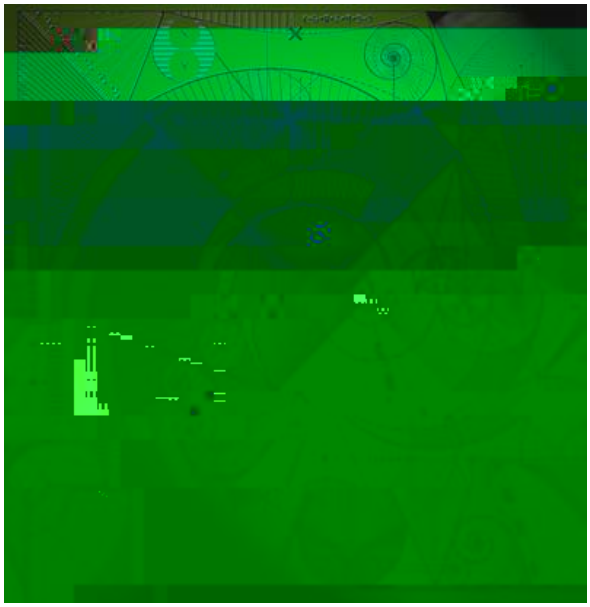
School of Mathematics, Meteorology
and Physics

Department of Mathematics

Preprint MPS_2009-04



MATHEMATICAL CONCEPTS OF DATA ASSIMILATION

N.K. NICHOLS
Department of Mathematics, University of Reading,
PO Box 220, Whiteknights, Reading, RG6 6AX UK
email: *n.k.nichols@reading.ac.uk*

In: *Data Assimilation: Making Sense of Observations*, eds. W Lahoz, R Swinbank and B Khattatov, Springer, 2009 (ISBN: 978-3-540-74702-4)[1]

1. Introduction

Environmental systems can be realistically described by mathematical and numerical models of the system dynamics. These models can be used to predict the future behaviour of the system, provided that the initial states of the system are known. Complete data de¯ning all of the states of a system at a speci¯c time are, however, rarely available. Moreover, both the models and the available initial data contain inaccuracies and random noise that can lead to signi¯cant di®erences between the predicted states and the actual states of the system. In this case, observations of the system over time can be incorporated into the model equations to derive `improved' estimates of the states and also to provide information about the `uncertainty' in the estimates.

The problem of state-estimation is an inverse problem and can be treated using observers and/or ¯lters derived by feedback design techniques from control theory (see, for example, Barnett and Cameron, 1985). For the very large nonlinear systems arising in the environmental sciences, however, traditional control techniques are not practicable and `data assimilation' schemes have been developed to generate accurate state-estimates

[1] *Link:* http://www.springer.com/earth+sciences/computer+%26+mathematical+applications/book/978-3-540-74702-4

(see, for example, Daley, 1993; Bennett, 1992). The aim of such schemes can be stated as follows.

lation problem for this case and examine its properties. Next we determine a best linear estimate of the solution to the nonlinear assimilation problem. The data assimilation scheme is then interpreted in a stochastic framework and the `optimal' state-estimate is derived using statistical arguments. We consider the case where the model includes errors in the system equations in a later section of this chapter.

2.1. BASIC LEAST-SQUARES FORMULATION FOR PERFECT MODELS

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

$$\mathbf{x}_{k+1} = \mathcal{M}_{k;k+1}(\mathbf{x}_k); \quad k = 0;:::;N-1; \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ denotes the vector of n model states at time t_k and $\mathcal{M}_{k;k+1} : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear operator describing the evolution of the states from time t_k to time t_{k+1}. The operator contains known inputs to the system including known external forcing functions that drive the system and known parameters describing the system.

Prior estimates, or `background estimates,' $\mathbf{x}_0^b$, of the initial states $\mathbf{x}_0$ at time t_0 are assumed to be known, usually provided by a previous forecast.

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

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \pm_k; \quad k = 0;:::;N; \tag{2}$$

where $\mathbf{y}_k \in \mathbb{R}^{p_k}$ is a vector of p_k observations at time t_k and $\mathcal{H}_k : \mathbb{R}^n \to \mathbb{R}^{p_k}$ is a nonlinear operator that includes transformations and grid interpolations. The observational errors $\pm_k \in \mathbb{R}^{p_k}$ consist of instrumentation errors and representativity errors (see Lahoz *et al.*, this book).

For the `optimal' analysis, we aim to ¯nd the best estimates $\mathbf{x}_k^a$ for the system states $\mathbf{x}_k;\ k = 0;:::;N;$ to ¯t the observations $\mathbf{y}_k;\ k = 0;:::;N;$ and the background state $\mathbf{x}_0^b$, subject to the model equations (1). We write the problem as a weighted nonlinear least-squares problem constrained by the model equations.

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

$$\mathcal{J} = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) +$$

$$+ \frac{1}{2}\sum_{k=0}^{N} (\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1} (\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)$$

subject to ¯nd/F210.91Tf5.121.78TD[( ) 11.78TD[(( ) J/F470.91

The model is assumed here to be `perfect' and the system equations are

the weighting matrix $\mathbf{B}_0$ is nonsingular, however, then, provided the operators $\mathcal{M}_{0;k}$ and $\mathcal{H}_k$ are continuously differentiable, the stationary points of the least-squares problem are well-defined. The weighted background term acts as a `regularization' term, ensuring the existence of a solution and also damping the sensitivity of the solution to the observational errors (Johnson *et al.*, 2005b, 2005a).

Under these conditions, the stationary points of the objective function (4) satisfy the gradient equation, given by

$$\nabla_{\mathbf{x}_0}\mathcal{J} = \mathbf{J}^T\mathbf{f}(\mathbf{x}_0) = 0; \tag{6}$$

where $\mathbf{J}$ is the Jacobian of the vector function $\mathbf{f}$ defined in (5). The Jacobian can be w-d652.741n2-1tf3uclMp12.-1tf3ufond

2.3. BEST LINEAR LEAST-SQUARES ESTIMATE

In general, explicit solutions to the nonlinear data assimilation problem, Problem 1, cannot be found. A `best' *linear* estimate of the solution to the nonlinear problem can, however, be derived explicitly. We assume that the departure of the estimated analysis $\mathbf{x}_0^a$ from the background $\mathbf{x}_0^b$ is a *linear* combination of the innovations $\mathbf{d}_k = \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^b)$, $k = 0;1;:::;N$, and ¯nd the estimate for $\mathbf{x}_0^a$ that solves the least-squares data assimilation problem as accurately as possible.

To determine the estimate, we linearize the assimilation problem about the nonlinear background trajectory $\mathbf{x}_k^b = \mathcal{M}_{0;k}(\mathbf{x}_0^b)$, $k = 1;:::;N$. We denote by the matrices $\mathbf{H}_k$ and $\mathbf{M}_{0;k}$ the linearizations of the observation and model operators $\mathcal{H}_k$ and $\mathcal{M}_{0;k}$, respectively, about the background trajectory; that is,

$$\mathbf{H}_k = \frac{@\mathcal{H}_k}{@\mathbf{x}}|_{\mathbf{x}_k^b}; \qquad \mathbf{M}_{0;k} = \frac{@\mathcal{M}_{0;k}}{@\mathbf{x}}|_{\mathbf{x}_0^b}:$$

The linearized least-squares objective function is then given by

$$\mathcal{J} = \frac{1}{2}\pm\mathbf{x}_0^T\mathbf{B}_0^{-1}\pm\mathbf{x}_0 + \frac{1}{2}\sum_{k=0}^{N-1}(\mathbf{H}_k\mathbf{M}_{0;k}\pm\mathbf{x}_0 - \mathbf{d}_k)^T\mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{M}_{0;k}\pm\mathbf{x}_0 - \mathbf{d}_k); \quad (8)$$

where $\pm\mathbf{x}_0 = (\mathbf{x}_0 - \mathbf{x}_0^b)$. Using the compact form of the Jacobian (7), the gradient equation of the linearized problem may be written

$$\begin{aligned}\nabla_{\mathbf{x}_0}\mathcal{J} &= \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \\ &\quad + \sum_{k=0}^{N}(\mathbf{H}_k\mathbf{M}_{0;k})^T\mathbf{R}_k^{-1}(\mathbf{H}_k\mathbf{M}_{0;k}(\mathbf{x}_0 - \mathbf{x}_0^b) - (\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^b)) \\ &= (\mathbf{B}_0^{-1} + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}})(\mathbf{x}_0 - \mathbf{x}_0^b) + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{d}} \\ &= 0; \end{aligned} \quad (9)$$

where $\hat{\mathbf{d}} = (\mathbf{d}_0^T;\mathbf{d}_1^T;:::;\mathbf{d}_N^T)^T$ is the vector of innovations.

The optimal *linear* state-estimate for $\mathbf{x}_0^a$ is then the solution to the gradient equation (9) and is given by

$$\mathbf{x}_0^a = \mathbf{x}_0^b + \hat{\mathbf{K}}\hat{\mathbf{d}}; \quad (10)$$

where

$$\hat{\mathbf{K}} = (\mathbf{B}_0^{-1} + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1} \equiv \mathbf{B}_0\hat{\mathbf{H}}^T(\hat{\mathbf{H}}\mathbf{B}_0\hat{\mathbf{H}}^T + \hat{\mathbf{R}})^{-1}: \quad (11)$$

The matrix $\hat{\mathbf{K}}$ is known as the *gain* matrix.

In addition to the basic statistical assumptions, the errors in the prior

In practice the error distributions may not be Gaussian and the assumptions underlying the estimates derived here may not hold. Ideally, we would like to be able to determine the full probability distributions for the true states of the system given the prior estimates and the observations. This is a major topic of research and new approaches based on sampling methods and particle ¯lters are currently being developed.

Techniques used in practice to solve the data assimilation problem, Problem 1, include sequential assimilation schemes and variational assimilation schemes. These methods are described in the next two sections.

3. Sequential Data Assimilation Schemes

We describe sequential assimilation schemes for discrete models of the form (1), where the observations are related to the states by the equations (2). We make the *perfect model assumption* here. We assume that at some time t_k, *prior* background estimates $\mathbf{x}_k^b$ for the states are known. The di®erences between the observations of the true states and the observations predicted by the background states at this time, $(\mathbf{y}_k - \mathcal{H}(\mathbf{x}_k^b))$, known as the innovations, are then used to make a correction to the background state vector in order to obtain improved estimates $\mathbf{x}_k^a$, known as the analysis states. The model is then evolved forward from the analysis states to the next time t_{k+1} where observations are available. The evolved states of the system at the time t_{k+1} become the background (or forecast) states and are denoted by $\mathbf{x}_{k+1}^b$. The background is then corrected to obtain an analysis at this time and the process is repeated.

Mathematically this procedure may be written

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}_k(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^b)); \tag{16}$$

$$\mathbf{x}_{k+1}^b = \mathcal{M}_{k;k+1}(\mathbf{x}_k^a): \tag{17}$$

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

The system (16)-(17) forms a modi¯ed dynamical system for the analysis states that can be written

$$\mathbf{x}_{k+1}^a = \mathcal{M}_{k;k+1}(\mathbf{x}_k^a) - \propto_{+1} \quad \mathcal{M}_{k;k+1}(\mathbf{x}_{k\text{-}376\ (\ po24.5est21b03.6243al7\ (\ ;k\ )\ ]TJno24.5est21b03.624\ )\ ]TJTf6.624.5TD}^a \quad \mathcal{J} \in \quad \mathcal{J}\propto \in \quad \{$$

its Jacobian, given by the matrix $(M_{k;k+1} + K_{k+1}H_{k+1}M_{k;k+1})$, where $H_k = \frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}|_{\mathbf{x}_k^a}$ and $M_{k;k+1} = \frac{\partial \mathcal{M}_{k;k+1}}{\partial \mathbf{x}}|_{\mathbf{x}_k^a}$. The choice of the gain matrices K_k, $k = 0, 1, \ldots,$ therefore determines the behaviour of the analysed states over time and this choice characterizes the data assimilation scheme.

3.1. OPTIMAL SEQUENTIAL ASSIMILATION SCHEME

For the `optimal' sequential assimilation scheme, the analysis x_k^a, given by (16), is taken to be the *best linear estimate* of the solution to the least squares assimilation problem

$$\min_{\mathbf{x}}[\frac{1}{2}(x - x_k^b)^T B_k^{-1}(x - x_k^b) + \frac{1}{2}(\mathcal{H}_k(x) - y_k)^T R_k^{-1}(\mathcal{H}_k(x) - y_k)] \quad (19)$$

at time t_k. The gain matrix K_k is then given by

$$K_k = B_k H_k^T (H_k B_k H_k^T + R_k)^{-1}, \quad (20)$$

with $H_k = \frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}|_{\mathbf{x}_k^b}$.

If we assume that the background errors are randomly distributed with mean zero and error covariance matrix

$$B_k = \mathcal{E}((x - x_k^b)(x - x_k^b)^T), \quad (21)$$

then the optimal analysis is equal to the BLUE, or best linear unbiased estimate, and minimizes the analysis error variance given, at the optimum, by

$$A_k \equiv \mathcal{E}((x - x_k^a)(x - x_k^a)^T) = (I_n - K_k H_k)B_k. \quad (22)$$

If the random background error vector has a Gaussian distribution, then the analysis is the maximum posterior Bayesian estimate. For linear systems, the solution (16),(20) gives the exact optimal analysis, but for nonlinear systems this solution gives only a ¯rst order approximation to the optimal due to the linearization (x)24TD[(data(x)24TD[(d1100.91Tf5.16.1473373(to)-371(for)-420o)-37obsererror)pro7303(

steps of the extended Kalman ¯lter assimilation scheme are then given as follows. For $k = 0; 1; \ldots$ ¯nd

$$x_k^a = x_k^b + K_k(y_k - \mathcal{H}_k(x_k^b)); \quad (23)$$

$$\text{where} \quad K_k = B_k H_k^T (H_k B_k H_k^T + R_k)^{-1}; \quad (24)$$

$$A_k = (I - K_k H_k) B_k; \quad (25)$$

$$x_{k+1}^b = \mathcal{M}_{k;k+1}(x_k^a); \quad (26)$$

$$B_{k+1} = M_{k;k+1} A_k M_{k;k+1}^T: \quad (27)$$

For systems where the model and observation operators are linear, the analysis x_N^a produced by the Kalman ¯lter at time t_N is exactly equal to the solution $x_N^a = \mathcal{M}_{0;N}(x_0^a)$ to the least-squares data assimilation problem, Problem 1, at the end of the time window. Furthermore, the analysis states produced by the Kalman ¯lter converge over time to the expected values of the true states. For nonlinear systems, however, the EKF only gives approximations to the optimal solution and the EKF may even become unstable as a dynamical system. The EKF is also sensitive to computational round-o® errors (Bierman, 1977).

For large geophysical and environmental systems the extended Kalman ¯lter is, in any case, impractical to implement due to the size of the covariance matrices that need to be propagated. For example, for global weather and ocean systems, the EKF requires the computation of matrices containing of the order of 10^{14} elements at every time step, making it computationally much too expensive to use for real-time state estimation.

The second di±culty in implementing the optimal assimilation scheme (16),(20) sequentially is that in order to compute the analysis x_k^a at each time step, we must ¯nd $B_k H_k^T w_k^a$, where w_k^a solves the linear equations

$$(H_k B_k H_k^T + R_k) w_k^a = (y_k - \mathcal{H}_k(x_k^b)): \quad (28)$$

This is a very large inverse problem with $O(10^5 - 10^6)$ variables to ¯nd. Moreover, the solution may be sensitive to small errors in the data if the matrix $(H_k B_k H_k^T + R_k)$ is ill-conditioned.

In practice most operational sequential assimilation schemes avoid these two di±culties by using approximations that can be implemented e±ciently. A summary of these methods is given in the next subsection.

3.2. PRACTICAL IMPLEMENTATION

A variety of sequential data assimilation schemes have been developed for practical implementation. These di®er mainly in the detailed steps of the procedures. Sequential assimilation schemes used operationally include (Nichols, 2003a):

{ *Successive Correction.* In these schemes, the feedback gain $\mathbf{K}_k$ is not chosen optimally, but is designed to smooth observations into the states at all spatial grid points within some radius of in°uence of each observation (Bergthorsen and Döös, 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 e®ective in data sparse regions.

{ *Optimal Interpolation or Statistical Interpolation.* These schemes approximate the optimal solution by replacing the background error covariance matrix $\mathbf{B}_k$ by a constant matrix $\bar{\mathbf{B}}$, which has a `¯xed' structure for all k. The gain matrix $\mathbf{K}_k$ in (16) is then taken to be

$$\mathbf{K}_k = \bar{\mathbf{B}}\mathbf{H}_k^T(\mathbf{H}_k\bar{\mathbf{B}}\mathbf{H}_k^T + \mathbf{R}_k)^{-1}: \tag{29}$$

(See Ghil and Malanotte-Rizzoli, 1991.) The matrix $\bar{\mathbf{B}}$ is generally de¯ned by an isotropic correlation function (dependent only on the distance between spatial grid points and observational points), with the correlation

{ *3DPSAS and 3D-Representer.* In these schemes iterative minimization methods are applied to the dual variational problem

$$\min_{\mathbf{w}}[\frac{1}{2}(\mathbf{w}^T \mathbf{H}_k \mathbf{B} \mathbf{H}_k^T + \mathbf{R}_k)\mathbf{w} - \mathbf{w}^T(\mathcal{H}_k(\mathbf{x}) - \mathbf{y}_k)]:$$

The iterates converge to the solution $\mathbf{w}_k^a$ of the system (28) with $\mathbf{B}_k$ replaced by $\mathbf{B}$. The resulting analysis states converge to $\mathbf{x}_k^a = \mathbf{B}\mathbf{H}_k^T\mathbf{w}_k^a$, which approximates the `optimal' solution to the variational problem (19), as in the 3DVAR scheme (Cohn *et al.*, 1998; Daley and Barker, 2001). 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 ¯xed over time and by simplifying the inversion problem and/or solving the inversion iteratively. Examples illustrating the application of these schemes to simpli¯ed models can be found in Martin *et al.*, (1999) and on the website of the Data Assimilation Research Centre at http://darc.nerc.ac.uk/ .

3.3. ENSEMBLE FILTERS AND SAMPLING METHODS

Newer approaches to sequential data assimilation known as *ensemble ¯lter* methods, based on classical Kalman or square-root ¯ltering, have recently received much attention. These methods use reduced rank estimation techniques to approximate the classical ¯lters and make the implementation feasible in real time. With these methods an ensemble consisting of a small number of analysis vectors (much less than the number of states n) is propagated simultaneously by the nonlinear model from one observation time to the next in order to provide an ensemble of background states. The background ensemble is updated with the observations to give a new ensemble of analysis vectors and the `optimal' analysis state and its error covariance matrix are determined using a ¯lter similar to the classical ¯lters. An advantage of these methods is that the model and observation operators are not approximated linearly. The accuracy of the estimated states depends, however, on the spread of the ensemble, which must be su±cient to capture the true behaviour of the system.

There are many variants of this technique under development; see, for example, Anderson (2001); Bishop *et al.*, (2001); Burgers *et al.*, (1998); Evensen (2003); Houtekamer and Mitchell (1998); Nerger

these methods retain the advantages of the classical Kalman and square-root filters while remaining feasible for application to large systems. Details of these techniques are described in a later chapter (Kalnay, this book).

Sampling and particle filter methods aim to determine the full probability distributions for the true states of the system. These methods allow for non-Gaussian behaviour of the errors in the prior estimates and the observations and are closely related to the ensemble methods; see for example, Anderson and Anderson (1999); Pham, (2001); Kim *et al.*, (2003); van Leeuwen, (2003); Apte *et al.*, (2007). Although these methods are not yet

tion 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, used in 4DPSAS and 4D-Representer methods, in which the minimization is performed in observation space, is also possible (Courtier, 1997; Xu *et al.*, 2005; Rosmond and Xu, 2006). In these schemes, as in the three dimensional 3DPSAS and 3D-Representer methods, a dual four-dimensional variational problem is solved using a gradient iteration method, and the analysis states are then reconstructed from the dual variables.

The primary difficulty in implementing variational assimilation schemes is the need to develop an adjoint model for the system. The adjoint equations are related theoretically to the linearized state equations, and the system matrix of the adjoint model is given directly by $\mathbf{M}_{k;k+1}^{T}$, where $\mathbf{M}_{k;k+1}$ 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; Giering and Kaminski, 1998). Alternatively an approximate adjoint system can be obtained by discretizing a continuous linear or adjoint model of the nonlinear dynamics (Lawless *et al.*, 2003). This approach has the advantage that additional approximations can be incorporated into the linearization of the system equations.

Other issues arising in the use of variational schemes are the need to cycle the scheme from one analysis time to the next and the length of the window to be used in each cycle. For each new cycle, the initial background weighting, or covariance, matrix $\mathbf{B}_0$ should depend on the current best estimate of the state, which is taken to be the optimal solution of the variational problem at the end of the previous assimilation window. The Hessian of the objective function at the end of the previous cycle can provide this information, but this information is expensive to extract. In practice a climatological or seasonal average is used for the weighting matrix to start each cycle. New research is now becoming available on flow dependent covariance matrices and on longer assimilation windows, in which the initial weighting matrix is expected to have less influence on the analysis (see ECMWF, 2007).

4.2. INCREMENTAL VARIATIONAL METHODS

To make the variational methods more e±cient, an `incremental' approach is generally used in which the nonlinear assimilation problem is replaced by a sequence of approximate linear least-squares problems (Courtier *et al.*, 1994).

At each step i of this method, a linear variational problem is solved to ¯nd an increment $\pm\mathbf{x}_0^{(i)}$ to the current best estimate of the analysis $\mathbf{x}_0^{(i)}$. From the analysis $\mathbf{x}_0^{(i)}$ we solve the nonlinear model equations (1) in order to determine the analysis states $\mathbf{x}_k^{(i)} = \mathcal{M}_{0;k}(\mathbf{x}_0^{(i)})$ and the corresponding innovations $\mathbf{d}_k^{(i)} = \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^{(i)})$ at time t_k. We then linearize the nonlinear assimilation problem about the analysis state trajectory. Initially we set $\mathbf{x}_0^{(0)} = \mathbf{x}_0^b$, for $i = 0$. The linearized variational problem becomes

$$\min_{\pm\mathbf{x}_0^{(i)}} \frac{1}{2}(\pm\mathbf{x}_0^{(i)} - [\mathbf{x}_0^b - \mathbf{x}_0^{(i)}])^{\mathsf{T}}\mathbf{B}_0^{-1}(\pm\mathbf{x}_0^{(i)} - [\mathbf{x}_0^b - \mathbf{x}_0^{(i)}]) + \frac{1}{2}\sum_{k=0}^{N}(\mathbf{H}_k\pm\mathbf{x}_k^{(i)} - \mathbf{d}_k^{(i)})^{\mathsf{T}}\mathbf{R}_k^{-1}(\mathbf{H}_k\pm\mathbf{x}_k^{(i)} - \mathbf{d}_k^{(i)}); \qquad (33)$$

subject to the tangent linear model (TLM) equations

$$\pm\mathbf{x}_{k+1}^{(i)} = \mathbf{M}_{k;k+1}\pm\mathbf{x}_k^{(i)}; \qquad (34)$$

where $\mathbf{M}_{k;k+1} \in \mathbb{R}^{n\times n}$ and $\mathbf{H}_k \in \mathbb{R}^{n\times p_k}$ are linearizations of the operators $\mathcal{M}_{k;k+1}$ and $\mathcal{H}_k$ about the states $\mathbf{x}_k^{(i)}$. A new estimate for the analysis $\mathbf{x}_0^{(i+1)} = \mathbf{x}_0^{(i)} + \pm\mathbf{x}_0^{(i)}$ is obtained by updating the current estimate of the analysis with the solution to the linear variational problem (33) and the process is then repeated.

The linearized problem (33) is solved by an `inner' iteration process. Each inner iteration requires one forward solution of the tangent linear model equations (34), and one backward solution of the corresponding linear adjoint equations to determine the gradient of the objective function. The full incremental variational procedure thus consists of an inner and outer iteration process. In practice, the inner linear least-squares problem is solved only approximately, using a relatively small number of inner iterations, and only a few outer loops of the process are carried out, due to computational time constraints.

The incremental approach is also used in the implementation of the 4D-Representer method (Xu *et al.*, 2005). The dual of the inner linear minimization problem is solved in observation space. The increments in physical space are then reconstructed from the dual variables at the end of the inner iteration and the outer loop is repeated.

Recently the incremental procedure has been shown to be equivalent to an approximate Gauss-Newton method and conditions for its convergence have been established (Lawless *et al.*, 2005; Gratton *et al.*, 2007). Approximations to the tangent linear model and to the corresponding adjoint may be used in the inner iteration without loss of convergence. Furthermore, the inner linear minimization problem does not need to be solved to full accuracy in each outer loop, thus avoiding unnecessary computation. Appropriate stopping criteria for the inner iteration process are presented in Lawless and Nichols (2006).

Additional techniques for increasing the e±ciency of the four-dimensional variational methods are discussed in the next subsections.

4.3. CONTROL VARIABLE TRANSFORMS

In the incremental variational assimilation scheme, transformations of the `control variables' may be applied in order to `decouple' the state variables, to simplify the computational work and to improve the conditioning of the minimization problem. The assimilation problem is written in terms of new variables $\hat{A}_0$, where

$$(x_0 - x_0^b) = U\hat{A}_0 : \tag{35}$$

The transformed linear variational problem (33) becomes

$$\min_{\hat{A}_0} [\frac{1}{2}\| B_0^{-1=2} U\hat{A}_0\|_2^2 + \frac{1}{2}\| \hat{R}^{-1=2}\hat{H}U\hat{A}_0 - \hat{R}^{-1=2}\hat{d}\|_2^2] : \tag{36}$$

where $\hat{H}$, $\hat{R}$ are de¯ned as in (7) and $\hat{d}$ is the vector comprised of the innovations. The conditioning of the optimization problem then depends on the Hessian of the objective function. Transforming the control variables alters the Hessian and changes the convergence properties of the inner iteration of the incremental method. The transformation thus acts as a preconditioner on the inner linearized least-squares problem. The transformation does not, however, a®ect the convergence of the outer loop of the incremental process.

If we choose $U = B_0^{1=2}$, where $B_0^{1=2}$ is the symmetric square root of B_0, the transformed problem (36) takes the form of a classical Tikhonov regularized inverse problem. The Hessian is then given by

$$I + B_0^{1=2}\hat{H}\hat{R}^{-1}\hat{H}B_0^{1=2} ; \tag{37}$$

which is essentially a low-rank update of the identity matrix. The matrix $\hat{R}^{-1=2}\hat{H}B_0^{1=2}$ is the *observability* matrix of the system and is key to the assimilation of information from the observations (Johnson *et al.*, 2005a, 2005b). In the transformed optimization problem (36), the state variables in the background (or regularization) term are weighted by the identity matrix

and thus are decoupled. From a statistical point of view, this means that the transformed variables are uncorrelated, identically distributed random variables. From a practical point of view, the computational work needed in the inversion of the Hessian is simpli¯ed and the inner iteration may be implemented more e±ciently. Additional preconditioners may also be applied to the gradient minimization algorithm in the incremental method to give further increases in the rates of convergence.

Operationally, control variable transforms may be used implicitly to de¯ne the background weighting, or covariance, matrix B_0 in the least-squares formulation of the assimilation problem. A set of control variables is selected that are assumed from physical arguments to be uncorrelated. An appropriate transformation U from these variables to the original variables $(x_0 - x_0^b)$ is then de¯ned and the matrix B_0 is implicitly constructed from this transformation together with information about the spatial autocorrelations of each control variable. By this method additional constraints can be built into the transformations to ensure balance relations hold between the variables, and spectral and other transformations can also be applied implicitly. Flow dependence is also introduced into the weighting

et al., 2008). More e±cient approaches using subspace iteration methods and rational interpolation techniques are currently under development. The latter approaches are promising as they allow for the practical reduction of unstable systems (Boess, 2008; Bunse-Gerstner *et al.*, 2007). E±cient new approximation methods based on proper orthogonal decomposition (POD) have also been developed recently for constructing the optimal projection operators (Willcox and Peraire, 2002).

Other new approaches aim to solve the full nonlinear variational problem in a low dimensional subspace spanned by basis functions generated using POD schemes from control theory or other similar methods. (See Cao *et al.*, 2007, and references therein.) The accuracy and e±ciency of these methods depends on how well the dynamics of the system can be captured in the low dimensional space. Similar techniques, which are adjoint free, have been developed for parameter estimation and model calibration (Vermeulen and Heemink, 2006). Research in this area is currently active.

In summary, four-dimensional variational data assimilation schemes are in operational use at major numerical weather forecasting centres and new theory and new implementation techniques for these schemes continue to be major areas for research. Examples illustrating the use of these schemes on simpli¯ed models can be found in Gri±th (1997) and Lawless, Gratton and Nichols (2005). Tutorial examples are also available on the website of the Data Assimilation Research Centre at http://darc.nerc.ac.uk/ .

5. Data Assimilation for Dynamical Systems with Model Errors

In the previous sections of this chapter, we have made the 'perfect' model assumption that the initial states of the model equations uniquely determine the future states of the system. In practice, however, the nonlinear dynamical model equations describing geophysical and environmental systems do not represent the system behaviour exactly and model errors arise due to lack of resolution (representativity errors) and inaccuracies in physical parameters, boundary conditions and forcing terms. Errors also occur due to discrete approximations and random disturbances. Model errors can be taken into account by treating the model equations as weak constraints in the assimilation problem.

A general least-squares formulation of the data assimilation problem latter appr82(,)9-301(Vting,

technique of state augmentation is developed (Nichols, 2003b) and applications are reviewed.

5.1. LEAST SQUARES FORMULATION FOR MODELS WITH ERRORS

We assume that the evolution of the dynamical system, taking into account model errors, is described by the discrete nonlinear equations

$$\mathbf{x}_{k+1} = \mathcal{M}_{k;k+1}(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k; \quad k = 0;:::;N-1; \tag{38}$$

where $\boldsymbol{\varepsilon}_k \in \mathbb{R}^n$ denotes model errors at time t_k. Prior estimates, or `background' estimates, $\mathbf{x}_0^b$, of the initial states $\mathbf{x}_0$ are assumed to be known and the observations are assumed to be related to the system states by the equations (2).

For the `optimal' analysis, we aim to find the best estimates $\mathbf{x}_k^a$ of the true states of the system, $\mathbf{x}_k$, given observations $\mathbf{y}_k; \; k = 0;:::;N;$ subject to the model equations (38) and prior estimates $\mathbf{x}_0^b$. The `optimal' assimilation problem is written as a weighted nonlinear least-squares problem where the square errors in the model equations, together with the square errors between the model predictions and the observed system states and between the background and initial states are minimized. The data assimilation problem is defined mathematically as follows.

Problem 2 *Minimize, with respect to* $\mathbf{x}_0$ *and* $\boldsymbol{\varepsilon}_k; \; k = 0;:::;N;$ *the objective function*

$$\begin{aligned}\mathcal{J} = & \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \\ & + \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k) + \\ & + \frac{1}{2}\sum_{k=0}^{N}\boldsymbol{\varepsilon}_k^T \mathbf{Q}_k^{-1}\boldsymbol{\varepsilon}_k;\end{aligned} \tag{39}$$

subject to $\mathbf{x}_k; \; k = 1;:::;N;$ *satisfying the system equations (38).*

The model equations (38) are treated here as *weak constraints* on the objective function. The initial states of the system and the model errors at every time step are the control parameters that must be determined.

If we assume that the errors in the prior estimates, in the observations and in the model equations are random variables, then the `optimal' solution to the weakly constrained data assimilation problem, Problem 2, can be interpreted in a statistical sense. We assume that the probability distribution of the random errors $(x_0 - x_0^b)$ between the true initial states and the prior background estimates is Gaussian with mean zero and covariance matrix $B_0 \in \mathbb{R}^{n \times n}$. The observational errors $\pm_k \in \mathbb{R}^{p_k}$, de¯ned in (2), are assumed to be unbiased, serially uncorrelated, Gaussian random vectors with covariance matrices $R_k \in \mathbb{R}^{p_k \times p_k}$. The model errors ^2_k, de¯ned in (38), are also assumed to be randomly distributed variables that are unbiased and serially uncorrelated, with zero means and covariance matrices given by $Q_k \in \mathbb{R}^{n \times n}$. The model errors, the observational errors and the errors in the prior estimates are assumed to be uncorrelated. Under these

the minimization is performed in observation space.

For very large stochastic systems, such as weather and ocean systems, these techniques for treating model errors are not practicable for `real-time' assimilation due to computational constraints. The four-dimensional variational and extended Kalman ¯lter data assimilation schemes are both generally too expensive for operational use due to the enormous cost of estimating all of the model errors in the variational approach or, alternatively, propagating the error covariance matrices in the Kalman ¯lter.

Promising practical approaches to solving the assimilation problem for models with stochastic forcing errors include the sequential ensemble ¯lter methods and the dual variational methods. The ensemble methods take the model errors into account in the low order equations for propagating the ensemble statistics. The dual variational methods solve the assimilation problem in observational space and estimate the model errors implicitly during the reconstruction of the states from the dual variables. Reduced order approaches to solving the variational problem in physical space also allow model errors to be taken into account.

In practice, model errors do not, however, satisfy the statistical assumptions made here. The model error is expected to depend on the model state and hence to be *systematic* and *correlated in time*. A more general form of the model error that includes both systematic and random elements is described in the next subsection.

5.3. SYSTEMATIC MODEL ERROR AND STATE AUGMENTATION

The problem of accounting for systematic model errors in a cost-e®ective way has recently received more attention. Techniques for treating bias errors in the forecast using sequential and four-dimensional variational assimilation schemes (Dee and da Silva, 1998; Derber, 1989; M¶nard, this bookbes and he7 p475ting mo75tiion,

described by the equations

$$z_k = T_k(e_k) + q_k; \tag{40}$$

$$e_{k+1} = \mathcal{G}_{k;k+1}(x_k; e_k); \tag{41}$$

where the vectors $e_k \in \mathbb{R}^r$ represent time-varying systematic components of the model errors and $q_k \in \mathbb{R}^n$ are random errors. The random errors

$$\mathbf{x}_{k+1} = \mathcal{M}_{k;k+1}(\mathbf{x}_k) + T_k(\mathbf{e}_k); \tag{42}$$

$$\mathbf{e}_{k+1} = \mathcal{G}_{k;k+1}(\mathbf{x}_k, \mathbf{e}_k); \tag{43}$$

for $k = 0, \ldots, N-1$, where the observations are related to the model states by the equations (2), as previously. It is assumed that prior estimates, or `background estimates,' $\mathbf{x}_0^b$ and $\mathbf{e}_0^b$ of $\mathbf{x}_0$ and $\mathbf{e}_0$ are known.

The augmented data assimilation problem is to minimize the weighted square errors between the model predictions and the observed system states, over the assimilation interval. The problem is written

Problem 3 *Minimize, with respect to* $(\mathbf{x}_0^T, \mathbf{e}_0^T)^T$, *the objective function*

$$\mathcal{J}$$

Nichols and Bell, 1999; Gri±th and Nichols, 1996, 2000). These techniques have been applied successfully in practice to estimate systematic errors in operational equatorial ocean models (Martin *et al.*, 2001; Bell *et al.*, 2004).

5.4. DATA ASSIMILATION FOR PARAMETER ESTIMATION

Model errors also arise from inaccurate parameters in the model equations. The parameters generally enter the problem nonlinearly, but since the required parameters are constants, the dynamics of the model errors in this case are simple. The error vector is usually also of small dimension relative to the dimension of the state variables. Using augmented forms of the equations, data assimilation can be applied directly to the estimation and calibration of the parameters. The augmented model equations take the form

$$\mathbf{x}_{k+1} = \mathcal{M}_{k;k+1}(\mathbf{x}_k; \mathbf{e}_k); \tag{45}$$

$$\mathbf{e}_{k+1} = \mathbf{e}_k; \tag{46}$$

where the vector $\mathbf{e}_0$ represents the unknown parameters in the model. The estimation problem is then to minimize the objective function (44), subject to the model equations (45){(46).

The standard sequential and variational assimilation schemes can be applied to solve the problem. In the sequential methods, the form of the weighting (or covariance) matrices becomes important due to the nonlinearity of the system equations. On the other hand, in the variational methods, the adjoint equations take a simple form and only the adjoints of the states are needed in order to ¯nd the gradients of the objective function with respect to both the states and the model errors. An application of a sequential scheme to the estimation of parameters in a simpli¯ed morphodynamic model for forecasting coastal bathymetry is described in Smith *et al.* (2008).

In summary, assimilation techniques for estimating random and systematic components of model errors along with the model states are described here. These techniques are e®ective and can lead to signi¯cantly improved forecasts (see Andersson and Th¶epaut, this book). For di®erent types of error, di®erent forms for the model error evolution are appropriate. E±cient methods for taking into account both random and systematic model errors are currently major topics of research.

6. Conclusions

The aims and basic concepts of data assimilation for geophysical and environmental systems are described here. Two approaches to the problem of

data assimilation, sequential and variational 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 the final section of the chapter we discuss data assimilation techniques for treating model errors of both types. Significant approximations are needed in order to implement these methods in `real-time,' due to computational constraints. Further research on data assimilation schemes is needed and there remain many open problems for investigation. Details of current work on data assimilation schemes are given in subsequent chapters of this book.

References

Anderson, J.L. (2001) An ensemble adjustment Kalman filter for data assimilation, *Mon.*

J. for Num. Meth. in Fluids, 53, 1571{1583.
Cohn, S. E., A. da Silva, J. Guo, M. Sienkiewicz and D. Lamich. (1998) Assessing the

Lawless, A.S., S. Gratton and N.K. Nichols. (2005) An investigation of incremental 4D-

Assim. Rev., 237, 112{11279.

Smith, P.J., M.J. Baines, S.L. Dance, N.K. Nichols and T.R. Scott. (2008) Data assimilation for parameter estimation with application to a simple morphodynamic model, University of Reading, Department of Mathematics, Mathematics Report 2/2008.

Talagrand, O. (1981) A study on the dynamics of four-dimensional data assimilation, *Tellus*, 33, 43{60.

Tippett, M.K., J.L. Anderson, C.H. Bishop, T.M. Hamil and J.S. Whitaker. (2003) Ensemble square root ¯lters, *Mon. Wea. Rev.*, 131, 1485{1490.

Tremolet, Y. (2005) Incremental 4D-Var Convergence Study, ECMWF Technical Memorandum, No. 469, European Centre for Medium-Range Weather Forecasts, Reading UK.

van Leeuwen, P. J. (2003) A variance minimizing ¯lter for large scale applications. *Mon. Wea. Rev.*, 131, 2071{2084.

Vermeulen, P.T.M., and A.W. Heemink. (2006) Model-reduced variational data assinilation, *Mon. Wea. Rev.*, 134, 2888{2899.

Weaver, A., and P. Courtier. (2001) Correlation modelling on the sphere using a generalized di®usion equation, *Quart. J. Roy. Met. Soc.*, 127, 1815{1846.