defined over an infinite domain $-\infty < x < \infty$, but Houghton and Kasahara use a periodic domain with boundaries far enough away from the obstacle such that the asymptotic conditions are established around the obstacle before any of the wave motions can feed back into this area. We also follow this approach.

The experiment we run is based on Case A of Houghton and Kasahara's paper. The height of the obstacle H is given by

$$H(x) = H_c \left(1 - \frac{x^2}{a^2}\right) \quad \text{for} \quad 0 < |x| < a, \tag{39}$$

and $H(x) = 0$ otherwise. H_c is the maximum height of the obstacle and a is half the length over which the base of the obstacle extends. The domain is defined to be periodic

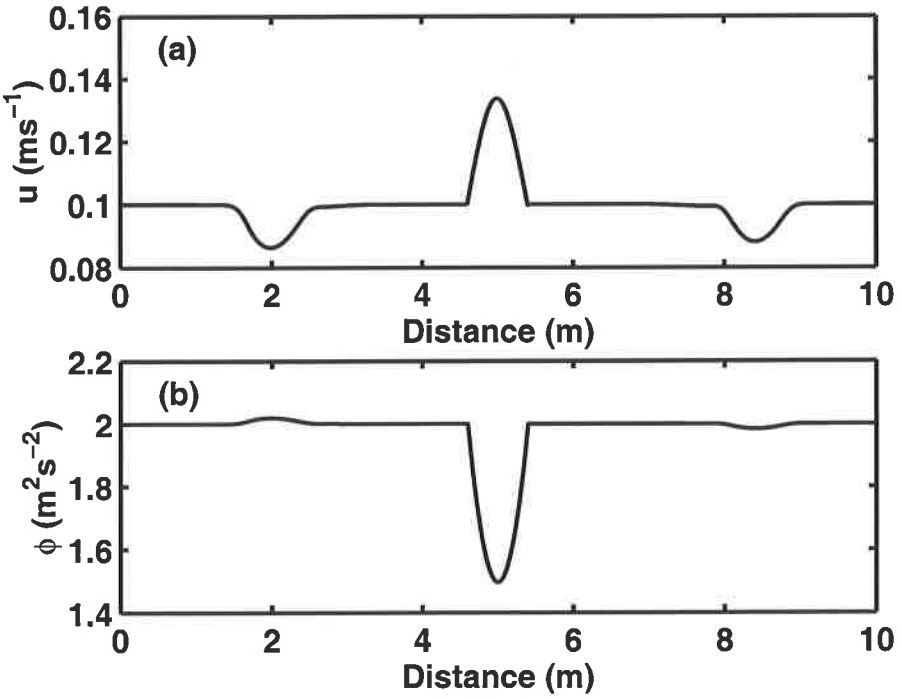

Figure 1: Fields of (a) u and (b) ϕ from the nonlinear base state run after 250 time steps.

over 1000 gridpoints, with a distance $\Delta x = 0.01m$ between them, so that $x \in [0m, 10m]$. The value of a is taken to be $40\Delta x = 0.4m$ and the height of the obstacle $H_c = 0.05m$. The parameters needed for the scheme we have chosen are taken to be $\alpha_1 = \alpha_2 = 0.6$ for the time weightings, the reference geopotential is $\Phi_{ref} = 1.5m^2s^{-2}$ and the extra time weightings for the PFM are $\alpha_3 = \alpha_4 = 0.6$. The gravitational constant g is set to $10ms^{-2}$ and the model time step Δt is $9.2 \times 10^{-3}s$.

For our nonlinear base state run we take the initial conditions at time $t = 0$ as $\phi(x) = \phi_0 - gH(x)$, with $\phi_0 = 2m^2s^{-2}$, and $u(x) = u_0 = 0.1ms^{-1}$. Perturbed nonlinear model runs are then generated by adding perturbations $(\gamma\delta u_0, \gamma\delta\phi_0)$ to the initial conditions, for various values of a scalar parameter γ. The initial perturbations are taken at $t = 0$ to be $\delta u_0 = 0.01ms^{-1}, \delta\phi_0 = -0.2m^2s^{-2}$, representing a change of 10% in each field. We compare the difference between the two nonlinear model runs with the output of both linear models initialized with the same perturbation. The run time is taken to be 250 time steps, corresponding to $2.3s$. The nonlinear base state fields at the end of the run are shown in Fig. 1. We first consider how well the two linear models represent the difference between the nonlinear model runs.

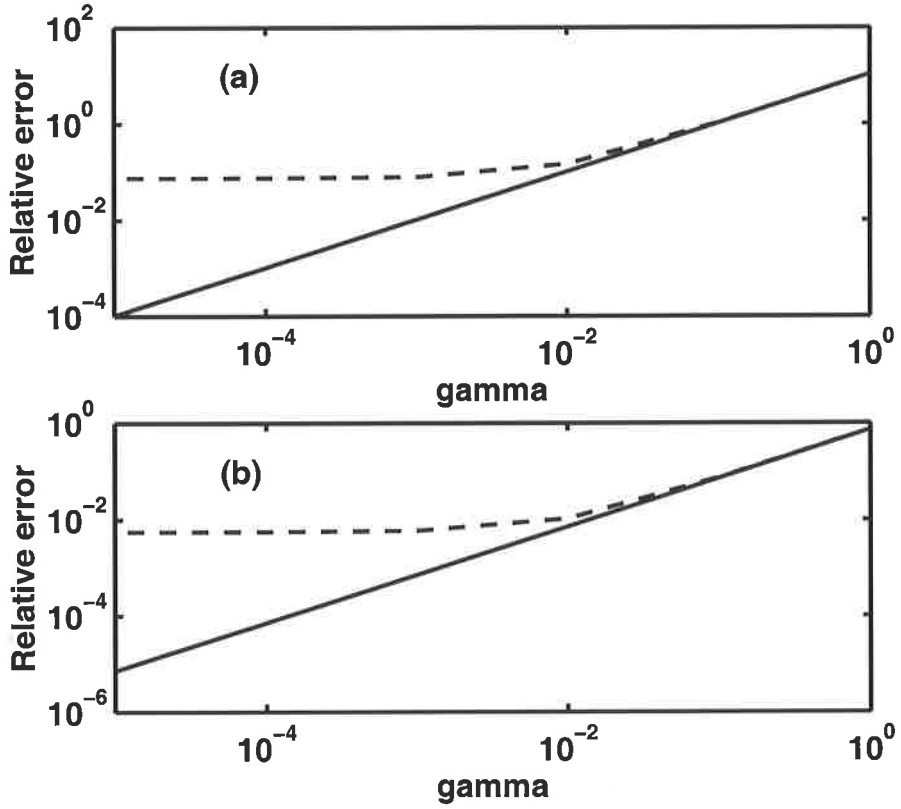

Figure 2: Plot of relative error E_R against perturbation size after 250 time steps for (a) u field and (b) ϕ field. The solid line is for the TLM and the dashed line for the PFM.

5.1 Tests of linear models

For the first experiment we examine the difference between the nonlinear and linear evolutions as the perturbation size is reduced. We set $\gamma = 10^p$ for $p = 0, -1, \ldots, -5$ and calculate the relative error (27) for the TLM and PFM after 250 time steps. In Fig. 2 we plot the relative error E_R against perturbation size for each linear model, calculated for the u and ϕ components of the error separately using the root mean square norm. We see that for the TLM the relative error tends linearly towards zero, showing that the model correctly represents the first order part of the discrete nonlinear model. For the PFM the error tends to a non-zero constant, as expected from the theory of Section 4, and so this test does not verify the correct coding of this model. However, it is encouraging that for the larger perturbations both linear models show the same relative error. This indicates that both methods of obtaining the linear model lead to equally valid approximations for reasonably sized perturbations.

We now consider the perturbation fields themselves for the experiment with the largest perturbations, that is for $\gamma = 1$. In Fig. 3 we compare the perturbation to the wind field

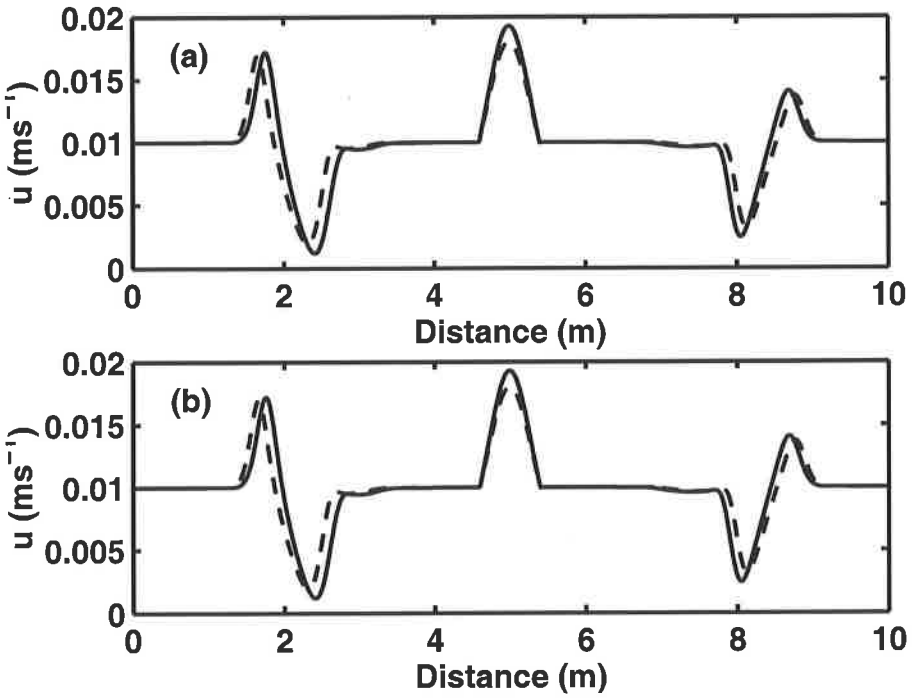

Figure 3: Comparison of nonlinear and linear perturbations of u after 250 time steps for (a) the TLM and (b) the PFM. In each case the solid line shows the nonlinear perturbation and the dashed line the perturbation from the linear model.

u from the nonlinear model at the end of the run with the perturbation fields from the two linear models. We see that the solution from both linear models follows the nonlinear evolution of a perturbation accurately. The models show very similar errors with respect to the nonlinear evolution, with an amplitude error in the stationary solution in the centre of the domain and a phase error in the outgoing gravity waves. Similar errors are also seen in the ϕ fields (not shown). We examine more closely the reasons for these errors in the two linear models.

5.2 Examination of linearization errors

The phase errors in the outgoing waves are found to arise solely from the perturbation to ϕ (further details can be found in Lawless 2001). We can understand this by calculating the dispersion relation for the original analytical system (1), (2), linearized around an equilibrium state U_0, Φ_0 which is constant in time and space. For this constant-coefficient system we find that the difference in phase speed between the two nonlinear model runs is given by

$$\delta U_0 \pm \left(\sqrt{\Phi_0 + \delta\Phi_0)} - \sqrt{\Phi_0} \right), \tag{40}$$

which is linear in U_0 and nonlinear in Φ_0. Thus we would not expect a linear model to represent correctly the phase of a gravity wave when the perturbation to ϕ is non-zero.

In order to understand the amplitude errors in the stationary solution, we refer back to the analytical properties of the equations stated in Section 2. There we found that the stationary solution of the nonlinear model satisfies

$$u(x)\phi(x) = K_2 \tag{41}$$

and for this problem we have $K_2 = u_0\phi_0$. From this we can define a nonlinear variation

$$\begin{aligned} V^N(x) &\equiv (\bar{u}(x) + \delta u(x))(\bar{\phi}(x) + \delta\phi(x)) - \bar{u}(x)\bar{\phi}(x) \\ &= \bar{u}(x)\delta\phi(x) + \delta u(x)\bar{\phi}(x) + \delta u(x)\delta\phi(x). \end{aligned} \tag{42}$$

Then using the asymptotic solutions we find that for the nonlinear problem the analytical value of $V^N(x)$ asymptotes to a constant value δK_2^N, given by

$$\delta K_2^N = u_0\delta\phi_0 + \delta u_0\phi_0 + \delta u_0\delta\phi_0. \tag{43}$$

We can define a similar variation for the linear models. The stationary solution satisfies (13) and for the problem being described in this section we have $\delta K_2 = u_0\delta\phi_0 + \delta u_0\phi_0$. Then defining the linear variation

$$V^L(x) \equiv \bar{u}(x)\delta\phi(x) + \delta u(x)\bar{\phi}(x), \tag{44}$$

we find that for perturbations calculated from the linear problem, the variation $V^L(x)$ asymptotes to δK_2. For the initial values used in the experiment of this section we find that $\delta K_2^N = -0.02$ and $\delta K_2 = 0$.

From the numerical solutions we calculate V^N and V^L around the centre of the domain. In Figure 4 we plot the variation over time of V^N calculated from the nonlinear perturbation and V^L calculated from the TLM over the interval $x \in [4.5m, 5.5m]$, together with their analytical asymptotic values. For each quantity we plot the absolute value. We see that after approximately 200 time steps both V^N and V^L asymptote to the analytical values of δK_2^N and δK_2 respectively. Also shown on Fig. 4 by the dotted line is the quantity V^N calculated using the perturbations from the TLM. The asymptotic value of this quantity does not equal that of the nonlinear model, but has a larger value. Thus above the orography the linear model asymptotes to a solution consistent with the linear equations. This explains the difference in amplitude between the linear and nonlinear solutions in the centre of the domain. Graphs of V^L and V^N from the PFM (not shown)

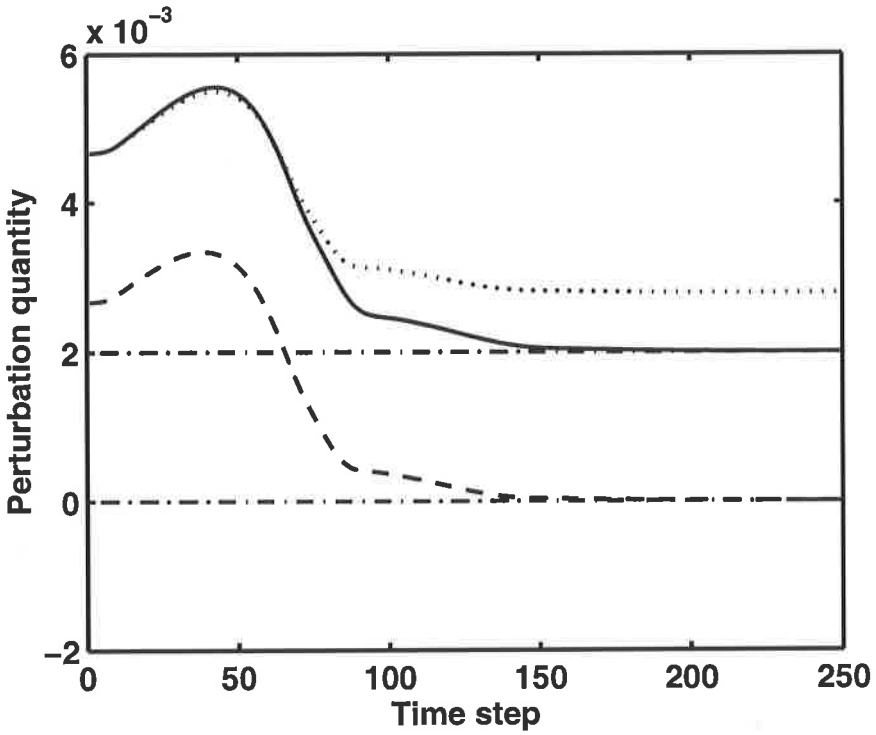

Figure 4: Variation of linear and nonlinear perturbations to $u\phi$ in the centre of the domain with time. The solid line is the value of V^N given by the nonlinear perturbations, the dashed line is the value of V^L given by the TLM and the dotted line is the value of V^N given by the TLM. The dot-dashed lines show the asymptotic values δK_2^N and δK_2. In each case the absolute value is shown.

are essentially the same as those for the TLM, indicating that the two linear models have the same underlying behaviour.

6 Verification of the linearization error estimate

We have illustrated by the theory of Section 4 and the experiments of Section 5(a) that the testing of a PFM is made more difficult by the fact that the linearization error does not reduce to zero as the perturbation size is reduced. However, we have also seen that for finite perturbations the PFM may be as good an approximation to the nonlinear evolution as a TLM. It is therefore important for a given perturbation to know how much of the linearization error we see from the PFM experiments is second order (and would therefore be present also with a TLM) and how much of the error is due to first order effects. The formulae derived in Section 4(b) were designed to measure this and since for this simple model we have the true TLM, we can measure the usefulness of these formulae.

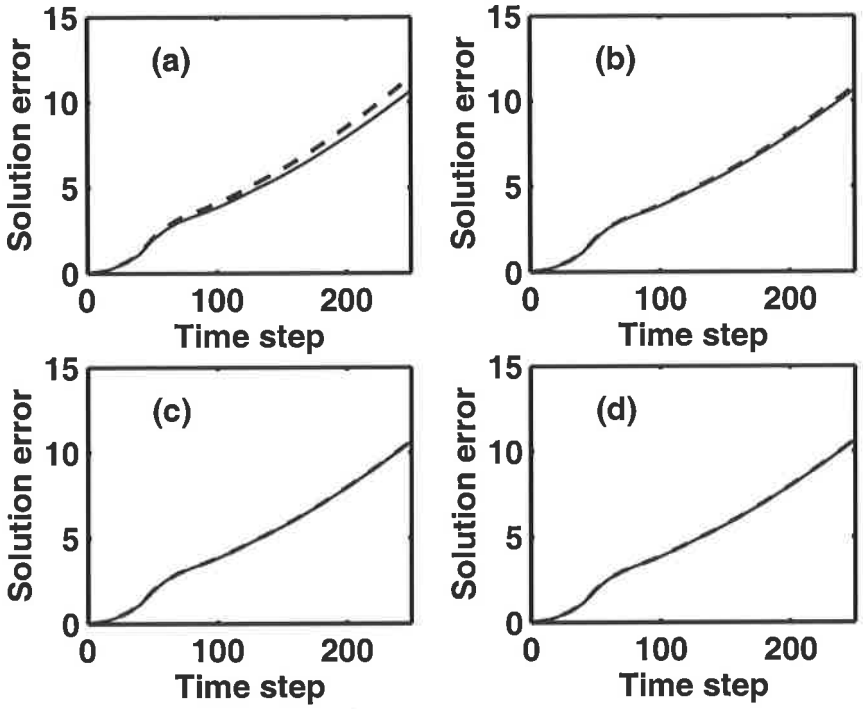

Figure 5: Comparison of the time evolution of the actual TLM solution error E_S (solid line) and the estimated TLM solution error $\hat{E}_S$ (dashed line) for estimates calculated using values of β equal to (a) 0.5, (b) 0.1, (c) 0.02 and (d) 0.01.

We run again the experiment of the previous section with a value of $\gamma = 1$. The non-linear model is then run again perturbed by a variation $(\beta\delta u_0, \beta\delta\phi_0)$, where we use values of $\beta = 0.5, 0.1, 0.02, 0.01$. For each value of β we calculate the estimated linearization error using (36) and the estimated solution error using (38). Figure 5 shows the time evolution of the estimated solution error $\hat{E}_S$ for the u field for the different values of β. We choose to plot the u field rather than the ϕ field since its solution errors are larger and thus we have a better test of the formulae. We see from Fig. 5 that for $\beta = 0.5$ the estimated solution error is a good approximation to the calculated TLM solution error only for approximately the first 70 time steps. As the value of β is reduced, the estimated solution error matches better the actual solution error calculated from the TLM.

One problem we have is how to choose a suitable value of β in practice when we do not have the true error to compare with. We must ensure that the value chosen is small enough to ensure an accurate estimate of the linearization error. However, inspection of (36) shows that if too small a value is chosen, then there may be problems with numerical rounding error. A method for choosing the parameter can be obtained by examining the evolution of the estimated solution error. We see from Fig. 5 that as the value of β is

reduced, the estimates of the solution error converge to the true value. Thus by calculating the estimated solution error for different values of β it is possible to see at what value the estimates begin to converge. This convergence can be seen even without reference to the true TLM value. The largest value of β for which the estimated solution error appears to have converged can then be taken as the most suitable value of the parameter with which to calculate the error in the fields. For this experiment we define convergence to be the point at which smaller values of β do not change the estimated solution error at the end of the run by more than 1%. This leads to a choice of $\beta = 0.02$.

We now look at the estimate of the linearization error itself. In Fig. 6 we plot the true linearization error $\mathbf{E}_n$ of the u field after 250 time steps and the estimated linearization error $\boldsymbol{\mathcal{E}}_n$ for values of $\beta = 0.1$ and 0.02. We see that qualitatively $\boldsymbol{\mathcal{E}}_n$ is a good estimate of the true linearization error for the chosen values of β. Figure 7 shows the differences between these estimates and the true error. We find that using a value of $\beta = 0.02$ compared to $\beta = 0.1$ introduces approximately a five-fold decrease in the maximum difference. The values of the maxima for these values of β are $3.43 \times 10^{-4} ms^{-1}$ and $6.78 \times 10^{-5} ms^{-1}$ respectively. This would be expected, since we see from (33) and (37) that

$$E_n^l - \mathcal{E}_n^l = -\beta \frac{1}{3!} \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \frac{\partial^3 S^l}{\partial x^i \partial x^j \partial x^k}(t_n, t_0, \mathbf{x}_0) \delta x_0^i \delta x_0^j \delta x_0^k + h.o.t. \tag{45}$$

which is proportional to β. The size of this difference should be compared to the maximum value of the linearization error over the domain, which is $4.5 \times 10^{-3} ms^{-1}$. The estimate of the TLM linearization error with the parameter value $\beta = 0.02$ is accurate to within 1.5% of the true value. For the stationary solution in the centre of the domain the linearization error is $1.3 \times 10^{-3} ms^{-1}$ and the estimate differs from this by approximately 0.25%. Thus it seems that with the value of $\beta = 0.02$, which was conjectured to be a suitable value of the parameter, the estimated error $\boldsymbol{\mathcal{E}}_n$ does indeed provide a good quantitative estimate of the TLM linearization error. Experiments with smaller perturbations have shown that this estimate remains valid as the perturbation size is reduced, provided that the solution error is greater than approximately 0.001%.

7 Conclusions

We have compared two methods for deriving a linear model, the discrete method, which produces a TLM, and the semi-continuous method, which produces a PFM. Using a semi-implicit semi-Lagrangian model of the one-dimensional shallow water equations we have

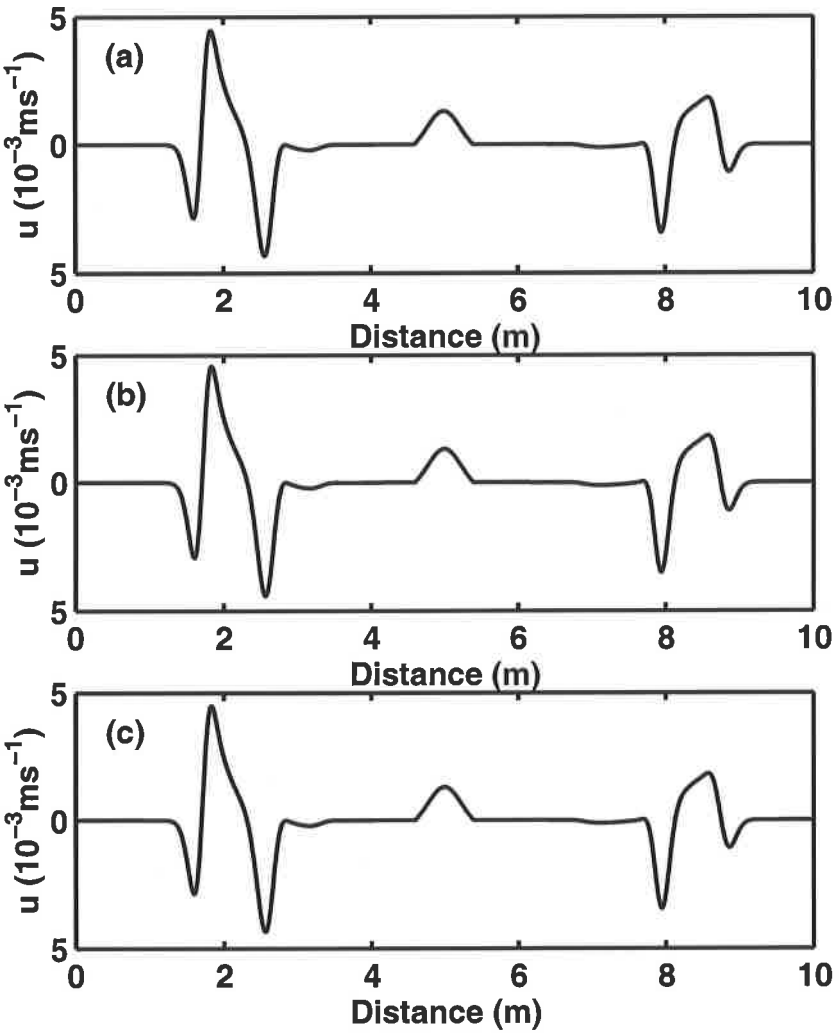

Figure 6: Linearization error for u field after 250 time steps. Plot (a) shows the true TLM linearization error $\mathbf{E}_n$. The other figures show the estimated TLM linearization error calculated using (36), with the parameter β equal to (b) 0.1 and (c) 0.02.

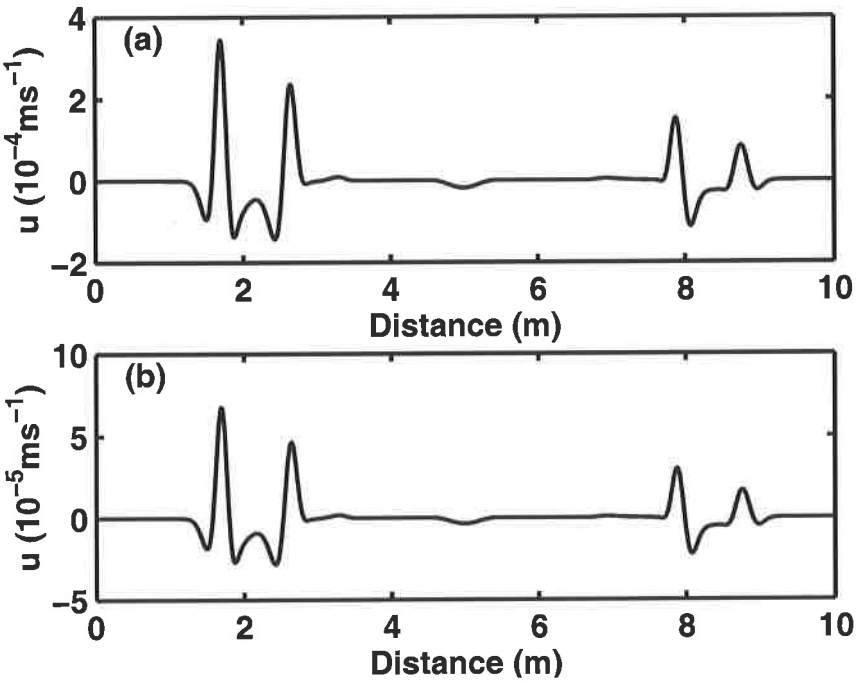

Figure 7: True TLM linearization error minus estimated TLM linearization error for values of β equal to (a) 0.1 and (b) 0.02.

shown that for finite perturbations the PFM can be as accurate as the TLM when compared to the difference between two runs of the nonlinear model.

We found that in the limit of small perturbations the PFM does not converge to the difference between the solution of two runs of the discrete nonlinear model, whereas the TLM usually does. In practice this is unlikely to cause a problem, since the linear model is either used to model finite-sized perturbations (in incremental 4D-Var) or to model explicitly a finite background error covariance (in the Kalman filter). It does however provide a problem for testing a PFM, since the usual method of showing that the linearization error tends to zero as the perturbation size is reduced cannot be applied. Within this study we have developed an alternative method of testing a PFM, which uses the nonlinear model to estimate the linearization error an exact TLM would give. This method has been illustrated theoretically and numerically and has since been used to verify tests of the three-dimensional PFM at the Met Office. Results from this work will be reported in a future paper.

References

Bartholomew-Biggs, M.C., Brown, S., Christianson, B. and Dixon, L., (2000). Automatic differentiation of algorithms. *Journal of Computational and Applied Mathematics*, **124,** 171–190.

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

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M. and Staniforth, A., (1998). The operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part I: Design considerations and formulation. *Mon. Weather Rev.*, **126,** 1373–1395.

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

Cullen, M.J.P., Davies, T., Mawson M.H., James, J.A., Coulter, S.C. and Malcolm, A., (1997). An overview of numerical methods for the next generation U.K. NWP and climate model, *in Numerical methods in atmospheric and oceanic modelling, eds. Lin, C.A., Laprise, R. and Ritchie, H.*, Canadian Meteorological and Oceanographic Society, 425–444.

Davies, T., Cullen, M.J.P., Mawson, M.H. and Malcolm, A.J., (1998). A new dynamical formulation for the UK Meteorological Office Unified Model, *in Recent developments in numerical methods for atmospheric modelling*, ECMWF seminar proceedings, September 1998, 202–225.

Errico, R.M. and Raeder, K.D., (1999). An examination of the accuracy of the linearization of a mesoscale model with moist physics. *Q. J. R. Meteorol. Soc.*, **125,** 169–195.

Ghil, M. and Malanotte-Rizzoli, P., (1991). Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33,** 141–266.

Giering, R. and Kaminski, T., (1998). Recipes for adjoint code construction. *ACM Trans. On Math. Software*, **24,** 437-474.

Houghton, D .D. and Kasahara, A., (1968). Nonlinear shallow fluid flow over an isolated ridge. *Comm. Pure Appl. Math.*, **21,** 1–23.

Lawless, A.S., (1996). A perturbation forecast model and its adjoint, *in Preprints of 11th Conference on Numerical Weather Prediction, American Meteorological Society*, 128–129.

Lawless, A.S., (2001). Development of linear models for data assimilation in numerical weather prediction. *PhD thesis*, Department of Mathematics, University of Reading.

Li, Y., Navon, I.M., Yang, W., Zou, X., Bates, J.R., Moorthi, S. and Higgins, R.W., (1994). Four-dimensional variational data assimilation experiments with a multilevel semi-Lagrangian semi-implicit general circulation model. *Mon. Weather Rev.*, **122,** 966–983.

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

Polavarapu, S., Tanguay, M., Menard, R. and Staniforth, A., (1996). The tangent linear model for semi-Lagrangian schemes: linearizing the process of interpolation. *Tellus*, **48A,** 74–95.

Polavarapu, S. and Tanguay, M., (1998). Linearizing iterative processes for four-dimensional data assimilation schemes. *Q. J. R. Meteorol. Soc.*, **124,** 1715–1742.

Pudykiewicz, J., Benoit, R. and Staniforth, A., (1985). Preliminary results from a partial LRTAP model based on an existing meteorological forecast model. *Atmos.-Ocean*, **23,** 267–303.

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

Rivest, C., Staniforth, A. and Robert, A., (1994). Spurious resonant response of semi-Lagrangian discretizations to orographic forcing: diagnosis and solution. *Mon. Weather Rev.*, **122,** 366–376.

Rostaing, N., Dalmas, S. and Galligo, A., (1993). Automatic differentiation in Odyssée. *Tellus*, **45A,** 558–568.

Tanguay M., Polavarapu S. and Gauthier P., (1997). Temporal accumulation of first-order linearization error for semi-Lagrangian passive advection. *Mon. Weather Rev.*, **125,** 1296–1311.

Temperton, C and Staniforth A., (1987). An efficient two-time-level semi-Lagrangian

semi-implicit integration scheme. *Q. J. R. Meteorol. Soc.*, **113,** 1025–1039.

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

Vukićević, T. and Bao, J-W., (1998). The effect of linearization errors on 4DVAR data assimilation. *Mon. Weather Rev.*, **126,** 1695–1706.