THE UNIVERSITY OF READING

Adjoint Methods for Treating Model Error in Data Assimilation

by

A.K. Griffith & N.K. Nichols

Numerical Analysis Report 5/98

DEPARTMENT OF MATHEMATICS

Adjoint Methods for Treating Model Error in Data Assimilation

by

A.K. Griffith & N.K. Nichols

Adjoint Methods for Treating Model Error in Data Assimilation

A. K. Griffith   N. K. Nichols
The University of Reading

1   Introduction

Mathematical models for simulating physical, biological and economic systems are now often more accurate than the data that is available to drive them. In particular, complete information describing the initial state of an evolutionary system is seldom known. In this case it is desirable to use the measured output data that is available from the system over an interval of time, in combination with the model equations, to derive accurate estimates of the expected system behaviour. The problem of constructing a state-estimator, or observer, is the dual of the feedback control design problem. For very large nonlinear systems arising in numerical weather prediction and in ocean circulation modelling, traditional control system design techniques are not practicable, and 'data assimilation' schemes are used instead to generate accurate state-estimates. The aim of these schemes is to incorporate observed data into computational simulations in order to improve the accuracy of the numerical forecasts.

Currently, variational data assimilation schemes are under development[7]. These schemes are attractive because they deliver the best statistically linear unbiased estimate of the model solution given the available observations and their error covariances. The problem is formulated as an optimal control problem where the cost functional measures the mismatch between the model predictions and the observed system states, weighted by the inverse of the covariance matrices. The model equations are treated as strong constraints and the controls to be determined are the initial states of the system. The constrained minimization problem is typically solved by a gradient iterative procedure for finding the optimal controls. The gradient directions needed in the iteration are obtained by solving the linear adjoint equations associated with the problem.

In practice the model equations do not represent the system behaviour exactly and model errors arise due to lack of resolution, to inaccurate physical parameters, or to errors in boundary conditions, in topography or in other forcing terms. To account for model error, the system equations can be treated as weak constraints in the optimization problem. The residual errors in the model equations at every time point are then treated as control parameters. Statistically the model error is assumed to be unbiased white noise which is uncorrelated in time. This approach is not practicable, however, due to the excessive size of the optimization problem and due to the need to propagate the covariance matrices of the model errors at each time step. Furthermore, the statistical assumptions made in this approach are not generally satisfied in practice, since the model errors are expected to be time-correlated.

Recently, the problem of accounting for model error in variational assimilation in a cost-effective way has begun to receive more attention [2], [6], [8], [1]. Studies on predictability in meteorological models have shown that the impact of model error on forecast error is indeed significant. The results given in [1] lead to the conclusion that the predictability limit of a forecast might be extended by two or three days if model error were eliminated. There is, however, a lack of quantitative information on model error in such forecast models.

Although the general form of the model error is not known, some simple assumptions

about the evolution of the error can be made. The control variables then reduce to the unknown initial values of the model error and the corresponding optimization problem can be solved efficiently. A major advantage of this approach is that the gradient directions with respect to the model errors can be obtained directly from the adjoint equations of the original problem at very little extra cost.

In the next section, a general representation of model error for use in data assimilation is introduced. The technique of state augmentation for estimating serially correlated components of model error is then described. The variational problem for the augmented state system is defined in Section 3 and the corresponding adjoint method is developed. Using simple models it is shown in Section 4 that a constant error, or bias error, can be used as a control to correct for model error in a source term. The extension of this approach to the treatment of time-correlated advection error is also demonstrated.

2 Model error and state augmentation

The system is modelled by a discrete nonlinear set of equations, given by

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \epsilon_k, \quad k = 0, \ldots, N-1, \tag{2.1}$$

where $\mathbf{x}_k, \epsilon_k \in \mathbb{R}^n$ are the model state and the model error at time t_k and $\mathbf{f}_k : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear function describing the evolution of the state from time t_k to time t_{k+1}. It is commonly assumed that the error ϵ_k is stochastic and that it is unbiased and serially uncorrelated with a known Gaussian probability distribution. A Kalman filter technique can then be used to solve the assimilation problem [5]. For large systems, such as weather and ocean systems, this method is generally too expensive for operational use due to the enormous cost of propagating the error covariance matrices.

Moreover, for such evolutionary systems the model error is likely to depend on the model state and hence to be correlated in time. We therefore introduce a more general form of the model error that includes both serially correlated and random elements. We write

$$\epsilon_{k+1} = T_k \mathbf{e}_k + \mathbf{q}_k, \tag{2.2}$$

where $\mathbf{q}_k \in \mathbb{R}^n$ are unbiased, serially uncorrelated, normally distributed random vectors and $\mathbf{e}_k \in \mathbb{R}^m$ represent serially correlated components of the model error. The matrices $T_k \in \mathbb{R}^{n \times m}$ are prescribed matrices, with rank$(T_k) = m$, that define the distribution of the serial error terms $\mathbf{e}_k$ in the model equations. The evolution of the serial error terms is assumed to satisfy the general equation

$$\mathbf{e}_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{e}_k), \tag{2.3}$$

where $\mathbf{g}_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is some function to be specified. In practice we know very little about the form of the model error and need to specify a simple form for the error evolution that reflects any available knowledge. Examples of simple forms of error evolution include:

- **Constant bias error :** $\mathbf{e}_k \equiv \mathbf{e} \quad \forall k$.
  This choice allows for a constant vector of unknown 'dynamical parameters' which, in the deterministic case (*i.e.* $\mathbf{q}_k = 0$), corresponds to the correction term of [2]. In the stochastic case, the constant correction $\mathbf{e}$ can be interpreted as a statistical bias in the model error, which needs to be estimated. This form is expected to be appropriate for representing average errors in source terms or in boundary conditions.

- **Evolving error with model evolution :** $e_{k+1} = F_k e_k$.
  Here $F_k \in \mathbb{R}^{m \times m}$ represents a simplified linear model of the state evolution. This choice is appropriate, for example, for representing discretization error in models that approximate continuous dynamical processes by discrete-time systems.

- **Spectral form of model error :** $e_k \equiv \left(I,\ \sin(\frac{k}{N\tau})I,\ \cos(\frac{k}{N\tau})I\right) e$.
  Here the constant vector e is partitioned into three component vectors $e = \left(e_1^T, e_2^T, e_3^T\right)^T$ and τ is a constant determined by the timescale on which the model error is expected to vary, for example, a diurnal timescale. This choice approximates the first order terms in a spectral expansion of the model error.

Other choices can be described using the general form (2.2)–(2.3), including piecewise constant error, linearly growing error, and combinations of any of these types of model error (see [3]).

Together the system equations and the model error equations (2.1)–(2.3) constitute an *augmented* state system model. The aim of the data assimilation problem for the augmented system is to estimate the expected values of the augmented states x_k and e_k for $k = 0, \ldots, N-1$, that fit the observations. The solution delivers the maximum likelihood estimate of the augmented system states, given the error covariances of both the observations and the model errors. Although this formulation takes into account the time evolution of the model errors, the data assimilation problem remains intractable for operational use. If the stochastic elements of the error are ignored and the augmented system is treated as a deterministic model, then the size of the problem is greatly reduced. The aim of the data assimilation, in this case, is to estimate the serially correlated components of the model error along with the dynamical states of the original system model. In the next section the data assimilation problem for the augmented deterministic problem is described and the adjoint method for solving the problem is discussed.

3   Data assimilation problem

The augmented state system for the model states and model errors is written

$$x_{k+1} = f_k(x_k) + T_k e_k, \tag{3.1a}$$
$$e_{k+1} = g_k(x_k, e_k), \tag{3.1b}$$

for $k = 0, \ldots, N-1$. The observations are related to the system states by the equations

$$y_k = h_k(x_k) + \delta_k, \quad k = 0, \ldots, N-1, \tag{3.2}$$

where $y_k \in \mathbb{R}^{p_k}$ is a vector of p_k observations at time t_k and $h_k : \mathbb{R}^n \to \mathbb{R}^{p_k}$ is a nonlinear function that includes transformations and grid interpolations. The observational errors $\delta_k \in \mathbb{R}^{p_k}$ are assumed to be unbiased, serially uncorrelated, Gaussian random vectors with covariance matrices $R_k \in \mathbb{R}^{p_k \times p_k}$. It is also assumed that prior estimates, or 'background estimates,' x_0^b and e_0^b of x_0 and e_0 are known and that the covariance matrices of the errors $(x_0 - x_0^b)$ and $(e_0 - e_0^b)$ are given, respectively, by $B_0 \in \mathbb{R}^n$ and $Q_0 \in \mathbb{R}^m$. The observational errors and the errors in the prior estimates are not correlated.

The aim of the data assimilation is to minimize the least square errors between the model predictions and the observed system states, weighted by the inverse of the covariance matrices, over the assimilation interval. The control variables are the initial values x_0 and e_0 of the model state and model error, which completely determine the response of the augmented

system (3.1). For the problem to be well-posed, in general, the mean square error between the prior estimate and the control variables must be included in the objective function. The data assimilation problem is then given by

Problem 1 *Minimize, with respect to* $\mathbf{x}_0$ *and* $\mathbf{e}_0$, *the objective function*

$$\mathcal{J} = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{j=0}^{N-1}(\mathbf{h}_j(\mathbf{x}_j) - \mathbf{y}_j)^T R_j^{-1}(\mathbf{h}_j(\mathbf{x}_j) - \mathbf{y}_j)$$

$$+\frac{1}{2}(\mathbf{e}_0 - \mathbf{e}_0^b)^T Q_0^{-1}(\mathbf{e}_0 - \mathbf{e}_0^b), \tag{3.3}$$

subject to the augmented system equations (3.1).

The constrained minimization problem can be converted into an unconstrained problem using the method of Lagrange multipliers. Necessary conditions for a solution to Problem 1 require that the system equations together with a set of adjoint equations be satisfied. The adjoints can be written

$$\boldsymbol{\lambda}_k = F_k^T(\mathbf{x}_k)\boldsymbol{\lambda}_{k+1} + G_k^T(\mathbf{x}_k, \mathbf{e}_k)\boldsymbol{\mu}_{k+1} - H_k^T(\mathbf{x}_k)R_k^{-1}(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k), \tag{3.4a}$$

$$\boldsymbol{\mu}_k = T_k^T\boldsymbol{\lambda}_{k+1} + \Gamma_k^T(\mathbf{x}_k, \mathbf{e}_k)\boldsymbol{\mu}_{k+1}, \tag{3.4b}$$

for $k = N - 1, \ldots, 0$, and

$$\boldsymbol{\lambda}_N = 0, \qquad \boldsymbol{\mu}_N = 0, \tag{3.4c}$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$, $\boldsymbol{\mu}_k \in \mathbb{R}^m$ are the adjoint variables and $F_k \in \mathbb{R}^{n \times n}$, $H_k \in \mathbb{R}^{n \times p_k}$ and $G_k \in \mathbb{R}^{m \times n}$ are the Jacobians of $\mathbf{f}_k$, $\mathbf{h}_k$ and $\mathbf{g}_k$ with respect to $\mathbf{x}_k$, respectively, and $\Gamma_k \in \mathbb{R}^{m \times m}$ is the Jacobian of $\mathbf{g}_k$ with respect to $\mathbf{e}_k$.

The gradients of the objective function (3.3) with respect to the initial data $\mathbf{x}_0$ and $\mathbf{e}_0$ are then given by

$$\nabla_{\mathbf{x}_0}\mathcal{J} = B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - \boldsymbol{\lambda}_0, \tag{3.5a}$$

$$\nabla_{\mathbf{e}_0}\mathcal{J} = Q_0^{-1}(\mathbf{e}_0 - \mathbf{e}_0^b) - \boldsymbol{\mu}_0. \tag{3.5b}$$

For the optimal it is required that the gradients (3.5) be equal to zero. Otherwise these gradients provide the local descent direction needed to find an improved estimate for the optimal initial values of the augmented system using a gradient minimization technique.

In the special case where the model error is assumed to be constant, the adjoint equations can be simplified. In this case $G_k = 0$ and only the values for the adjoint variables $\boldsymbol{\lambda}_k$ need to be calculated. The gradient of the objective function is then given simply by

$$\nabla_{\mathbf{e}_0}\mathcal{J} = Q_0^{-1}(\mathbf{e}_0 - \mathbf{e}_0^b) - \sum_{j=1}^{N-1}T_{j-1}^T\boldsymbol{\lambda}_j. \tag{3.6}$$

Hence there is little extra computational effort needed to compute the gradients of the objective function in the case where the controls consist of the initial data for the model state and the model error.

4   Applications

The performance of data assimilation with the augmented system is examined for two cases using the initial state, the model error and both together as control vectors. In the first case

a constant bias error correction is applied and in the second case an evolving model error correction is developed. In both cases the minimization problem is solved using the conjugate gradient method. The convergence criterion for the iteration is given by $\|\nabla_{\mathbf{u}} \mathcal{J}\| \leq 10^{-6}$, where $\mathbf{u}$ denotes the control variables. (Here $\|\cdot\|$ denotes the L_2 - norm.)

The results are presented in Figs 1–4. In all figures a solid line indicates the solution to the 'true' system, from which the observations are taken; the observations are error-free and are denoted by $+$; a dotted line shows the unassimilated solution to the 'imperfect' model equations; and a dashed line represents the analysed solution to the data assimilation problem. The assimilation is applied on the interval $[0, 0.5]$ and a forecast is produced on the interval $[0.5, 1]$, starting from the assimilated solution at time $t = 0.5$. The covariance matrices of the prior estimates and the observations are taken, respectively, to be $B_0 = 0$, $Q_0 = qI$ and $R_k = \frac{2}{N}I$, $\forall k$.

4.1 Example 1

In the first case the system is derived from a standard explicit finite difference approximation to the heat equation

$$v_t = \sigma v_{zz} + s(z), \tag{4.1}$$

with zero boundary conditions at $z = 0, 1$ and a point source $s(z) = (1/3)\delta(z - 0.25)$, where δ denotes the Dirac delta function. The model equations are given by

$$x_j^{k+1} - x_j^k = \sigma \Delta t \left(x_{j-1}^k - 2x_j^k + x_{j+1}^k \right) / \Delta z^2 + s_j \Delta t, \tag{4.2a}$$

$$x_0^k = 0, \qquad x_J^k = 0, \tag{4.2b}$$

for $j = 0, 1, .., J$, $k = 0, 1, .., N$, where the model variables x_j^k approximate $v(j\Delta z, k\Delta t)$ with $\Delta t = (1/N)$, $\Delta z = (1/J)$. The discretized source term is given by $s_{J/4} = 1/(3\Delta z)$ and $s_j = 0$, $\forall j \neq J/4$.

The 'true' states, from which the observations are taken, are the solutions to the discrete equations (4.2) with initial values $x_j^0 = 1$, where $\Delta t = (1/80)$, $\Delta z = (1/16)$ and $\sigma = 0.1$. The positions of the observations, shown in Figs 1–2, do not coincide with the finite difference grid and the function $\mathbf{h}_k(\mathbf{x}_k) \equiv C\mathbf{x}_k$, where $C \in \mathbb{R}^{p \times n}$, defines a fixed linear interpolation between the model grid and the observation positions. In the model equations, the source term is omitted, making the model 'imperfect.' It is assumed, however, that the prior estimate of the initial values is exact. The aim is to estimate the state of the 'true' system using the observations and the 'imperfect' model.

Fig. 1 shows the assimilated solution obtained by using the the initial state alone as the control variable. At the initial point the assimilation does not produce the 'true' initial state, but instead generates initial values that compensate for the model errors and ensure that the assimilated solution is as close as possible to the observations over the whole interval. The estimated state at the end of the assimilation interval ($t = 0.5$) is therefore closer to the true state than the background (unassimilated) solution. The forecast from this position is still poor, however, due to the inaccuracy of the model.

Fig. 2 shows the results of the assimilation using the augmented system where the model error is assumed to be a constant bias error and $q = 0$. In this case the assimilated solution exactly matches the true solution on the assimilation interval. (Theoretically this is expected since the system is completely observable and the model error is constant in time.) Retaining the computed model error correction over the forecast interval then gives a perfect forecast. Equally good results are obtained if the correction terms are confined to a region around the

Figure 1: Example 1. Variational assimilation using the initial data as the control vector.
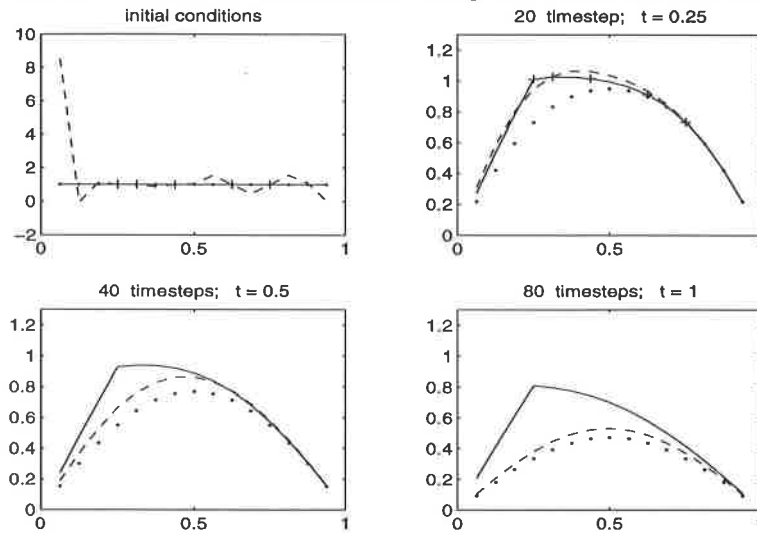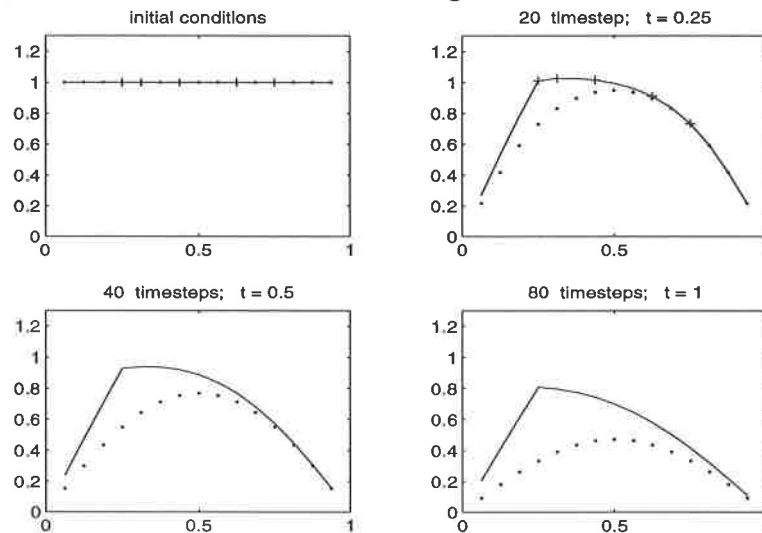

Figure 2: Example 1. Variational assimilation using the constant error as the control vector.

source term. The dimension of the model error vector can thus be reduced and the efficiency improved, if the location of the source is known.

Additional results are presented in [3], including examples where the prior estimate of the initial data is incorrect and where the initial state and the constant bias error are both used together as the control.

4.2 Example 2

In the second case the system is obtained from an upwind approximation to the linear advection equation

$$v_t + v_z = 0,$$ (4.3)

Figure 3: Example 2. Variational assimilation using the initial data as the control vector.
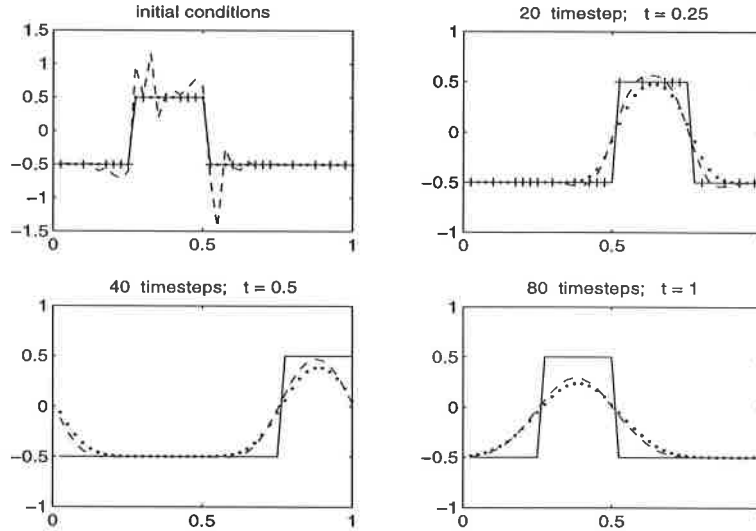
with periodic boundary conditions on the interval $z \in [0, 1]$. Initially the solution is a square wave defined by

$$v(z, 0) = \alpha(z) = \begin{cases} 0.5 & 0.25 < z < 0.5, \\ -0.5 & z < 0.25 \ \ \text{or} \ \ z > 0.5. \end{cases} \tag{4.4}$$

Over the time interval $[0, 1]$ this square wave is advected all the way around the model domain and back to its starting position.

The model equations are defined by

$$x_j^{k+1} - x_j^k = -\frac{\Delta t}{\Delta z}(x_j^k - x_{j-1}^k), \tag{4.5a}$$

$$x_j^0 = \alpha(j\Delta z), \qquad x_0^k = x_J^k, \tag{4.5b}$$

for $j = 1, .., J$, $k = 0, 1, .., N$, with model variables $x_j^k \approx v(j\Delta z, k\Delta t)$ and $\Delta z = (1/J)$, $\Delta t = (1/N)$.

The 'true' states in this case are the exact solutions to the continuous advection problem with the given initial conditions. (These are generated as solutions to the model equations with $\Delta t = \Delta z = 1/80$.) The observations are taken from the 'true' states at 20 grid points on the assimilation interval. The positions of the observations are shown in Figs 4–5. The model states are generated from the exact initial states using $\Delta t = 1/80$ and $\Delta z = 1/40$. With this choice of stepsizes, the discretization introduces model error and the upwind scheme exhibits numerical dissipation, which smears the shock fronts.

The aim of the data assimilation is to reconstruct the 'true' states of the system, and in particular the steep shock fronts, using the observations and the 'imperfect' model. In this case taking the model error to be a constant bias error does not give any improvement in the solution, since the average error over the time interval introduced by the discretization is zero. The model error now depends on the true system state and hence the evolving model error correction is used here. The error is assumed to satisfy the same linear dynamical equations as the model states.

In Fig. 3 the results of the assimilation are shown for the case where the initial state alone is used as the control variable and the error is not modelled. As noted previously, at the initial point the assimilation does not reproduce the correct initial data, but generates

Figure 4: Example 2. Variational assimilation using the evolving error as the control vector.
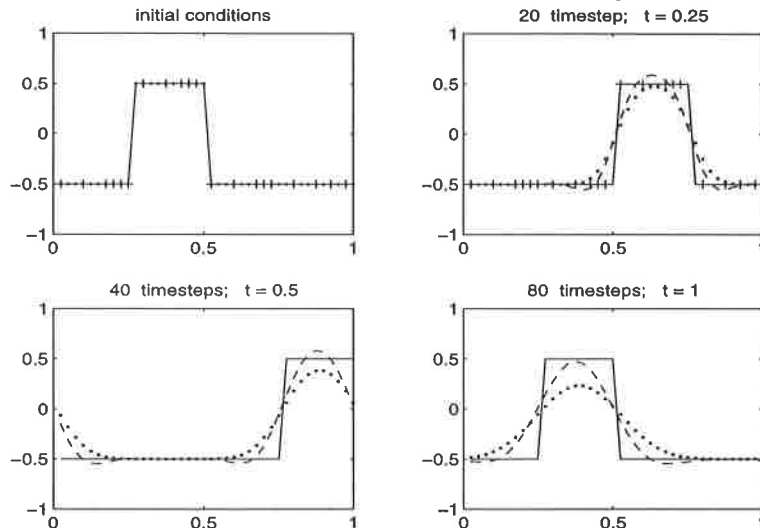
an initial solution that compensates for the impact of the model error over the assimilation interval. At the end of the interval, the assimilated solution is closer to the true solution, estimating the amplitude slightly more accurately than the 'background' model solution, but the forecast remains poor.

In Fig. 4 the assimilated solution found using the evolving model error correction with $q = 10$ is shown. A much better approximation to the true state of the system is obtained than in the case where the initial state is used as the control vector. Evolving the model error along with the model state over the forecast interval then gives a considerably improved prediction of the true state of the system.

The results of further tests on this example are given in [3].

4.3 Nonlinear example

The techniques described here for treating model error in data assimilation have also been tested on a one dimensional nonlinear shallow water model [3],[4]. In addition to error in the initial states of the system, various types of model error have been investigated, including error in the topography and error in rotation (incorrect Coriolis parameter). Assimilation with noisy data has also been examined. In these cases, the constant bias error corrrection gives good estimates of the true solution over the assimilation interval. By retaining the model error correction over the forecast interval, the forecast is improved significantly. Assimilation using both the initial state and the model error as control variables to correct simultaneously for initial and model errors has also been successful.

5 Conclusions

A technique for treating model error in data assimilation is described here. The aim of the technique is to estimate the serially correlated components of the model error along with the dynamical model states. A simple form for the evolution of the model error is assumed and an augmented system for both the model state and model error is obtained. For different types of error, it is found that different forms for the model error evolution are appropriate. The initial states of the augmented system are used as control variables in the assimilation process. A

modified objective function is minimized to determine the solution of the augmented system that best fits the available observations over the assimilation interval. It is shown that this technique is effective and leads to significantly improved forecasts.

Acknowledgements

References

[1] A. Dalcher and E. Kalnay, Error growth and predictability in operational ECMWF forecasts, *Tellus*, 39A:474–491, 1987.

[2] J.C. Derber, A variational continuous assimilation technique, *Monthly Weather Review*, 117:2437–2446, 1989.

[3] A.K. Griffith. *Data Assimilation for Numerical Weather Prediction Using Control Theory*. The University of Reading, Department of Mathematics, PhD Thesis, 1997.

[4] A.K. Griffith and N.K. Nichols, Accounting for model error in data assimilation using adjoint methods, *Computational Differentiation: Techniques, Applications and Tools*, (Eds. M. Berz, C. Bischof, G. Corliss and A. Greiwank), SIAM, Philadelphia, 195-204, 1996.

[5] R.E. Kalman, A new approach to linear filtering and prediction problems, *Transactions of the ASME, Series D*, 83:35–44, 1961.

[6] R. Menard and R. Daley, The application of Kalman smoother theory to the estimation of 4DVAR error statistics, *Tellus*, 48A:221–237, 1996.

[7] O. Talagrand and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation I: Theory, *Quart.J.Roy.Met.Soc.*, 113:1311–1328, 1997.

[8] J.J. Tribbia and D.P. Baumhefner, The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies, *Monthly Weather Review*, 108:2276–2288, 1988.