THE UNIVERSITY OF READING

ACCOUNTING FOR MODEL ERROR
IN
DATA ASSIMILATION
USING ADJOINT METHODS

by

A K Griffith & N K Nichols

DEPARTMENT OF MATHEMATICS

Accounting for Model Error in Data Assimilation using Adjoint Methods*

Anne K. Griffith[†]   Nancy K. Nichols [†]

Abstract

We consider a modification to the variational assimilation method, in which a correction term is added to the model equations. This correction term is determined in the assimilation procedure and produces a solution that is closer to the observed data and compensates to some extent for model error. The 1-d shallow water equations are used to illustrate the variational assimilation method, with and without the correction term, in the presence of model error.

1   Introduction

1.1   Data assimilation

In meteorological or oceanographic modelling, data assimilation is a means for combining observational data with a numerical model. The aim of data assimiltion is to produce a better estimate of the true atmospheric or oceanic state for climate records or to provide improved initial conditions for a model forecast. It is well known that nonlinear models such as weather forecast models can be very sensitive to their initial conditions, and so putting much effort into finding the right initial conditions is justifiable.

Various different methods for data assimilation have been developed over the last few decades. They range from simple schemes of inserting observed values directly into the model and smoothing in space, to sophisticated and expensive filtering methods, such as the Kalman Filter. Two particularly desirable features of a data assimilation scheme are that it should give a result which is consistent with the model dynamics, and that it should take into account the relative accuracies of the observed data and of the numerical scheme.

The *variational* method of data assimilation is attractive because it finds the statistically most likely solution of the model given the available observations and their error covariances. Many meteorological centres are currently developing this method for operational use. The variational assimilation method minimizes some criterion (a cost functional) expressing the difference between the observations and their corresponding model values, giving the observations weights inversely proportional to their error covariances. This minimization is subject to the constraint that the optimal solution must satisfy the model equations exactly.

The variational assimilation method involves iterating on the model initial conditions to solve the optimization problem. This iteration process uses the gradient of the cost functional with respect to the initial conditions. The required gradient can be expressed in terms of the model's adjoint equations, and so each iteration involves a run of the model followed by a run of its adjoint. Once the optimal initial conditions have been found,

---

[†]Department of Mathematics, University of Reading, U.K.

the optimal model state over the entire assmilation period can be found. This process of iterating to the optimal initial conditions is known as using the initial conditions as the *control variable*. Section 2 gives a more detailed mathematical formulation of the variational assimilation problem.

1.2   Model error

One disadvantage of the variational method is that the solution fits the model equations exactly, and hence doesn't allow for the fact that the model is imperfect. The difference between a forecast model state and the true atmospheric (or oceanic) state is known as *forecast error*. Forecast error can be split into two components, one from errors in the initial conditions, and the second from the way the model evolution misrepresents the true atmospheric evolution. This second type of error is called model error, and in general we have little idea of its form (Dee [2]). However, it is possible to identify particular examples of model error, which might constitute part of the total model error. Specific examples of model error include misspecification of model forcing terms (such as topography) or other model parameters, resolution errors, linearisation errors and errors from the boundary conditions or from problems with the poles. More general errors might arise from simplification or lack of knowledge of the physics.

In some cases, it might be known that a particular model is prone to a certain type of error, which it is not easy to correct due to a lack of knowledge of the true behaviour of the system or because of limited computing resources. In all cases, however, we can expect there to be a general component of model error which is unspecified and unmodelled.

Other data assimilation schemes, which do not hold the model equations as a strong constraint, do account for model error. In statistical schemes such as "Optimal Interpolation", the solution found depends on the relative accuracies of the observations and of a previous model forecast. In the Kalman Filter, model error is assumed to take the form of noise added to the model state at each timestep. The model error is assumed to be unbiased, independent of the model error at previous timesteps, and independent of the model state itself. It is realised that these assumptions are not very realistic, and yet it is hard to do better because little is known about such errors. For the same reasons, it is also difficult to construct an error covariance matrix for model error. However, the inclusion of model error in the Kalman Filter does allow the solution to "drift" from the model solution a certain amount depending on the size of the model error covariances, relative to the size of the observation error covariances. This motivates what we wish to do with the variational assimilation method; to slightly weaken the condition that the model equations must hold exactly, and to allow some drift of the solution towards the observations.

In Section 2 we firstly give a mathematical overview of variational assimilation, using the model initial conditions as the control variable. We then discuss the use of a correction to the model equations to account for model error, and follow Derber's method [3] of using a constant correction term instead of the initial conditions as the control variable. Section 3 describes the experiments carried out in an idealized situation to compare the performance of the two different control variables in the presence of different types of forecast error. Section 4 sumarises the results and Section 5 gives conclusions.

2   Variational Assimilation

In variational assimilation, the objective is to find that solution of the model which is as close as possible to the observations, taking into account the statistical likelihood of the observations. This is done by minimizing a cost functional expressing the difference between the observations and their model counterparts, with weighting matrices containing the error covariances of the different observations.

If the available observations (over the assimilation time period) are not sufficient to determine the optimal model state uniquely, it may be necessary to include in the cost functional the condition that the initial model state should also be close to some "background" initial state, maybe from a previous model forecast. Indeed, incorporating information from a previous forecast is usually desirable, and in this case weighting matrices can be chosen to satisfy Bayesian criteria for the most likely state given the likelihood of the observations and of the previous model forecast.

2.1   Mathematical formulation

We introduce a generalized nonlinear model

$$(1) \qquad \mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k),$$

where $\mathbf{x}_k \in \mathbb{R}^n$ represents the model state at time t_k, $\mathbf{u}_k \in \mathbb{R}^m$ represents model inputs at time t_k, (which might be model parameters or external forcing terms), and $\mathbf{f}_k$ represents the nonlinear evolution of the state from time t_k to time t_{k+1}. The variational data assimilation problem may be written

PROBLEM 1. *Minimize*

$$(2) \qquad \mathcal{J} = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{j=0}^{N-1} (\mathbf{h}\,(\mathbf{x}_j) - \mathbf{y}_j)^T \mathbf{O}_j^{-1}(\mathbf{h}(\mathbf{x}_j) - \mathbf{y}_j)\frac{\Delta t}{2}$$

subject to (1), where $\mathbf{x}_0^b$ is a background initial state and $\mathbf{B}_0$ is its error covariance matrix, $\mathbf{y}_j \in \mathbb{R}^p$ is a vector of observations at time t_j, and $\mathbf{h}$ is a function which interpolates to the observation positions and transforms to observed variables. The matrix $\mathbf{O}_j$ contains error covariances of the observational errors and the errors in the transformation $\mathbf{h}$ at time t_j.

We can reduce this constrained minimization problem to an unconstrained problem using Lagrange multipliers. Finding a minimum of $\mathcal{J}$ subject to (1) is equivalent to finding a saddle point of the Lagrangian functional $\mathcal{L}$, where

$$\mathcal{L} = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \sum_{j=0}^{N-1} (\mathbf{h}(\mathbf{x}_j) - \mathbf{y}_j)^T \mathbf{O}_j^{-1}(\mathbf{h}(\mathbf{x}_j) - \mathbf{y}_j)\frac{\Delta t}{2} + \boldsymbol{\lambda}_{j+1}^T(\mathbf{x}_{j+1} - \mathbf{f}_j(\mathbf{x}_j, \mathbf{u}_j)),$$

(3)

and the $\boldsymbol{\lambda}_j$ are Lagrange multipliers.

This still represents a huge optimization problem, and we seek to solve it iteratively. However, instead of iterating on the model state $\mathbf{x}_k$ for the whole time interval $[t_0, t_N]$, we iterate to the optimal initial condition. This makes sense since once the model equations and input $\mathbf{u}_k$ have been specified, the choice of initial condition uniquely determines the model state for the interval $[t_0, t_N]$. This technique is known as "reducing the control variable", and was introduced by LeDimet and Talagrand [5]. In this case the control variable is the initial condition.

To iterate to the optimal initial condition (using a minimization package), we need an expression of the gradient of $\mathcal{L}$ with respect to the initial condition, denoted here by $\nabla_{\mathbf{x}_0}\mathcal{L}$. This is given in the relationship

$$(4) \qquad \delta\mathcal{L} = <\nabla_{\mathbf{x}_0}\mathcal{L}, \delta\mathbf{x}_0>,$$

where $\delta\mathcal{L}$ is the first variation of $\mathcal{L}$, $\delta\mathbf{x}_0$ is a perturbation in $\mathbf{x}_0$ and $<.,.>$ is the relevant inner product (here the L_2 inner product in $\mathbb{R}^n$).

We now take the first variation of $\mathcal{L}$ and enforce that the following adjoint equations hold

$$(5) \qquad \boldsymbol{\lambda}_k = \mathbf{F}_k^T(\mathbf{x}_k, \mathbf{u}_k)\boldsymbol{\lambda}_{k+1} - \mathbf{H}_k^T\mathbf{O}_k^{-1}(\mathbf{h}(\mathbf{x}_k) - \mathbf{y}_k)\Delta t,$$

with

$$(6) \qquad \boldsymbol{\lambda}_N = 0,$$

where the matrix $\mathbf{F}_k$ is the Jacobian of $\mathbf{f}_k$ with respect to $\mathbf{x}_k$ and the matrix $\mathbf{H}_k$ is the Jacobian of $\mathbf{h}$ with respect to $\mathbf{x}_k$. The Lagrange multipliers $\boldsymbol{\lambda}_k$ are also called adjoint variables in this context. Then the expression for the first variation reduces to

$$(7) \qquad \delta\mathcal{L} = \{B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - \boldsymbol{\lambda}_0\}^T\delta\mathbf{x}_0,$$

from which we see that

$$(8) \qquad \nabla_{\mathbf{x}_0}\mathcal{L} = B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - \boldsymbol{\lambda}_0.$$

We now have everything we need to iterate to the optimal initial condition. From a guess of the initial condition, we run the model forwards and hence calculate the $(\mathbf{h}(\mathbf{x}_k) - \mathbf{y}_k)$ terms. Using these we run the adjoint equations backwards to find $\boldsymbol{\lambda}_0$, and so evaluate the gradient of the Lagrangian with respect to the initial condition. The gradient can be used in a minimization algorithm to improve the guess of the initial condition, from which the iteration can be repeated.

Each iteration towards the optimal initial condition involves a run of the model and a backwards run of its adjoint, and hence is potentially expensive. However, it is still an attractive method for the data assimilation problem, and the cost can be reduced by carrying out the iteration on a coarser grid, or by using a simplified model.

The adjoint equations involve the transpose of the Jacobian of the model equations and are forced by the difference between the observations and their model equivalent values. It is also possible to find the adjoint of a model directly from its computer code, using techniques described in [1], for example. This avoids the need to explicitly work out the adjoint of the model's discrete equations, but for a large meteorological or oceanographic model (which might involve 10^6 or 10^7 unknowns), it is still a huge task. For this reason there is much interest in automating the process using computational differentiation.

2.2   Using a correction term to account for model error

Our aim is to compensate for model error in variational assimilation by adding a correction term to the model equations. The model (1) now becomes

$$(9) \qquad \mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k) + \boldsymbol{\varepsilon}_k,$$

where $\boldsymbol{\varepsilon}_k$ is a correction term representing model error at time t_k. This time we take the "background" for $\boldsymbol{\varepsilon}_k$ to be zero, and the covariance matrix for the model error to be $\mathbf{Q}$.

The variational assimilation problem is now modified to

PROBLEM 2. *Minimize*

$$\tilde{\mathcal{J}} = \frac{1}{2}(\mathbf{x}_0-\mathbf{x}_0^b)^T\mathbf{B}_0^{-1}(\mathbf{x}_0-\mathbf{x}_0^b)+\frac{1}{2}\sum_{j=0}^{N-1}\varepsilon_j^T\mathbf{Q}^{-1}\varepsilon_j\Delta t+\sum_{j=0}^{N-1}+(\mathbf{h}(\mathbf{x}_j)-\mathbf{y}_j)^T\mathbf{O}_j^{-1}(\mathbf{h}(\mathbf{x}_j)-\mathbf{y}_j)\frac{\Delta t}{2},$$

(10)

subject to (9).

The solution to this minimization problem is the statistically most likely solution of the model (9) given the covariances of the observation error, the initial condition background error and the model error. It can be shown that for a linear model, the solution to this problem at time t_N is equal to the solution of the Kalman Filter at time t_N, [8].

Instead of using just $\mathbf{x}_0$ as the control variable as in Section 2.1, the minimization now involves using $\mathbf{x}_0$ and each of the ε_k as control variables. To make this approach more practical, it is necessary to reduce the control variables ε_k also, by forming some deterministic relationship between the ε_k. A general "model" for the evolution of the model error ε_k might take the form

(11) $$\varepsilon_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{u}_k, \varepsilon_k).$$

Now that we have assumed that there is a deterministic model for the ε_k, the model error no longer has a similar role to the role it has in the Kalman Filter, and the two methods no longer give an equivalent solution at time t_N. We could now write our model as the augmented system of equations

(12) $$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k) + \varepsilon_k$$
(13) $$\varepsilon_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{u}_k, \varepsilon_k),$$

and use the augmented initial vector $(\mathbf{x}_0^T, \varepsilon_0^T)^T$ as the control variable in variational assimilation.

As mentioned earlier, we have in general little idea of the nature of model error, and to proceed at all must make some big simplification about its form. The simplest approach is to assume that the model error is constant on the assimilation interval $[t_0, t_N]$. This approach to variational assimilation involving model error was introduced by Derber [3], who used a constant correction term instead of the initial condition as the control variable in his "variational continuous assimilation" method. In this case, the evolution of model error is given by

(14) $$\varepsilon_k = \varepsilon,$$

where ε is the quantity to be determined, a control variable.

So the variational assimilation problem is now to minimize (10) subject to (9) and (14), using the augmented vector $(\mathbf{x}_0^T, \varepsilon^T)^T$ as a control variable. Since the equation (14) is in this case trivial, it is not necessary to introduce extra adjoint equations. If $\tilde{\mathcal{L}}$ is the Lagrangian associated with (9) and (10), then the gradient of $\tilde{\mathcal{L}}$ with respect to $\mathbf{x}_0$ is as given in (8). The gradient of $\tilde{\mathcal{L}}$ with respect to ε can be expressed in terms of the same adjoint variables

(15) $$\nabla_\varepsilon\tilde{\mathcal{L}} = \mathbf{Q}^{-1}\varepsilon - \sum_{n=1}^{N-1}\lambda_n,$$

and the gradient of the "augmented" control variable $(\mathbf{x}_0^T, \varepsilon^T)^T$ is $(\nabla_{\mathbf{x}_0}\tilde{\mathcal{L}}^T, \nabla_\varepsilon \tilde{\mathcal{L}}^T)^T$.

The method of using both the initial condition and the correction term ε as control variables is now a straight forward modification of the method described in Section 2.1. However, the minimization algorithm must now be carried out on a vector of dimension $2n$, a large increase in the amount of work and memory required of the computing resources.

Other simple forms of correction term can be derived, in which (14) is modified to

$$(16) \qquad\qquad\qquad \varepsilon_k = s_k \varepsilon,$$

where the s_k are predetermined values used to give the correction term greater influence at certain parts of the assimilation period (see [3] for examples). Alternatively, the function $\mathbf{g}_k$ could represent piecewise constant model error, or model error with a spectral form, giving less restriction on the form of model error, but at the cost of increasing the dimension of the control variable. For these different forms of correction term, the process for finding the gradients would be unchanged, and the expression for the gradients would be a modification of (15).

In this study, we compare the variational assimilation method using different control variables; the initial condition, a constant correction term, and both together. This is carried out in an idealized situation in which model error is introduced artificially, as described in the next section.

3   Experimental set-up

The finite difference model used here is a nonlinear 1-d shallow water equation model including rotation and topography terms. It was introduced by Parrett and Cullen [7], who showed it to be suitable for simulating hydraulic jumps. Lorenc also used this model with variational assimilation [6].

The minimization procedure used is the M1QN3 code from the MODULOPT library of INRIA, a limited memory quasi-Newton method [4].

3.1   Introduction of model error

In this case, we suppose that the "true" model state is defined from a model run with particular initial conditions, topography and model parameters, and observations are taken from these true values. We then suppose that we have an imperfect version of the model or wrong initial conditions, and use the observations in a data assimilation algorithm to try to generate a model state closer to the true state.

In this case, there is an observation for each model variable, and so there is no need for a background term in the initial conditions, and this is omitted in this study. We also do not include observation errors, and the covariance matrices $\mathbf{O}_j$ and $\mathbf{Q}$ are equal to the identity matrix. Although an operational data assimilation scheme would have to account for observation error statistics, the aim here is simply to investigate the effects of particular types of forecast error using the different control variables.

For two different types of introduced model error, data assimilation is carried out using the constant correction term, initial condition or both as control variables. When using just the initial condition as the control variable, the correction term is fixed at zero, and when just the correction term is used as the control variable, the initial condition is assumed to be known. In operational data assimilation, of course, model error does not just involve misspecification of model parameters or forcing, but this gives some way of testing the
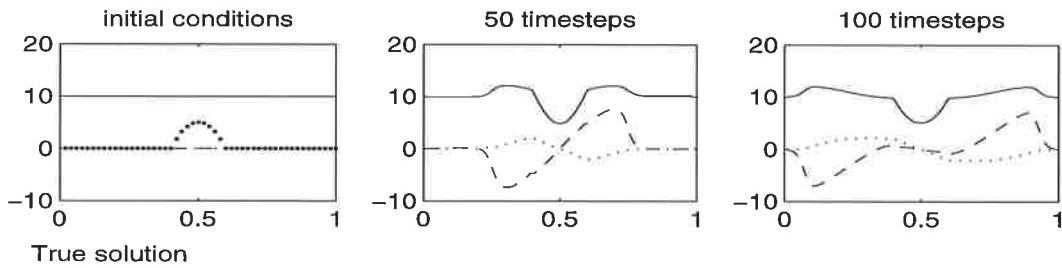
FIG. 1. *The "true" solution. Solid line: depth*gravity; Dashed line: u-field of velocity; Dotted line: v-field of velocity. The heavy dotted line in the first plot shows topography.*

different control variables with particular types of model error. This gives some insight into how an operational data assimilation scheme might cope with model error.

3.2   The experiments

The model was run with 100 grid points for 100 or 200 timesteps, and periodic boundary conditions were used. For the true run, the initial values of the u- and v-components of velocity were set to zero, and the initial water depth was $1m$ over the whole domain. The bottom topography was flat with a "mountain" of height $0.5m$ in the middle of the domain (see Figure 1).

The following situations were examined using the initial condition, the correction term, and then both together as control variables.

Experiment (a) Perfect model, wrong initial conditions. The initial water depth is taken to be $1.5m$ instead of $1m$. The initial u- and v-fields are unaltered.

Experiment (b) Omission of topography. In this case, the topography is unspecified, and so the initial water height is constant throughout the domain, and hence no motion is initiated in the subsequent model run.

Experiment (c) Omission of rotation. Without the Coriolis forcing, the v-component of velocity remaines zero throughout the model run.

4   Results

Figure 1 shows the "true" solution of the model over 100 timesteps, for the u- and v-components of velocity in ms^{-1}, and for ϕ, water depth times gravity in $m^2 s^{-2}$. Figures 2, 3 and 4 show the errors in these fields for experiments (a), (b) and (c) respectively, for a model run without assimilation, and for model runs using the initial condition and the correction term as control variables.

4.1   Wrong initial conditions

The results from experiment (a) (Figure 2) illustrate how using the initial condition as the control variable can exactly reconstruct the true initial condtions when there is no model error and when there are sufficient observations with no error. The results also show that using the correction term as the control variable to some extent compensates for the effect that errors in the initial conditions have on the subsequent model run. The largest improvement to the results in this case is in the middle of the assimilation period. When
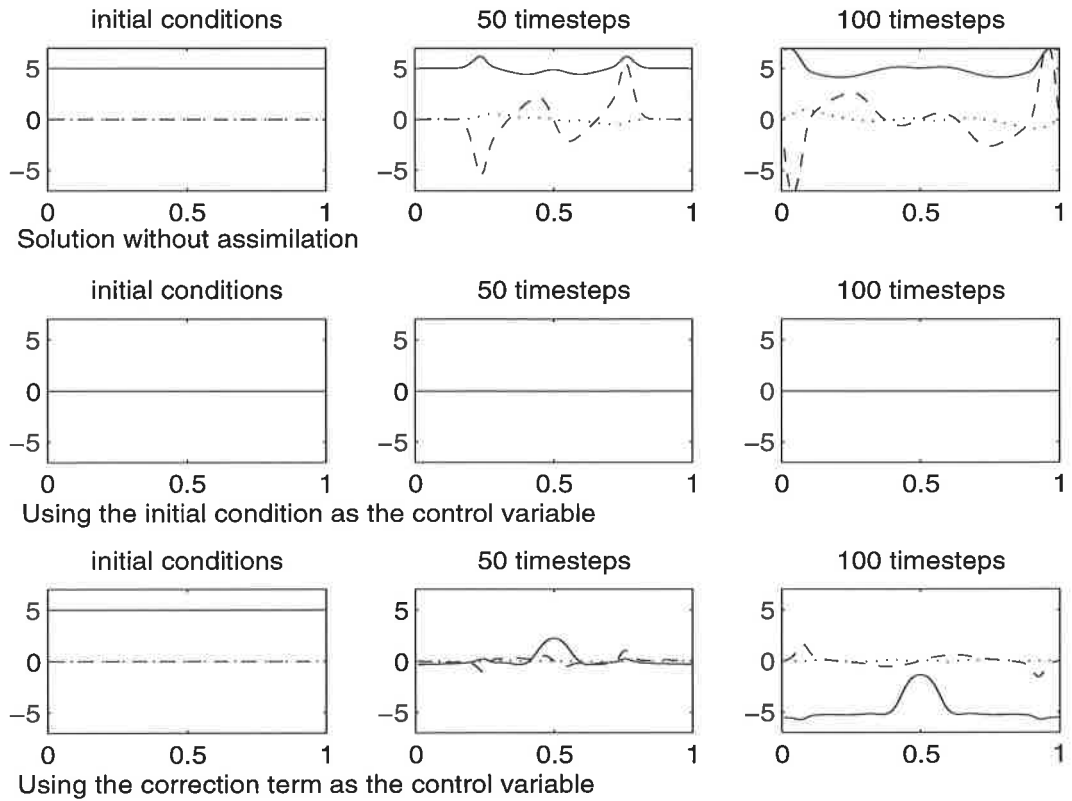
FIG. 2.   *Experiment (a). Error fields before assimilation, and after using the two different control variables. Solid line: depth*gravity; Dashed line: u-field of velocity; Dotted line: v-field of velocity.*

data assimilation is used operationally to produce improved initial conditions for the next forecast, accurate results at the end of the current assimilation period are more desirable. Additional weighting in the cost functional could be chosen to force the results to be closer to the true solution at this time. When both control variables were used together, the same results were achieved as when the initial condition only was used as the control variable, as desired. A much larger number of iterations of the minimization algorithm were needed, however, in this case.

4.2   Model error

The results from experiments (b) and (c) show that using the initial condition as the control variable reduces the errors at the middle and at the end of the assimilation period, and so compensates for the effects of model error at these times. Hence, the optimal initial condition is not equal to the initial condition of the true model solution, but compensates for the effects of model error to give a solution which is overall closer to the true solution.

In experiment (b), both the correction term and the initial condition correct very effectively for the errors in the velocity fields, but the correction term makes a better correction of the water depth (Figure 3), and so is the more effective control variable in this case. In experiment (c) both control variables produce a similar reduction in error in the latter part of the assimilation period (Figure 4).

When the assimilation period was extended to 200 timesteps, using the correction term
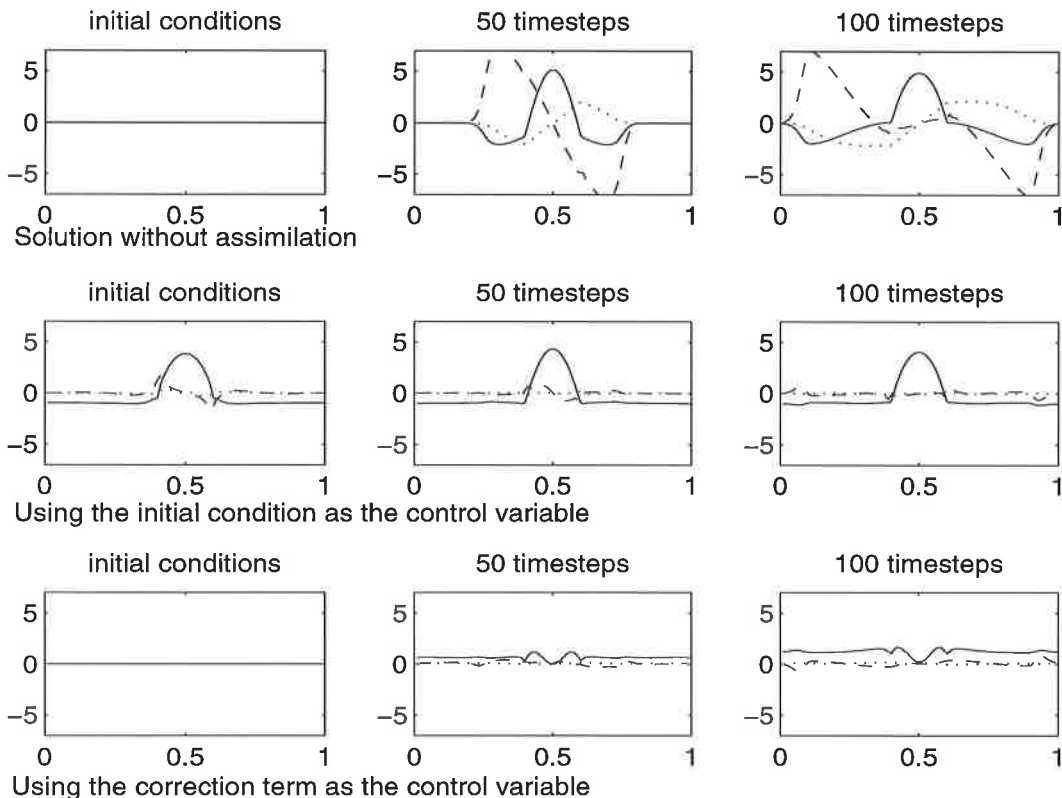
FIG. 3. *Experiment (b). Error fields before assimilation, and after using the two different control variables. Solid line: depth*gravity; Dashed line: u-field of velocity; Dotted line: v-field of velocity.*

as the control variable gave less improvement than in the 100 timestep run. In experiment (b) the results were still better than when the initial condition was used as the control variable, but in experiment (c) they were slightly worse. This illustrates a limitation of the correction term technique: the longer the assimilation period, the less effective a constant correction term is likely to be, since it represents an "average" correction over this time. However, the 100 timestep run was long enough for the effects of model error to propagate all the way across the domain, and so it is significant that the correction term compensates for model error on this time scale. It is also worth noting that the model error in experiments (b) and (c) involves the true solution in each case, and this is not constant in time.

Using both control variables together further reduced the errors slightly, but again, this took a large number of iterations of the minimization algorithm.

5   Conclusion

The techniques of variational data assimilation involving adjoint equations can easily be modified to use a correction term as the control variable instead of, or as well as, the initial condition. The correction term provides a way to compensate for model error and the results from this study show that in an idealized situation, using a constant correction term as the control variable over a significant though limited time period can successfully compensate for model error. Further work is being carried out to investigate other forms
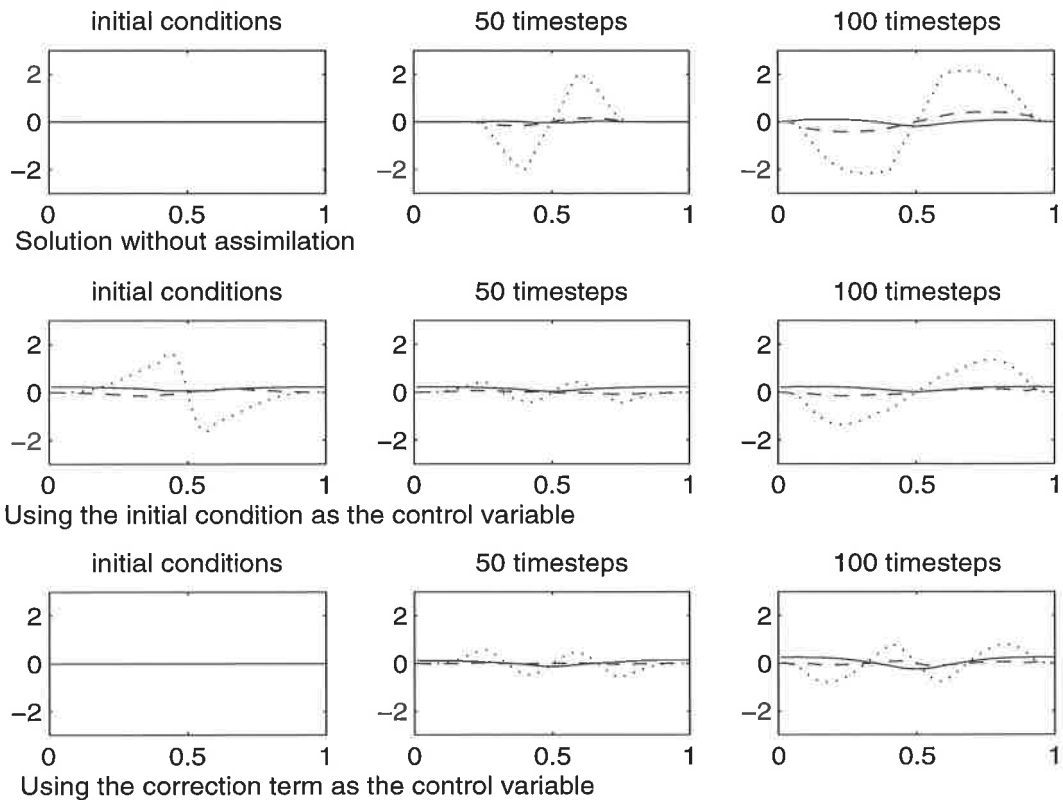
FIG. 4. *Experiment (c). Error fields before assimilation, and after using the two different control variables. Solid line: depth*gravity; Dashed line: u-field of velocity; Dotted line: v-field of velocity.*

of correction term to improve the technique. It is also noteworthy that using the initial condition as the control variable can also reduce the effect of model error in assimilation. Using both control variables together has theorectical advantages, but in this study required many more iterations of the minimization algorithm.

Acknowledgements

References
[1] W. C. Chao and L-P. Chang, *Development of a four-dimensional analysis system using the adjoint method at GLA. Part 1: dynamics*, Monthly Weather Review, 120 (1992), pp. 1661–1673.

[2] D. P. Dee, *On-line estimation of error covariance parameters for atmospheric data assimilation*, Monthly Weather Review, 123 (1995), pp. 1128–1145.

[3] J. C. Derber, *A variational continuous assimilation technique*, Monthly Weather Review, 117 (1989), pp. 2437–2446.

[4] J. Ch. Gilbert, and C. Lemaréchal, *Some numerical experiments with variable-storage quasi-Newton algorithms*, Mathematical Programming, 45 (1989), pp. 407–435.

[5] F-X. Le Dimet and Talagrand, *Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects*, Tellus, 38A (1986), pp. 97–110.

[6] A. C. Lorenc, *Optimal nonlinear objective analysis*, Q.J.R. Meteorol. Soc, 114 (1988), pp. 205–240.

[7] C. A. Parrett and M. J. P. Cullen, *Simulation of hydraulic jumps in the presence of rotation and mountains*, Q.J.R. Meteorol. Soc, 110 (1984), pp. 147–165.

[8] W. C. Thacker, *Relationships between statistical and deterministic methods of data assimilation*, in Variational Methods in the Geosciences, edited by Y. K. Sasaki, Elsevier, New York, 1986.