

DATA ASSIMILATION: AIMS AND BASIC CONCEPTS

N.K.NICHOLS

*Department of Mathematics, The University of Reading,
Box 220, Reading, RG6 6AX United Kingdom*

1. Introduction

Atmosphere and ocean systems can be simulated effectively by discrete numerical models and, provided that the initial states of the system are known, accurate forecasts of future dynamical behaviour can be determined. Complete information defining all of the states of the system at a specified time are, however, rarely available. Moreover, both the models and the measured data contain inaccuracies and random noise. In this case, observations of the system measured over an interval of time can be used in combination with the model equations to derive estimates of the expected values of the states. The problem of constructing a ‘state-estimator,’ or ‘observer,’ for these systems can be treated by using feedback design techniques from control theory. For the very large nonlinear systems arising in climate, weather and ocean prediction, however, traditional control techniques are not practicable and ‘data assimilation’ schemes are used instead to generate accurate state-estimates (see, for example, Daley, 1994; Bennett, 1992).

The aim of data assimilation is 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 problem can be stated as follows.

Problem 1 *Given a (noisy) discrete model of the dynamics of a system, find estimates of the system states from (noisy) observations.*

The most significant properties of the data assimilation problem are that the models are very large and nonlinear, with order $O(10^7)$ state variables. The number of observations is also large, of order $O(10^5 - 10^6)$.

There are two approaches to this problem. The first uses a ‘dynamic observer,’ which gives a *sequential data assimilation scheme*, and the second uses a ‘direct observer,’ which gives a *four-dimensional data assimilation*

scheme. In the next section these two approaches are illustrated for a simple linear time-invariant system. In the following sections data assimilation schemes for a full nonlinear system are developed and examples are presented.

2. Data Assimilation in Linear Systems

We first examine a discrete linear time-invariant system model. The dynamical equations describing the evolution of the states from time t_k to time t_{k+1} are given by

$$\mathbf{x}_{k+1} = F\mathbf{x}_k + G\mathbf{u}_k, \quad k = 0, \dots, N-1, \quad (1)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ denotes the model states and $\mathbf{u}_k \in \mathbb{R}^m$ denotes the model inputs at time t_k , and $F \in \mathbb{R}^{n \times n}$, $G \in \mathbb{R}^{n \times m}$ are the system matrices. The input vector contains the known forcing functions that drive the system, and it is assumed that for given inputs and for any initial state \mathbf{x}_0 given at time t_0 , the system equations uniquely determine the states of the system for all future times.

The observations are linearly related to the system states by the equations

$$\mathbf{y}_k = H\mathbf{x}_k, \quad k = 0, \dots, N-1, \quad (2)$$

where $\mathbf{y}_k \in \mathbb{R}^p$ is a vector of p observations at time t_k and $H \in \mathbb{R}^{p \times n}$ is the observation matrix, which includes transformations and grid interpolations.

In the case where \mathbf{x}_0 is not known (accurately), the objective of the assimilation schemes is to use measured data over time to estimate correctly the states of the system at times t_k , $k > 0$. Two classes of methods can be applied.

2.1. SEQUENTIAL DATA ASSIMILATION

With sequential assimilation, *a priori* estimates for the initial states \mathbf{x}_0 are chosen and the model is evolved forward to the time t_k where the first observations are available. The predicted states of the system at this time are known as the background states and are denoted by \mathbf{x}_k^b . The difference between the predicted observation vector given by the background states and the vector of measured observations at this time ($H\mathbf{x}_{k+1}^b - \mathbf{y}_{k+1}$), known as the ‘innovation vector,’ is then used to make a correction to the background state vector in order to obtain improved state estimates \mathbf{x}_k^a , known as the analysis states. The model is then evolved forward again from the analysis states to the next time where an observation is available and the process is repeated.

Mathematically this procedure may be written

$$\mathbf{x}_{k+1}^b = F\mathbf{x}_k^a + G\mathbf{u}_k, \quad (3)$$

$$\mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^b + K(H\mathbf{x}_{k+1}^b - \mathbf{y}_{k+1}). \quad (4)$$

The matrix $K \in \mathbb{R}^{n \times p}$, known as the ‘gain matrix,’ must be chosen to ensure that the analysis states converge to the true states of the system over time. This is always possible if the system is ‘observable.’ Conditions for this property to hold are well-known. (See, for example, Barnett and Cameron, 1985.)

The system (3)-(4) forms a modified dynamical system for the analysis states which can be written

$$\mathbf{x}_{k+1}^a = (F + KHF)\mathbf{x}_k^a + (G + KHG)\mathbf{u}_k - K\mathbf{y}_{k+1}. \quad (5)$$

This system has different properties from the original discrete system model (1) and the stability of its response depends upon the spectrum of the modified system matrix $F + KHF$. The choice of the gain matrix K therefore determines the behaviour of the analysed states over time and this choice characterizes the data assimilation scheme.

2.2. FOUR-DIMENSIONAL DATA ASSIMILATION

In contrast to sequential data assimilation, which evolves the model one step at a time and updates the estimated states each time an observation is available, the four-dimensional assimilation schemes use all the observations available over a given time window to provide improved estimates for all the states in that window. Thus ‘time’ is the fourth dimension in the problem.

Mathematically we may rearrange the system equations (1)-(2) in order to express the observations over time in terms of the initial (unknown) states \mathbf{x}_0 . Using the relation

$$\mathbf{y}_k = H\mathbf{x}_k = HF\mathbf{x}_{k-1} + HG\mathbf{u}_{k-1}, \quad (6)$$

we may rewrite the system equations at points t_k , $k = 0, 1, \dots, N - 1$, in the form

$$H\mathbf{x}_0 = \mathbf{y}_0, \quad (7)$$

$$HF\mathbf{x}_0 = \mathbf{y}_1 - HG\mathbf{u}_0, \quad (8)$$

$$HF^2\mathbf{x}_0 = \mathbf{y}_2 - HG\mathbf{u}_1 - HFG\mathbf{u}_0, \quad (9)$$

$$\vdots \quad (10)$$

$$HF^{N-1}\mathbf{x}_0 = \mathbf{y}_{N-1} - \sum_{j=0}^{N-2} HF^j G\mathbf{u}_{N-2-j}. \quad (11)$$

These equations form an $Np \times N$ over-determined system of linear equations for the unknown initial states that may be solved by the method of least squares. The problem may be written

$$\min_{\mathbf{x}_0} \|V\mathbf{x}_0 - \mathbf{r}\|_2^2, \quad (12)$$

where $V^T = [H^T, (HF)^T, (HF^2)^T, \dots, (HF^{N-1})^T]$, \mathbf{r} denotes the right-hand-side of the equations (7)–(11) and $\|\mathbf{z}\|_2$ denotes the L_2 -vector norm of \mathbf{z} , given by $(\mathbf{z}^T \mathbf{z})^{\frac{1}{2}}$. The matrix V is of full rank if the system is ‘observable’ and $N \geq n$. The least square problem then has a unique solution.

Using the system equations, we may write the linear least-square problem (12) as a constrained least square problem

$$\min_{\mathbf{x}_0} \frac{1}{2} \sum_{k=0}^{N-1} (H\mathbf{x}_k - \mathbf{y}_k)^T (H\mathbf{x}_k - \mathbf{y}_k), \quad (13)$$

subject to the conditions that \mathbf{x}_k , $k = 1, \dots, N - 1$, satisfy the model equations (1) starting with initial data \mathbf{x}_0 . The problem is very large and may be solved using an iterative procedure.

In this approach the initial states are treated as parameters that must be selected to minimize the mean square errors between the observations predicted by the model and the measured observations over the entire time window. The initial data is adjusted to different positions in order to achieve the best fit, using an efficient iterative minimization algorithm.

2.3. SUMMARY

Two approaches to the problem of data assimilation are illustrated here in the case of a discrete linear deterministic dynamical model. These same approaches apply to models of the atmosphere and oceans, but the models are nonlinear and contain uncertainties. In practice the model equations are *not accurate*, the input data are *not accurate*, the observational data are *not accurate*, and the computation is *not accurate*! In the next section we develop sequential and four-dimensional assimilation schemes for treating nonlinear systems with observations that contain random noise. In a subsequent chapter we consider nonlinear stochastic systems containing both random and systematic model errors.

3. Data Assimilation in Nonlinear Dynamical Systems

A variety of models are used to describe systems arising in atmosphere and ocean applications, as well as in other physical, biological and economic fields. These range from simple linear, deterministic, continuous ordinary

differential equation models to sophisticated nonlinear stochastic partial differential or discrete models. The data assimilation schemes, with minor modifications, can be applied to any general model.

Data assimilation schemes are described here for a system modelled by the discrete nonlinear equations

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k), \quad k = 0, \dots, N - 1, \quad (14)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ denotes the model states and $\mathbf{u}_k \in \mathbb{R}^{m_k}$ denotes the m_k known inputs to the system at time t_k , and $\mathbf{f}_k : \mathbb{R}^n \times \mathbb{R}^{m_k} \rightarrow \mathbb{R}^n$ is a nonlinear function describing the evolution of the states from time t_k to time t_{k+1} .

The observations are assumed to be related to the system states by the equations

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \boldsymbol{\delta}_k, \quad k = 0, \dots, N - 1, \quad (15)$$

where $\mathbf{y}_k \in \mathbb{R}^{p_k}$ is a vector of p_k observations at time t_k and $\mathbf{h}_k : \mathbb{R}^n \rightarrow \mathbb{R}^{p_k}$ 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 $R_k \in \mathbb{R}^{p_k \times p_k}$.

Prior estimates, or ‘background estimates,’ \mathbf{x}_0^b of the initial states \mathbf{x}_0 are assumed to be known and the initial random errors $(\mathbf{x}_0 - \mathbf{x}_0^b)$ are 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 data assimilation scheme, or ‘state-estimator,’ for this model is defined by the system

$$\mathbf{x}_{k+1}^b = \mathbf{f}_k(\mathbf{x}_k^a, \mathbf{u}_k) \quad (16)$$

$$\mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^b + K_{k+1}(\mathbf{h}_k(\mathbf{x}_{k+1}^b) - \mathbf{y}_{k+1}), \quad (17)$$

for $k = 0, \dots, N - 1$. As for the linear model, these equations represent a prediction for the background states from the model equations, followed by an update for the analysed states using the innovations $(\mathbf{h}_k(\mathbf{x}_{k+1}^b) - \mathbf{y}_{k+1})$. The gain matrices K_k must be chosen to ensure that the analysed states \mathbf{x}_k^a converge to the expected values $E(\mathbf{x}_k)$ of the true states as $k \rightarrow \infty$ for *any* initial background state \mathbf{x}_0^b . The choice of the gain matrices characterizes the data assimilation scheme. The primary question in designing the scheme therefore concerns how K_k should be selected.

For the ‘optimal’ analysis, we aim to find the best estimates \mathbf{x}_k^a for the expected values of the true states \mathbf{x}_k , $k = 0, \dots, N - 1$, given observations \mathbf{y}_k , $k = 0, \dots, N - 1$, subject to the model equations (14) and prior

estimates \mathbf{x}_0^b . Under the statistical assumptions made here, the optimal analysis is given by the maximum likelihood *a priori* Bayesian estimate of the system states. The problem reduces to minimizing the square error between the observations predicted by the model and the measured observations, weighted by the inverse of the covariance matrices, over the assimilation window (Lorenç, 1986). The data assimilation problem is then defined explicitly as follows.

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

$$\mathcal{J} = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T B_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2} \sum_{k=0}^{N-1} (\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k)^T R_k^{-1}(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k), \quad (18)$$

subject to \mathbf{x}_k , $k = 1, \dots, N - 1$, satisfying the system equations (14) with initial states \mathbf{x}_0 .

The model is assumed to be ‘perfect’ and the system equations are treated as strong constraints on the minimization problem. The states \mathbf{x}_k that satisfy the model equations (14) are uniquely determined by the initial states of the system and therefore can be written explicitly in terms of \mathbf{x}_0 , as in the linear case. Substituting into the objective function (18) then allows the optimization problem to be expressed in terms of the initial states alone. The assimilation problem, Problem 2, thus becomes an unconstrained weighted least squares problem where the initial states are the required control variables in the optimization.

The aim is to determine the analysis of the initial states that produces the best fit, in a statistical sense, between the observations predicted by the analysis and the measured observations over the assimilation window. The first term in the objective function constrains the choice of the initial states and acts as a ‘regularizing’ term. Since the covariance matrices B_0 , R_k , $k = 0, 1, \dots, N - 1$, are of full rank, the resulting least square problem necessarily has a unique solution.

The data assimilation problem, Problem 2, can be solved directly to give a sequential assimilation scheme, or it can be solved indirectly to give a four-dimensional ‘variational’ assimilation scheme. These two techniques are described in the next subsections.

3.1. SEQUENTIAL ASSIMILATION SCHEMES

To derive the ‘optimal’ sequential assimilation scheme for the nonlinear system, we assume that we have prior estimates for the background states

\mathbf{x}_k^b at time step t_k with error covariance matrix

$$B_k = E((\mathbf{x} - \mathbf{x}_k^b)(\mathbf{x} - \mathbf{x}_k^b)^T). \quad (19)$$

The optimal analysis \mathbf{x}_k^a then solves the problem

$$\min_{\mathbf{x}} \left[\frac{1}{2} (\mathbf{x} - \mathbf{x}_k^b)^T B_k^{-1} (\mathbf{x} - \mathbf{x}_k^b) + \frac{1}{2} (\mathbf{h}_k(\mathbf{x}) - \mathbf{y}_k)^T R_k^{-1} (\mathbf{h}_k(\mathbf{x}) - \mathbf{y}_k) \right]. \quad (20)$$

The optimal solution is given by

$$\mathbf{x}_k^a = \mathbf{x}_k^b + K_k (\mathbf{h}_k(\mathbf{x}_k^b) - \mathbf{y}_k), \quad (21)$$

where

$$K_k = B_k H_k^T (H_k B_k H_k^T + R_k)^{-1} \quad (22)$$

and $H_k = \frac{\partial \mathbf{h}_k}{\partial \mathbf{x}} \Big|_{\mathbf{x}_k^b}$. (The matrix H_k is the ‘Jacobian’ of the function \mathbf{h}_k with respect to \mathbf{x} evaluated at \mathbf{x}_k^b , and thus $H_k(\mathbf{x} - \mathbf{x}_k^b)$ approximates $(\mathbf{h}_k(\mathbf{x}) - \mathbf{h}_k(\mathbf{x}_k^b))$ for small $\|\mathbf{x} - \mathbf{x}_k^b\|_2$).

For linear systems the solution (21)-(22) gives the exact optimal analysis, but for nonlinear systems this solution gives only a first order approximation to the optimal due to the linearization H_k of the nonlinear observation operator that is used (Lorenc, 1986, 1988; Thacker, 1996).

In evolving the ‘optimal’ analysis sequentially, two computational difficulties arise. The first is that the background covariance matrices B_k are required at each time step. These matrices can be propagated forward in time from the initial background error covariance matrix B_0 using an extended Kalman filter (EKF) technique (Kalman, 1961). For full-scale weather and ocean systems, however, the EKF requires the computation of matrices of order $O(10^7 \times 10^7)$ at every time step, making it very expensive to implement. Moreover, for nonlinear models the EKF is highly sensitive to computational round-off errors and may become unstable as a dynamical system (Bierman, 1977).

The second difficulty in implementing the optimal assimilation scheme sequentially is that in order to compute the analysis \mathbf{x}_k^a at each time step, we must find $B_k H_k^T \mathbf{w}_k^a$, where \mathbf{w}_k^a solves the linear equations

$$(H_k B_k H_k^T + R_k) \mathbf{w}_k^a = (\mathbf{h}_k(\mathbf{x}_k^b) - \mathbf{y}_k). \quad (23)$$

This is a very large inverse problem with $O(10^5 - 10^6)$ variables to find.

In practice most operational sequential assimilation schemes avoid these two difficulties by using approximations that can be implemented efficiently. A variety of schemes have been developed that differ mainly in the detailed steps of the procedures.

Sequential data assimilation schemes used operationally include:

– *Successive Correction*. In these schemes, the feedback gain K_k is not chosen optimally, but is designed to smooth observations into the states at all spatial grid points within some radius of influence of each observation (Bergthorsen and Doos, 1955). An iterative process is used to determine the analysis. The Cressman scheme is an example (Cressman, 1959). The iterations converge to a result that is consistent with observational error but may not be consistent with the dynamical system equations. Over time the analysis states may not converge to the expected values of the true states. These schemes are generally not effective in data sparse regions.

– *Optimal Interpolation or Statistical Interpolation*. These schemes approximate the optimal solution by replacing the optimal gain matrix K_k in (21) by the approximation

$$\tilde{K}_k = \tilde{B}H_k^T(H_k\tilde{B}H_k^T + R_k)^{-1},$$

where \tilde{B} has a ‘fixed’ structure for all k . (See Ghil and Malanotte-Rizzoli, 1991.) The matrix \tilde{B} is generally defined by an isotropic correlation function (dependent only on the distance between spatial grid points and observational points), with the correlation lengths adjusted empirically. To simplify the inversion step, the gain is further modified to have a block structure by using innovations only in small regions around grid points to obtain the analysis states. The inversion problem then reduces to solving a number of much smaller systems of equations.

– *Analysis Correction*. In these schemes, approximations to the optimal analysis states are computed iteratively, as in the Successive Correction method. The procedure is designed, however, to ensure that the iterates converge to the approximate ‘optimal’ analysis that is obtained by replacing the optimal gain K_k in (21) by the gain \tilde{K}_k , as defined for the optimal interpolation scheme (Bratseth, 1986; Lorenc, Bell and MacPherson, 1991). This scheme is effective across data sparse regions and the analysis produced remains consistent with the dynamical equations.

– *3DVAR*. These schemes apply iterative minimization methods directly to the variational problem (20) (Rabier *et al.*, 1993). The covariance matrix B_k is replaced by the approximation \tilde{B} , as defined for optimal interpolation. The solution converges to the analysis obtained by replacing the optimal gain K_k by \tilde{K}_k in (21). Minimization techniques used commonly are Pre-conditioned Conjugate Gradient methods and Quasi-Newton methods. The properties of the analysis are similar to those obtained by the Analysis Correction method, but the iteration procedure is more efficient.

– *3DPSAS*. In these schemes iterative minimization methods are applied

to the dual variational problem

$$\min_{\mathbf{w}} \left[\frac{1}{2} (\mathbf{w}^T H_k \tilde{B} H_k + R_k) \mathbf{w} - \mathbf{w}^T (\mathbf{h}_k(\mathbf{x}) - \mathbf{y}_k) \right].$$

The iterates converge to the solution \mathbf{w}_k^a of the system (23) and the resulting analysis states converge to $\mathbf{x}_k^a = \tilde{B} H_k^T \mathbf{w}_k^a$, which approximates the ‘optimal’ solution as in the 3DVAR scheme (Cohn *et al.*, 1998). The advantage is that this scheme operates in the ‘observation space,’ which is of lower dimension than the state space. Additional work is needed, however, in order to reconstruct the analysis states.

In summary, most operational sequential data assimilation schemes aim to approximate the optimal analysis by replacing the background error covariance matrix by an approximation that is fixed over time and by simplifying the inversion problem and/or solving the inversion iteratively. Simple examples illustrating the use of these schemes can be found in Martin *et al.* (1999).

3.2. FOUR-DIMENSIONAL ASSIMILATION SCHEMES

The full four-dimensional data assimilation problem is currently treated in operational centres using four-dimensional variational schemes (4DVAR) (Sasaki, 1970; Talagrand, 1981; Rabier *et al.*, 2000). In these schemes the constrained minimization problem, Problem 2, is solved iteratively by a gradient optimization 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 (14) must be satisfied. The adjoint equations are given by

$$\boldsymbol{\lambda}_N = 0, \quad (24)$$

$$\boldsymbol{\lambda}_k = F_k^T(\mathbf{x}_k) \boldsymbol{\lambda}_{k+1} - H_k^T R_k^{-1} (\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k), \quad k = N - 1, \dots, 0, \quad (25)$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$, $j = 0, \dots, N$, 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 adjoint variables $\boldsymbol{\lambda}_k$ measure the sensitivity of the objective function (18) to changes in the solutions \mathbf{x}_k of the state equations for each k .

The gradient of the objective function (18) with respect to the initial data \mathbf{x}_0 is then given by

$$\nabla_{\mathbf{x}_0} \mathcal{J} \equiv B_0^{-1} (\mathbf{x}_0 - \mathbf{x}_0^b) - \boldsymbol{\lambda}_0. \quad (26)$$

At the optimal, the gradient (26) 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.

A dual approach (4DPSAS) in which the minimization is performed in observation space is also possible (Courtier, 1987). In this scheme, as in the three dimensional 3DPSAS method, a dual four-dimensional variational problem is solved using a gradient iteration method, and the analysis states are then reconstructed from the dual variables.

To make these variational methods more efficient, an ‘incremental’ approach may be used in which the forward solution of the nonlinear model equations in the iterative procedure is replaced by the forward solution of an approximate ‘incremental’ linear system (Courtier, Thepaut and Hollingsworth, 1994). In the variational problem the unknown states are replaced by $\mathbf{x}_k = \mathbf{x}_k^b + \delta\mathbf{x}_k$ and the innovations are denoted $\mathbf{d}_k = -(\mathbf{h}_k(\mathbf{x}_k) - \mathbf{y}_k)$. The variational problem then becomes (to first order)

$$\min_{\delta\mathbf{x}_0} \left[\frac{1}{2} \delta\mathbf{x}_k^T B_0^{-1} \delta\mathbf{x}_k + \frac{1}{2} \sum_{k=0}^{N-1} (H_k \delta\mathbf{x}_k - \mathbf{d}_k)^T R_k^{-1} (H_k \delta\mathbf{x}_k - \mathbf{d}_k) \right], \quad (27)$$

subject to the tangent linear model (TLM) equations

$$\delta\mathbf{x}_{k+1} = F_k \delta\mathbf{x}_k, \quad (28)$$

where $F_k \in \mathbb{R}^{n \times n}$ and $H_k \in \mathbb{R}^{n \times p_k}$ are the Jacobians of \mathbf{f}_k and \mathbf{h}_k .

The incremental procedure requires first a forward solution to the nonlinear model (14), starting from the initial background states \mathbf{x}_0^b , in order to determine the background states \mathbf{x}_k^b and the innovations \mathbf{d}_k for each k . Then an initial choice for $\delta\mathbf{x}_0$ is made and a forward solution to the tangent linear model (28) is found. The backward solution of the corresponding adjoint system and the gradient of the linearized objective function (27) are then determined, and the minimization algorithm is used to find an improved estimate for $\delta\mathbf{x}_0$. The linear process is repeated to obtain further updates to $\delta\mathbf{x}_0$. The success of this procedure depends on how well the tangent linear model approximates the nonlinear model. After a number of iterations of the linearized model, the background solution to the nonlinear model is usually updated by a forward solve of the nonlinear equations, starting from the incremented initial states. The full incremental variational procedure thus consists of an inner and outer iteration process.

The primary difficulty in implementing variational assimilation schemes is the need to develop a full adjoint model for the system. The adjoint equa-

tions are related theoretically to the linearized state equations, and the system matrix of the adjoint model is given directly by F_k^T , where F_k is the system matrix of the linearized model. The adjoint equations can thus be generated directly from the linearized system equations. Automatic differentiation techniques can be applied to the forward solution code to generate the adjoint code (Griewank and Corliss, 1991). Alternatively an approximate adjoint system can be obtained by discretizing a continuous linear or adjoint model of the nonlinear dynamics (Lawless, 2001). This approach has the advantage that additional approximations can be incorporated into the linearization of the system equations.

Four-dimensional variational data assimilation schemes are currently in operational use and are under further development. Examples illustrating the use of these schemes can be found in Griffith (1997) and Lawless (2001).

4. Conclusions

The aims and basic concepts of data assimilation for atmosphere and ocean systems are described here. Two approaches to the problem of data assimilation, sequential and four-dimensional assimilation, are introduced. A variety of assimilation schemes for discrete nonlinear system models are derived and practical implementation issues are discussed. For all of these schemes, the model equations are assumed to be ‘perfect’ representations of the true dynamical system. In practice the models contain both systematic errors and random noise. In a subsequent chapter in this volume we discuss assimilation techniques for treating stochastic model error and systematic bias errors.

References

- Barnett, S. and Cameron, R.G. (1985) *Introduction to the Mathematical Theory of Control*, 2nd edition, Clarendon Press, Oxford, UK.
- Bennett, A.F. (1992) *Inverse Methods in Physical Oceanography*, Cambridge University Press, Cambridge, UK.
- Bergthorsson, P. and Doos, B.R. (1955) Numerical weather map analysis, *Tellus*, **7**, 329–340.
- Bierman, G.L. (1977) *Factorization Methods for Discrete Sequential Estimation*, Mathematics in Science and Engineering, V. 128, Academic Press, New York.
- Bratseth, A.M. (1986) Statistical interpolation by means of successive corrections, *Tellus*, **38A**, 439–447.
- Cohn, S. E., da Silva, A., Guo, J., Sienkiewicz, M. and Lamich, D. (1998) Assessing the effects of data selection with the DAO physical-space statistical analysis system, *Monthly Weather Review*, **126**, 2913–2926.
- Courtier, P. (1997) Dual formulation of four-dimensional variational assimilation, *Quart. J. Roy. Met. Soc.*, **123**, 2449–2461.
- Courtier, P., Thepaut, J-N. and Hollingsworth, A. (1994) A strategy for operational implementation of 4D-Var, using an incremental approach, *Quart. J. Roy. Met. Soc.*, **120**, 1367–1387.

- Cressman, G. (1959) An optimal objective analysis system, *Monthly Weather Review*, **87**, 367–374.
- Daley, R. (1994) *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, UK.
- Ghil, M. and Malanotte-Rizzoli, P. (1991) Data assimilation in meteorology and oceanography, *Adv. Geophys.*, **33**, 141–266.
- Griewank, A. and Corliss, G.F. (1991) *Automatic Differentiation of Algorithms*, SIAM, Philadelphia.
- Griffith, A.K. (1997) *Data Assimilation for Numerical Weather Prediction Using Control Theory*, The University of Reading, Department of Mathematics, PhD Thesis.
- Kalman, R.E. (1961) A new approach to linear filtering and prediction problems, *Transactions of the ASME, Series D*, **83**, 35–44.
- Lawless, A.S. (2001) *Development of Linear Models for Data Assimilation in Numerical Weather Prediction*, Department of Mathematics, The University of Reading, Ph.D. Thesis.
- Lorenc, A.C. (1986) Analysis methods for numerical weather prediction, *Quart. J. Roy. Met. Soc.*, **112**, 1177–1194.
- Lorenc, A.C. (1988) Optimal nonlinear objective analysis, *Quart. J. Roy. Met. Soc.*, **114**, 205–240.
- Lorenc, A.C., Bell, R.S. and Macpherson, B. (1991) The Met. Office analysis correction data assimilation scheme, *Quart. J. Roy. Met. Soc.*, **117**, 59–89.
- Martin, M.J., Nichols, N.K. and Bell, M.J. (1999) *Treatment of Systematic Errors in Sequential Data Assimilation*, Meteorological Office, Ocean Applications Division, Tech. Note, No. 21.
- Rabier, F., Courtier, P., Pailleux, J., Talalgrand, O. and Vasiljevic, D. (1993) A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional variational analysis, *Quart. J. Roy. Met. Soc.*, **119**, 845–880.
- Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. (2000) The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Quart. J. Roy. Met. Soc.*, **126**, 1143–1170.
- Sasaki, Y. (1970) Some basic formulisms on numerical variational analysis, *Monthly Weather Review*, **98**, 875–883.
- Talagrand, O. (1981) A study on the dynamics of four-dimensional data assimilation, *Tellus*, **33**, 43–60.
- Thacker, W.C. (1996) Relationships between statistical and deterministic methods of data assimilation, in Y.K. Sasaki (ed.), *Variational Methods in the Geosciences*, Elsevier, New York.