THE UNIVERSITY OF READING

Adjoint Methods in Data Assimilation for Estimating Model Error

by

A.K. Griffiths and N.K. Nichols

Numerical Analysis Report 9/99

DEPARTMENT OF MATHEMATICS

THE UNIVERSITY OF READING
Department of Mathematics

Adjoint Methods in Data Assimilation for Estimating Model Error

by

A.K. Griffiths and N.K. Nichols

Numerical Analysis Report 9/99

Invited paper for the ERCOFTAC Workshop on Adjoint Methods, Institute de Mechaniques de Fluides de Toulouse, Toulouse, France, June, 1999

Adjoint Methods in Data Assimilation for Estimating Model Error

A. K. Griffith N. K. Nichols
Department of Mathematics,
The University of Reading

Abstract

Data assimilation aims to incorporate measured observations into a dynamical system model in order to produce accurate estimates of all the current (and future) state variables of the system. The optimal estimates minimize a variational principle and can be found using adjoint methods. The model equations are treated as strong constraints on the problem. In reality, the model does not represent the system behaviour exactly and errors arise due to lack of resolution and inaccuracies in physical parameters, boundary conditions and forcing terms. A technique for estimating systematic and time-correlated errors as part of the variational assimilation procedure is described here. The modified method determines a correction term that compensates for model error and leads to improved predictions of the system states. The technique is illustrated in two test cases. Applications to the 1-D nonlinear shallow water equations demonstrate the effectiveness of the new procedure.

Keywords Data assimilation, adjoint methods, model error, bias estimation, nonlinear shallow-water equations.

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 [12]. 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 objective 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 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], [10], [14], [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.

A new technique for treating model errors is presented here. The aim of the technique is to estimate the systematic, time-correlated components of the model error along with the dynamical model states as part of the variational assimilation procedure. Although the general form of the model error is not known, some simple assumptions about the evolution of the error can be made. An augmented system for both the model states and model errors is thus derived. The control variables are reduced to the unknown initial values of the model states and model errors 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 from the adjoint equations of the original problem at very little extra cost. Preliminary results using this technique have been presented at conferences and workshops [5], [6], [7]. A comprehensive development of the procedure is given here, together with new applications and results.

In the next section, the variational data assimilation procedure is introduced. A general representation of model error for use in data assimilation is defined in Section 3 and the technique of state augmentation for estimating serially correlated components of model error is described. The variational problem for the augmented state system is derived in Section 4 and the corresponding adjoint method is developed. In Section 5, a simple diffusion model is used to show that a constant error, or bias error, can be taken as the control in order to correct for model error in a source term. The extension of this approach to the treatment of time-correlated advection error by an evolving model error is demonstrated with another test example. In Section 6, the augmented assimilation procedure is applied to estimate model errors in a discretized form of the 1-D nonlinear shallow water equations. Concluding remarks are presented in Section 7.

2 Variational data assimilation

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

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

where $\mathbf{x}_k$ is the model state 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}. The observations are related to the system states by the equations

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\delta}_k, \qquad k = 0, \ldots, N-1, \tag{2.2}$$

where $\mathbf{y}_k \in \mathbb{R}^{p_k}$ is a vector of p_k observations at time t_k and $\mathbf{h}_k : \mathbb{R}^n \to \mathbb{R}^{p_k}$ is a nonlinear function that includes transformations and grid interpolations. The observational errors $\boldsymbol{\delta}_k \in \mathbb{R}^{p_k}$ are assumed to be unbiased, serially uncorrelated, Gaussian random vectors with covariance matrices $R_k \in \mathbb{R}^{p_k \times p_k}$. A prior estimate, or 'background estimate,' $\mathbf{x}_0^b$ of the initial state $\mathbf{x}_0$ is assumed to be known and the initial random error $(\mathbf{x}_0 - \mathbf{x}_0^b)$ is assumed to be Gaussian with covariance matrix $B_0 \in \mathbb{R}^{n \times n}$. The observational errors and the errors in the prior estimates are assumed to be uncorrelated.

The aim of the data assimilation is to find the maximum likelihood Bayesian estimate of the system states given the observations and the prior estimate of the initial state. This problem reduces to minimizing the square error between the model predictions and the observed system states, weighted by the inverse of the covariance matrices, over the assimilation interval. The model is assumed to be 'perfect' and the system equations are treated as strong constraints on the objective function. The model states that satisfy the system equations are uniquely determined on the assimilation interval by the initial states of the system. The initial states can thus be treated as the required control variables in the optimization. The data assimilation problem is defined explicitly as follows.

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

$$\mathcal{J} = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{j=0}^{N-1}(\mathbf{h}_j(\mathbf{x}_j) - \mathbf{y}_j)^T R_j^{-1}(\mathbf{h}_j(\mathbf{x}_j) - \mathbf{y}_j) \tag{2.3}$$

subject to the system equations (2.1).

In practice the constrained minimization problem is solved iteratively by a gradient method. The problem is first reduced to an unconstrained problem using the method of Lagrange. Necessary conditions for the solution to the unconstrained problem then require that a set of adjoint equations together with the system equations must be satisfied. The adjoint equations are given by

$$\boldsymbol{\lambda}_N = 0, \tag{2.4a}$$

$$\boldsymbol{\lambda}_k = F_k^T(\mathbf{x}_k)\boldsymbol{\lambda}_{k+1} - H_k^T R_k^{-1}(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k), \qquad k = N-1, \ldots, 0, \tag{2.4b}$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$, $j = 0, \ldots, N$, are the adjoint variables and $F_k \in \mathbb{R}^{n \times n}$ and $H_k \in \mathbb{R}^{n \times p_k}$ are the Jacobians of $\mathbf{f_k}$ and $\mathbf{h_k}$ with respect to $\mathbf{x}_k$.

The gradient of the objective function (2.3) with respect to the initial data $\mathbf{x}_0$ is then given by

$$\nabla_{\mathbf{x}_0}\mathcal{J} = B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - \boldsymbol{\lambda}_0, \tag{2.5}$$

At the optimal, the gradient (2.5) is required to be equal to zero. Otherwise this gradient provides the local descent direction needed in the iteration procedure to find an improved estimate for the optimal initial states. Each step of the gradient iteration process requires one forward solution of the model equations, starting from the current best estimate of the initial states, and one backward solution of the adjoint equations. The estimated initial conditions are then updated using the computed gradient direction. This process is expensive, but it is operationally feasible, even for very large systems, such as weather and ocean systems, which may involve as many as 10^7 state variables.

In reality, the system models are not 'perfect' and the model equations do not represent the system behaviour exactly. Both systematic and random errors affect the states of the system. Recent studies on predictability in meteorological models have shown that the impact of model error on forecast error is significant [1]. In the next sections, a general form for the error that includes both serially correlated and random components is proposed and an augmented system model is introduced that enables estimates of the model errors to be found. The aim of the assimilation is then to estimate both the model states and the systematic model errors using the observed data.

3 Model error and state augmentation

In order to take errors into account, the system is now modelled by the discrete nonlinear equations

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \boldsymbol{\epsilon}_k, \qquad k = 0, \ldots, N-1, \tag{3.1}$$

where $\boldsymbol{\epsilon}_k \in \mathbb{R}^n$ denotes the model error at time t_k. The observations $\mathbf{y}_k$ are related to the system by the equations (2.2) and a prior 'background' estimate $\mathbf{x}_0^b$ for the initial states, defined as in Section 2, is known.

Commonly the model errors $\boldsymbol{\epsilon}_k$ are assumed to be stochastic variables that are unbiased and serially uncorrelated, with a known Gaussian distribution. The data assimilation problem then reduces to minimizing the square error in the model equations, together with the square error between the model predictions and the observed system states, all weighted by the inverses of the covariance matrices, over the assimilation interval. The model equations (3.1) are thus treated as weak constraints in the objective function. As well as the initial states of the system, the model errors at every time point are the control parameters that must be determined.

An extended Kalman filter technique can now be used to solve the assimilation problem [8]. Alternatively, the problem can be solved by a gradient iterative procedure where the descent directions are determined from the associated adjoint equations. In this case, the converged estimate of the system state at the end of the assimilation period is equivalent to that obtained using the Kalman filter [13]. For large systems, such as weather and ocean systems, these methods are generally too expensive for operational use due to the enormous cost of propagating the error covariance matrices in the Kalman filter or, alternatively, estimating all of the model errors.

In any case, for evolutionary systems, the model error is expected to depend on the model state and hence to be *correlated in time*. Thus the statistical assumptions needed in this formulation of the assimilation problem are not generally satisfied. A more general form of the model error that includes both serially correlated and random elements is, therefore, now introduced here.

It is assumed that the evolution of the model error can be described by the equation

$$\boldsymbol{\epsilon}_{k+1} = T_k \mathbf{e}_k + \mathbf{q}_k, \tag{3.2}$$

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

$$\mathbf{e}_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{e}_k), \tag{3.3}$$

where $\mathbf{g}_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is some function to be specified.

In practice very little is known about the form of the model error and a simple form for the error evolution that reflects any available knowledge needs to be specified. Examples of simple forms of error evolution include:

- **Constant bias error :** $\mathbf{e}_{k+1} = \mathbf{e}_k, \quad T_k = I.$
This choice allows for a constant vector $\mathbf{e} \equiv \mathbf{e}_0$ of unknown 'dynamical parameters' to be found. In the deterministic case (*i.e.* $\mathbf{q}_k = 0$), the constant error $\mathbf{e}$ 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 :** $\mathbf{e}_{k+1} = F_k\mathbf{e}_k, \quad T_k = I.$
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 :** $\mathbf{e}_{k+1} = \mathbf{e}_k, \quad T_k = \left(I, \sin(\frac{k}{N\tau})I, \cos(\frac{k}{N\tau})I\right).$
In this case the constant vector $\mathbf{e} \equiv \mathbf{e}_0$ is partitioned into three component vectors, $\mathbf{e} = \left(\mathbf{e}_1^T, \mathbf{e}_2^T, \mathbf{e}_3^T\right)^T$, 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 (3.2)–(3.3), including piecewise constant error, linearly growing error, and combinations of any of these types of model error (see [4]).

Together the system equations and the model error equations (3.1)–(3.3) 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 $\mathbf{x}_k$ and $\mathbf{e}_k$ 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, however, and the augmented system is treated as a 'perfect' 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 'perfect' augmented problem is described and the adjoint method for solving the problem is discussed.

4 Augmented data assimilation problem

The augmented system equations for the model states and model errors are now written

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + T_k\mathbf{e}_k, \tag{4.1a}$$

$$\mathbf{e}_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{e}_k), \tag{4.1b}$$

for $k = 0, \ldots, N-1$. As in previous sections, the observations are related to the model states by the equations

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\delta}_k, \qquad k = 0, \ldots, N-1. \tag{4.2}$$

The covariance matrices R_k of the observational errors are assumed to be known. It is also assumed that prior estimates, or 'background estimates,' $\mathbf{x}_0^b$ and $\mathbf{e}_0^b$ of $\mathbf{x}_0$ and $\mathbf{e}_0$, respectively, are known and that the covariance matrices of the errors $(\mathbf{x}_0 - \mathbf{x}_0^b)$ and $(\mathbf{e}_0 - \mathbf{e}_0^b)$ are given 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 square errors between the model predictions and the observed system states, weighted by the inverse of the covariance matrices, over the assimilation interval. The augmented system equations (4.1) are treated as strong constraints on the problem. The initial values $\mathbf{x}_0$ and $\mathbf{e}_0$ of the model state and model error completely determine the response of the augmented system and are taken, therefore, to be the control variables in the optimization. The problem is well-posed, in general, if the square errors between the prior estimates and the control variables are included in the objective function. The data assimilation problem is now given by

Problem 2 *Minimize, with respect to* $\mathbf{x}_0$ *and* $\mathbf{e}_0$, *the objective function*

$$\mathcal{J} = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{j=0}^{N-1}(\mathbf{h}_j(\mathbf{x}_j) - \mathbf{y}_j)^T R_j^{-1}(\mathbf{h}_j(\mathbf{x}_j) - \mathbf{y}_j)$$
$$+\frac{1}{2}(\mathbf{e}_0 - \mathbf{e}_0^b)^T Q_0^{-1}(\mathbf{e}_0 - \mathbf{e}_0^b), \tag{4.3}$$

subject to the augmented system equations (4.1).

The constrained minimization problem can again be converted into an unconstrained problem using the method of Lagrange. Necessary conditions for a solution to Problem 2 require that the system equations together with a set of adjoint equations be satisfied. The adjoint equations are given by

$$\boldsymbol{\lambda}_N = 0, \qquad \boldsymbol{\mu}_N = 0, \tag{4.4a}$$

and

$$\boldsymbol{\lambda}_k = F_k^T(\mathbf{x}_k)\boldsymbol{\lambda}_{k+1} + G_k^T(\mathbf{x}_k, \mathbf{e}_k)\boldsymbol{\mu}_{k+1} - H_k^T R_k^{-1}(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k), \tag{4.4b}$$
$$\boldsymbol{\mu}_k = T_k^T\boldsymbol{\lambda}_{k+1} + \Gamma_k^T(\mathbf{x}_k, \mathbf{e}_k)\boldsymbol{\mu}_{k+1}, \tag{4.4c}$$

for $k = N-1, \ldots, 0$, 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 (2.3) with respect to the initial data $\mathbf{x}_0$ and $\mathbf{e}_0$ are then given by

$$\nabla_{\mathbf{x}_0}\mathcal{J} = B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - \boldsymbol{\lambda}_0, \tag{4.5a}$$
$$\nabla_{\mathbf{e}_0}\mathcal{J} = Q_0^{-1}(\mathbf{e}_0 - \mathbf{e}_0^b) - \boldsymbol{\mu}_0. \tag{4.5b}$$

For the optimal it is required that the gradients (4.5) be equal to zero. Otherwise these gradients provide the local descent direction needed by the iteration procedure to find an improved estimate for the optimal initial values of the augmented system. In each step of the gradient iteration the augmented equations are solved in the forward direction, starting from the current best estimate of the initial conditions, and the corresponding adjoint equations are solved in the reverse direction. The estimated initial values are then updated using the computed gradients.

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 with respect to $\mathbf{e}_0$ is then given simply by

$$\nabla_{\mathbf{e}_0}\mathcal{J} = Q_0^{-1}(\mathbf{e}_0 - \mathbf{e}_0^b) - \sum_{j=1}^{N-1} T_{j-1}^T \boldsymbol{\lambda}_j. \tag{4.6}$$

In general, little extra computational effort is needed to compute the gradients of the objective function for the augmented system, since the controls consist only of the initial data for both the model states and the model errors.

5 Test examples

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 $\|\cdot\|_2$ 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 $R_k = \frac{2}{N}I$, $\forall k$.

5.1 Example 1

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

$$v_t = \sigma v_{zz} + s(z), \tag{5.1}$$

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

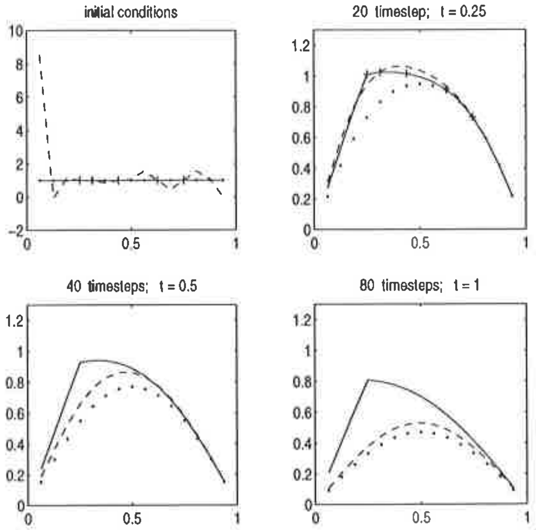
$$x_j^{k+1} - x_j^k = \sigma\Delta t\left(x_{j-1}^k - 2x_j^k + x_{j+1}^k\right)/\Delta z^2 + s_j\Delta t, \tag{5.2a}$$

$$x_0^k = 0, \qquad x_J^k = 0, \tag{5.2b}$$

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 (5.2) 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

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



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 the standard procedure described in Section 2, which uses the initial state alone as the control variable. At the initial point the assimilation does not reproduce 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, as described in Section 4, 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 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 [4], 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 controls.

5.2 Example 2

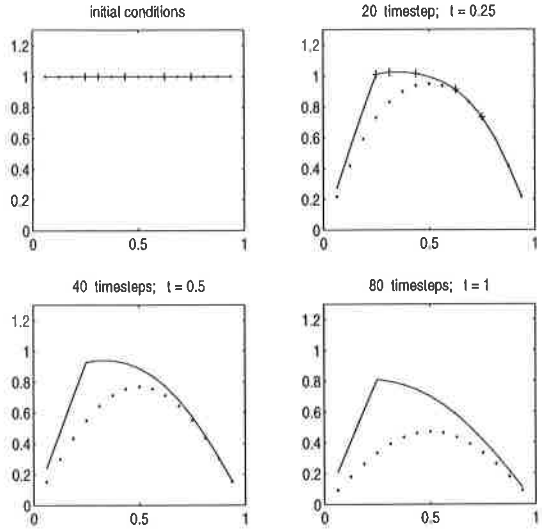
In the second case the system is obtained from an upwind approximation to the linear advection equation

$$v_t + v_z = 0, \tag{5.3}$$

with periodic boundary conditions on the interval $z \in [0, 1]$. Initially the solution is a square wave defined by

$$v(z, 0) = \alpha(z) = \begin{cases} 0.5, & 0.25 < z < 0.5, \\ -0.5, & z < 0.25 \ \text{ or } \ z > 0.5. \end{cases} \tag{5.4}$$

Figure 2: Example 1: Variational assimilation using the constant bias error as the control vector, with $q = 0$.



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

$$x_j^{k+1} - x_j^k = -\frac{\Delta t}{\Delta z}(x_j^k - x_{j-1}^k), \tag{5.5a}$$

$$x_j^0 = \alpha(j\Delta z), \qquad x_0^k = x_J^k, \tag{5.5b}$$

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 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.

Figure 3: Example 2: Variational assimilation using the initial data as the control vector.

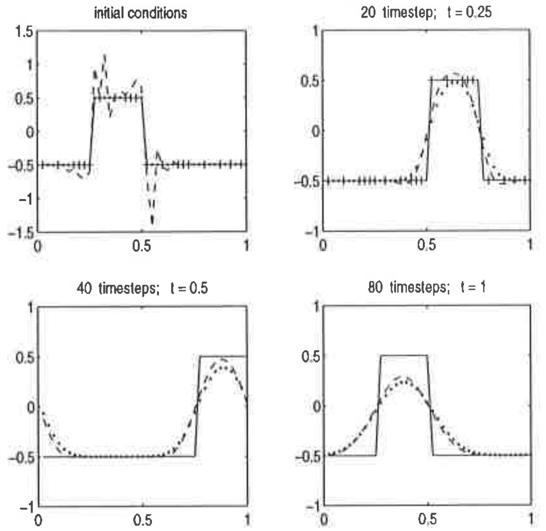
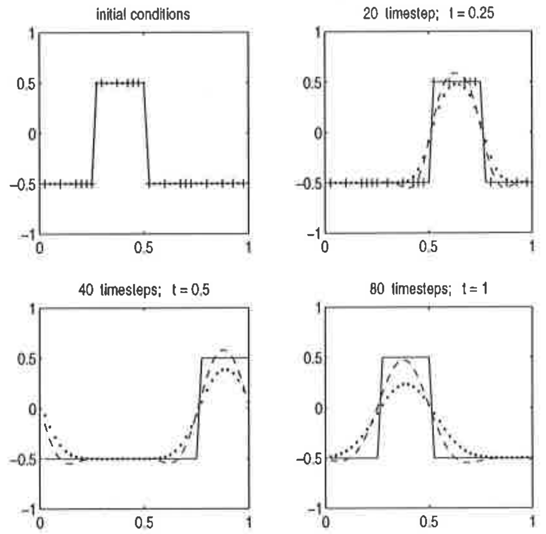


Figure 4: Example 2: Variational assimilation using the evolving error as the control vector.



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 [4].

6 Application to the nonlinear shallow water equations

The technique described here for treating model error in data assimilation is now applied to a discretized form of the one dimensional nonlinear shallow water equations. The flow described by these equations exhibits several features present in the dynamics of the atmosphere and oceans and the system is, therefore, used frequently in test problems.

6.1 The model

The system equations include rotation and bottom topography and are given by

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial \phi}{\partial x} = fv - g\frac{\partial H}{\partial x}, \quad (6.1a)$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} = -fu, \quad (6.1b)$$

$$\frac{\partial \phi}{\partial t} + u\frac{\partial \phi}{\partial x} + \phi\frac{\partial u}{\partial x} = 0, \quad (6.1c)$$

where $x \in [0, 2\pi L]$ and $t \in [0, T]$. The system variables $u = u(x,t)$ and $v = v(x,t)$ are the eastward and northward components of velocity and $\phi = \phi(x,t)$ is the geopotential, defined by $\phi = g\eta(x,t)$, where g is the acceleration due to gravity and $\eta(x,t)$ is the depth of the fluid. The height of the bottom topography is represented by $H = H(x)$, and f is the Coriolis parameter. Periodic boundary conditions are assumed. The equations are nonlinear and describe flow that may develop hydraulic jumps.

The model equations are obtained by applying the finite difference scheme derived in [11] to the flux form of the equations (6.1). The scheme uses artificial diffusion to eliminate spurious oscillations and is suitable for simulating hydraulic jumps. The model has also been used in [9] to investigate nonlinear data assimilation techniques. The discretization scheme uses centred time and space differencing, except for the diffusion terms, where forward time differencing is used for stability. The discrete model is given by

$$\begin{aligned} m_j^{k+1} = m_j^{k-1} & -\frac{\Delta t}{2\Delta x}\{(u_{j+1}^k + u_j^k)(m_{j+1}^k + m_j^k) - (u_j^k + u_{j-1}^k)(m_j^k + m_{j-1}^k) \\ & +((\phi_{j+1}^k)^2 - (\phi_{j-1}^k)^2)\} \\ & -g\frac{\Delta t}{2\Delta x}\{(\phi_{j+1}^k + \phi_j^k)(H_{j+1} - H_j) + (\phi_j^k + \phi_{j-1}^k)(H_j - H_{j-1})\} \\ & +2\Delta t f n_j^k + 2\frac{\Delta t}{\Delta x^2}K(m_{j+1}^{k-1} - 2m_j^{k-1} + m_{j-1}^{k-1}), \end{aligned} \quad (6.2a)$$

$$\begin{aligned} n_j^{k+1} = n_j^{k-1} & -\frac{\Delta t}{2\Delta x}\{(v_{j+1}^k + v_j^k)(m_{j+1}^k + m_j^k) - (v_j^k + v_{j-1}^k)(m_j^k + m_{j-1}^k)\} \\ & -2\Delta t f m_j^k + 2\frac{\Delta t}{\Delta x^2}K(n_{j+1}^{k-1} - 2n_j^{k-1} + n_{j-1}^{k-1}), \end{aligned} \quad (6.2b)$$

$$\phi_j^{k+1} = \phi_j^{k-1} - \frac{\Delta t}{\Delta x}(m_{j+1}^k - m_{j-1}^k) + 2\frac{\Delta t}{\Delta x^2}K(\phi_{j+1}^{k-1} - 2\phi_j^{k-1} + \phi_{j-1}^{k-1}), \quad (6.2c)$$

for $k = 1, .., N-1$, $j = 0, .., J-1$, where

$$m_j^k = \phi_j^k u_j^k, \qquad n_j^k = \phi_j^k v_j^k, \qquad \forall j,\ k. \quad (6.2d)$$

At the first time level the model equations are specified using one-step forward time differences. The periodic boundary conditions

$$u_J^k = u_0^k, \quad v_J^k = v_0^k, \quad \phi_J^k = \phi_0^k, \qquad k = 0, ..N-1, \quad (6.3)$$

are imposed. The model states $\phi_j^k \approx \phi(j\Delta x, k\Delta t)$, $u_j^k \approx u(j\Delta x, k\Delta t)$, $v_j^k \approx v(j\Delta x, k\Delta t)$, with $\Delta x = 2\pi L/J$, $\Delta t = T/N$, give approximations to the continuous variables satisfying the equations (6.1).

The 'true' states of the system are taken to be the solutions to the discrete equations (6.2) with the Coriolis parameter $f = 7.292 \times 10^{-5}\ s^{-1}$ and $L = 3.189 \times 10^6\ m$, which are suitable for a model approximating the earth's atmosphere at latitude 30^O North. The numbers of grid points and time steps are given by $J = 100$ and $N = 100$, respectively, and the time step satisfies $\Delta t/\Delta x = 0.1$, which ensures the stabilty of the scheme. The value of K is specified to be $K = 500\ m^2 s^{-1}$. The bottom topography consists of a ridge in the middle of the domain, defined by the function

$$H(x) = 0.5(1 - (x - L/2)^2/a^2), \qquad 0 \leq\ (x - L/2)\ \leq a, \tag{6.4}$$

where $a = 10\Delta x$. The initial values of the states are given by

$$m_j^0 = 0\ m^3 s^{-3}, \qquad n_j^0 = 0\ m^3 s^{-3}, \qquad \phi_j^0 = 10\ m^2 s^{-2}, \tag{6.5}$$

for $j = 0, ..., J - 1$. From these initial states, motion is initiated as the fluid flows down from the ridge in the centre of the domain and a wave travels in each direction across the domain.

6.2 Experimental results

Assimilation with noisy data is examined in the experiments. The observations are drawn from the 'true' solution of the discrete system corrupted by uniformly-distributed, unbiased, sequentially-uncorrelated random noise. The covariance matrices of the observations are taken to be $R_k = I,\ \forall k$, and it is assumed that observations are available at every grid point and at every time step. The prior estimates of the initial states are assumed to be exact. Model error is introduced by omitting the topography from the dynamical equations, making the model 'imperfect.' Because the topography is missing, if data assimilation is not applied, the model states remain in equilibrium with constant height and zero velocity fields and no motion is initiated. The aim of the data assimilation is then to estimate accurately the states of the 'true' system using the observations and the 'imperfect' model. The minimization problem is solved by the limited memory quasi-Newton procedure M1QN3 from the INRIA MODULOPT library [3]. The convergence criterion for the iteration is given by $\|\nabla_{\mathbf{u}} \mathcal{J}(\mathbf{u}^i)\| \,/\, \|\nabla_{\mathbf{u}} \mathcal{J}(\mathbf{u}^0)\| \leq 10^{-4}$, where $\mathbf{u}^i$ denotes the control variables at the i^{th} step of the iteration.

The noise corrupted 'true' data are shown in Figure 5a at times $t = 0,\ T/2,$ and T. The results of the assimilation are shown in Figure 5b, at the same times, for the case where the initial state is the only control variable and the error is not modelled. As in the previous examples, the assimilation does not reproduce 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. At the initial time, the height of the fluid is increased to compensate for the missing topography. At the middle and end of the interval, however, the assimilated solution produces better estimates of the flow variables than in the case where no assimilation is applied.

Figure 5c shows the results of the assimilation using the *augmented* system. The model error is assumed to be a constant bias error with initial covariance matrix $Q_0 = qI$ and $q = 1$. The assimilated solutions obtained by taking into account the model error are greatly improved over the whole time interval. The solutions at the end of the assimilation period, in particular, provide very good estimates of the true solution. In this case the model error is not, in fact, constant, but is serially correlated, and during the assimilation interval the effects of the model error propagate across the spatial domain. The bias error represents an average correction to the solution and adds significantly to the accuracy of the estimated flow over the entire assimilation period.

Figure 6 shows the forecast of the flow variables over the interval $t \in [T, 2T]$ in three cases. The results shown in Figures 6a and 6b are initiated from the assimilated solution obtained at time $t = T$ using the augmented system model, which takes into account the model error. The exact 'true' dynamics are shown in Figure 6c. In Figure 6a the 'imperfect' model is used to propagate the flow. The predictions, in this case, rapidly diverge from the true solution due to the unmodelled topography. In Figure 6b the bias error computed during the assimilation period is retained over the forecast interval and the flow variables are propagated using the augmented system. Retaining the bias correction is seen to give a significantly better forecast.

6.3 Further discussion

In addition to error in the topography of the system, other forms of error have been investigated in the shallow water model. Detailed results are presented in [4]. In the case where errors occur in the model rotation, for example, due to an incorrect Coriolis parameter, the constant bias error corrrection technique produces good estimates of the true solution over the assimilation interval and in the forecast.

Assimilation taking both the initial states and the model errors as control variables to correct simultaneously for initial and model errors is also successful. In the case where the prior estimate of the initial states is incorrect and model errors in both the topography and the rotation occur, excellent results are obtained by using the constant bias error correction together with the initial states as the controls. The accuracy of the estimated solutions is then improved greatly by the assimilation, but the convergence of the optimization procedure is found to be slow.

The case where observations are available at fewer spatial positions has also been examined. With noisy observational data and model error in the topography, the assimilated solutions are much rougher over intervals where there are fewer observations, and the noise in the data is reflected to a much greater extent in the estimated states. Increasing the weighting factor q acts to smooth the assimilated solutions and gives good estimates of the true states of the system, but accuracy is lost for large weightings.

Retaining the model error correction over the forecast interval can improve the preditions significantly, as shown in the experiments presented here. The intervals over which the constant bias error correction is effective are, however, limited. In both the assimilation and the forecast, extending the periods over which the constant bias error is applied can lead to a deterioration in the results. The impact of model error estimation thus depends on the length of the assimilation and forcast periods and attention is needed to determine the most effective time-scales over which to apply the error estimation procedures.

7 Conclusions

A new 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

This research was supported in part by the UK Meteorological Office and the UK Engineering and Physical Sciences Research Council as a CASE (Co-operative Awards in Science and Engineering) Project.

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] J.Ch. Gilbert and C. Lemarechal, Some numerical experiments with variable-storage quasi-Newton algorithems, *Mathematical Programming*, 45:407–435, 1989.

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

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

[6] A.K. Griffith and N.K. Nichols, Adjoint methods for treating error in data assimilation, in *Numerical Methods for Fluid Dynamics VI*, (Ed. M.J. Baines), ICFD, Oxford, 335-344, 1998.

[7] A.K. Griffith, M.J. Martin and N.K. Nichols, Techniques for treating systematic model error in 3D and 4D data assimilation, in *Proceedings of the Third WMO Int. Symposium on Assimilation of Observations in Meteorology and Oceanography*, World Meteorological Organization, (to appear).

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

[9] A.C. Lorenc, Optimal nonlinear objective analysis, *Quart.J.Roy.Met.Soc.*, 114:205–240, 1988.

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

[11] C.A. Parrett and M.J.P. Cullen, Simulation of hydraulic jumps in the presence of rotation and mountains, *Quart.J.Roy.Met.Soc.*, 110:147–165, 1984.

[12] 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.

[13] W.C. Thacker, Relationships between statistical and deterministic methods of data assimilation, in *Variational Methods in the Geosciences*, (Ed. Y.K. Sasaki), Elsevier, New York, 1996.

[14] 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.

Figure 5: Shallow Water Model: (a) true solution - observations corrupted by random error; (b) assimilated solution without error correction; (c) assimilated solution with error correction. Solutions are shown at times t = 0, T/2, T. Dotted line: ϕ-field; dashed line: n-field; solid line: m-field.

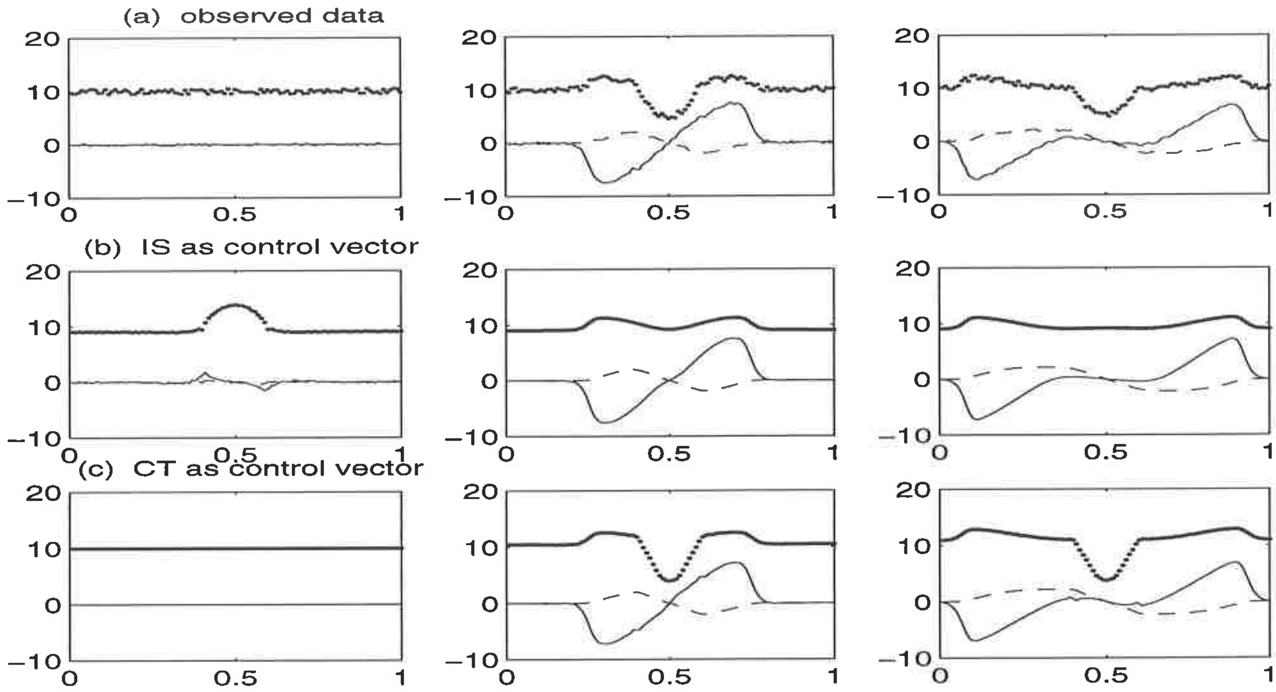


Figure 6: Shallow Water Model: Forecast on interval $t \in [T, 2T]$ – (a) forecast from assimilated solution without error correction; (b) forecast from assimilated solution with error correction included; (c) forecast from true solution with exact dynamics. Soutions are shown at times t = T, 3T/2, 2T. Dotted line: ϕ-field; dashed line: n-field; solid line: m-field.