School of Mathematics, Meteorology
and Physics

University of Reading

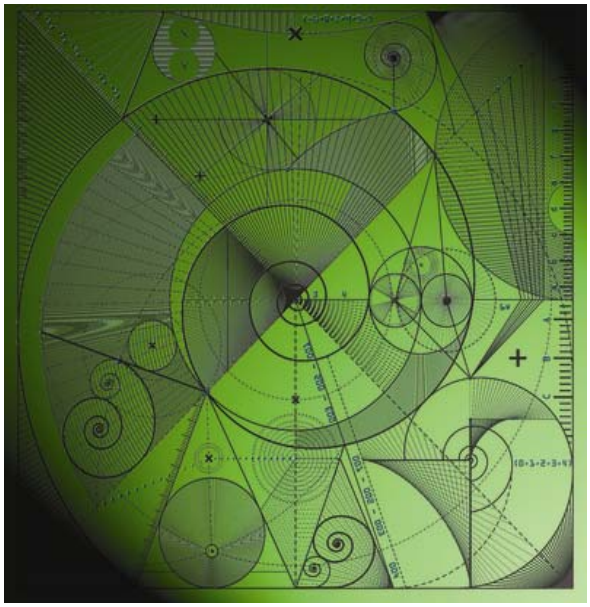
Department of Mathematics

Preprint MPS_2009-04

Mathematical Concepts of Data Assimilation

by

N.K. Nichols



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 defining all of the states of a system at a specific time are, however, rarely available. Moreover, both the models and the available initial data contain inaccuracies and random noise that can lead to significant differences 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 filters 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.

The aim of a data assimilation scheme is to use measured observations in combination with a dynamical system model in order to derive accurate estimates of the current and future states of the system, together with estimates of the uncertainty in the estimated states.

The most significant properties of the data assimilation problem are that the models are very large and nonlinear, with order $O(10^7 - 10^8)$ state variables. The dynamics are multi-scale and often unstable and/or chaotic. The number of observations is also large, of order $O(10^5 - 10^6)$ for a period of 6 hours, but the data are not evenly distributed in time or space and generally have 'holes' where there are no observations (see Lahoz *et al.*, this book). In practice the assimilation problem is generally ill-posed and the state estimates may be sensitive to errors.

There are two basic 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 first case, the observations are 'fed-back' into the model at each time these are available and a best estimate is produced and used to predict future states. In the second case a feasible state trajectory is found that best fits the observed data over a time window, and the estimated states at the end of the window are used to produce the next forecast. Under certain mathematical assumptions these processes solve the same 'optimal' state-estimation problem. In operational systems, solving the 'optimal' problem in 'real-time' is not always possible, and many different approximations to the basic assimilation schemes are employed.

In the next section the data assimilation problem is formulated mathematically. In subsequent sections various techniques for solving the assimilation problem are discussed.

2. Data Assimilation for Nonlinear Dynamical Systems

A variety of models are used to describe systems arising in environmental 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 continuous or discrete models. The data assimilation schemes, with minor modifications, can be applied to any general model.

We begin by assuming that for any given initial states and given inputs, the equations modelling the dynamical system uniquely determine the states of the system at all future times. This is known as the 'perfect' model assumption. In the following subsections we define the data assimi-

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

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

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

$$\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \boldsymbol{\delta}_k, \qquad k = 0, \ldots, N, \tag{2}$$

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 $\boldsymbol{\delta}_k \in \mathbb{R}^{p_k}$ consist of instrumentation errors and representativity errors (see Lahoz *et al.*, this book).

For the 'optimal' analysis, we aim to find the best estimates $\mathbf{x}_k^a$ for the system states $\mathbf{x}_k$, $k = 0, \ldots, N$, to fit the observations $\mathbf{y}_k$, $k = 0, \ldots, N$, 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*

$$\begin{aligned}\mathcal{J} = {} & \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \\ & + \frac{1}{2}\sum_{k=0}^{N}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k)^T \mathbf{R}_k^{-1}(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k),\end{aligned} \tag{3}$$

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

The model is assumed here 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 (1) are uniquely determined by the initial states and therefore can be written explicitly in terms of $\mathbf{x}_0$. Substituting into the objective function (3) then allows the optimization problem to be expressed in terms of the initial states alone. The assimilation problem, Problem 1, thus becomes an unconstrained weighted least squares problem where the initial states are the required control variables in the optimization.

The weighting matrices $\mathbf{B}_0 \in \mathbb{R}^{n\times n}$ and $\mathbf{R}_k \in \mathbb{R}^{p_k\times p_k}$, $k = 0, 1 \ldots, N$, are taken to be symmetric and positive definite and are chosen to give the problem a 'smooth' solution. They represent, respectively, the uncertainty in the background states (prior estimates) and the observations. The objective function (3) can then be written in the compact form:

$$\mathcal{J}(\mathbf{x}_0) \;=\; \frac{1}{2}\| \mathbf{f}(\mathbf{x}_0) \|_2^2 \;\equiv\; \frac{1}{2}\mathbf{f}(\mathbf{x}_0)^{\mathrm{T}}\mathbf{f}(\mathbf{x}_0)\,, \tag{4}$$

where

$$\mathbf{f}(\mathbf{x}_0) = \begin{pmatrix} \mathbf{B}_0^{-1/2}(\mathbf{x}_0 - \mathbf{x}_0^b) \\ \mathbf{R}_0^{-1/2}(\mathcal{H}_0(\mathbf{x}_0) - \mathbf{y}_0) \\ \vdots \\ \mathbf{R}_N^{-1/2}(\mathcal{H}_N(\mathbf{x}_N) - \mathbf{y}_N) \end{pmatrix}, \tag{5}$$

and $\mathbf{x}_k = \mathcal{M}_{0,k}(\mathbf{x}_0)\,,\ k = 1, \ldots, N,$ satisfy the system equations (1) with initial states $\mathbf{x}_0$ at time t_0. (See Lawless, Gratton and Nichols, 2005.) The matrices $\mathbf{B}_0^{-1/2}$ and $\mathbf{R}_k^{-1/2}$ denote the inverses of the symmetric square roots of $\mathbf{B}_0$ and $\mathbf{R}_k$, respectively.

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

2.2. PROPERTIES OF THE BASIC LEAST-SQUARES FORMULATION

The solution $\mathbf{x}_0^a$ to the least-squares problem (4) is known as the *analysis*. The analysis may not be well-defined if $\mathbf{B}_0^{-1} = 0$, that is, if no background state is specified. In that case the number and locations of the observations may not be sufficient to determine all the degrees of freedom in the optimization problem; in other words, the system may not be 'observable.' If

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 written in the compact form

$$\mathbf{J} = \begin{pmatrix} \mathbf{B}_0^{-1/2} \\ \hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}} \end{pmatrix}\,, \quad \hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{0,1} \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{0,N} \end{pmatrix}, \tag{7}$$

where $\hat{\mathbf{R}} = \mathrm{diag}\{\mathbf{R}_k\}$ is a block diagonal matrix containing the weighting matrices $\mathbf{R}_k$ on the diagonal. The matrices $\mathbf{M}_{0,k}$ and $\mathbf{H}_k$ denote the Jacobians of the model and observation operators $\mathcal{M}_{0,k}$ and $\mathcal{H}_k$, respectively; that is,

$$\mathbf{M}_{0,k} = \frac{\partial\mathcal{M}_{0,k}}{\partial\mathbf{x}}|_{\mathbf{x}_0}\,, \qquad \mathbf{H}_k = \frac{\partial\mathcal{H}_k}{\partial\mathbf{x}}|_{\mathcal{M}_{0,k}(\mathbf{x}_0)}\,.$$

If $\mathbf{B}_0$ is non-singular, then the Jacobian $\mathbf{J}$, given by (7), is of full rank and the stationary points satisfying the gradient equation (6) are well-defined. Stationary points are not unique, however, and may not yield a minimum of the nonlinear assimilation problem. If a stationary point is such that the Hessian $\nabla^2_{\mathbf{x}_0}\mathcal{J}$, of the objective function (3) (or equivalently, (4)) is positive-definite at that point, then the stationary point is a local minimum of the assimilation problem (see Gratton *et al.*, 2007). It should be noted that multiple local minima of the assimilation problem may exist.

We remark that the sensitivity of the analysis to small perturbations in the data depends on the 'conditioning' of the Hessian, $\nabla^2_{\mathbf{x}_0}\mathcal{J}$, that is, on the sensitivity of the inverse of the Hessian to small perturbations. If small errors in the Hessian lead to large errors in its inverse, then the computed solution to the data assimilation problem may be very inaccurate. In designing data assimilation schemes, it is important, therefore, to ensure that the conditioning of the Hessian is as small as feasible, or to use 'preconditioning' techniques to improve the conditioning.

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 $\mathbf{d}_k = \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^b)\,,\ k = 0, 1, \ldots, N$, and find 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 $\mathbf{x}_k^b = \mathcal{M}_{0,k}\,(\mathbf{x}_0^b)\,,\ k = 1, \ldots, N$. 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,

$$\mathbf{H}_k = \frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}|_{\mathbf{x}_k^b}, \qquad \mathbf{M}_{0,k} = \frac{\partial \mathcal{M}_{0,k}}{\partial \mathbf{x}}|_{\mathbf{x}_0^b}\,.$$

The linearized least-squares objective function is then given by

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

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

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

where $\hat{\mathbf{d}} = (\mathbf{d}_0^T, \mathbf{d}_1^T, \ldots, \mathbf{d}_N^T)^T$ 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.

For systems where the model and observation operators are linear, the analysis (10)–(11) is an exact, unique, stationary point of the data assimilation problem, Problem 1. For nonlinear systems multiple stationary points of the objective function (3) may exist and the analysis (10)–(11) is only a first order approximation to an optimal solution, due to the linearization of the nonlinear model and observation operators.

The Hessian of the linearized objective function (8) at the analysis (10)–(11) is given by

$$\nabla^2_{\mathbf{x}_0} \tilde{\mathcal{J}} \;=\; (\mathbf{B}_0^{-1} \,+\, \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})\,. \tag{12}$$

If $\mathbf{B}_0$ is nonsingular, then the matrix (12) is symmetric and positive-definite and (10)–(11) provides the 'best' linear estimate of the minimum of the data assimilation problem, Problem 1, in a region of the state space near to the background.

2.4. STATISTICAL INTERPRETATION

The data assimilation problem, as formulated in Problem 1, determines a least-squares fit of the model predictions to the observations, subject to constraints. An estimate of the 'uncertainty' in this analysis would be valuable. If additional assumptions about the stochastic nature of the errors in the initial state estimates and the observations are made, then the solution to the data assimilation problem can be interpreted in statistical terms and the uncertainty in the analysis can be derived.

To obtain a statistical formulation of the data assimilation problem, we assume that the errors $(\mathbf{x}_0 - \mathbf{x}_0^b)$ between the true initial states $\mathbf{x}_0$ and the *prior* background estimates $\mathbf{x}_0^b$ are randomly distributed with mean zero and covariance matrix $\mathbf{B}_0 \in \mathbb{R}^{n\times n}$. The observational errors $\boldsymbol{\delta}_k \in \mathbb{R}^{p_k}$, $k = 0, \ldots, N$, defined in (2), are assumed to be unbiased, serially uncorrelated, randomly distributed vectors with zero means and covariance matrices $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$. The observational errors and the errors in the prior estimates are assumed to be uncorrelated.

Under these basic statistical assumptions, given the prior estimates $\mathbf{x}_0^b$, and the observations $\mathbf{y}_k$, $k = 0, \ldots, N$, the 'best linear unbiased estimate,' or BLUE, of the true state $\mathbf{x}_0$ at time t_0 equals the best least squares estimate (10)–(11) for the analysis $\mathbf{x}_0^a$. The uncertainty in this estimate is described by the analysis error covariance, which is given by

$$\mathbf{A} = (\mathbf{I}_n - \hat{\mathbf{K}}\hat{\mathbf{H}})\mathbf{B}_0\,. \tag{13}$$

Over all linear combinations of the innovations of form (10), the BLUE minimizes the analysis error covariance and is thus the solution to the assimilation problem with *minimum variance*. The analysis given by (10)–(11) is therefore the 'optimal' linear estimate in this sense.

In addition to the basic statistical assumptions, the errors in the prior estimates and in the observations are commonly assumed to have Gaussian probability distributions, which are fully defined by the means and covariances specified. In this case, the solution to the data assimilation problem, Problem 1, is equal to the *maximum a posteriori Bayesian estimate* of the system states at the initial time. From Bayes Theorem we have that the posterior probability of $(\mathbf{x}_0 - \mathbf{x}_0^b)$, given the departures from the observations $(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))$, $k = 0, \ldots, N$, satisfies

$$\begin{aligned}\rho(\mathbf{x}_0 - \mathbf{x}_0^b \,|\, \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k),\ k = 0, \ldots, N) \ = \\ = \alpha\rho(\mathbf{x}_0 - \mathbf{x}_0^b)\rho(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k),\ k = 0, \ldots, N \,|\, \mathbf{x}_0 - \mathbf{x}_0^b)\,, \qquad (14)\end{aligned}$$

where $\rho(\mathbf{x}_0 - \mathbf{x}_0^b)$ is the prior probability of $(\mathbf{x}_0 - \mathbf{x}_0^b)$ and $\rho(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k), k = 0, \ldots, N \,|\, \mathbf{x}_0 - \mathbf{x}_0^b)$ is the conditional joint probability of $(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))$, $k = 0, \ldots, N$, given $(\mathbf{x}_0 - \mathbf{x}_0^b)$. The scalar α is a normalizing constant that ensures that the value of the posterior probability is not greater than unity. The 'optimal' analysis is then the initial state that maximizes the posterior probability.

From the assumption that the probability distributions are Gaussian, we have that

$$\rho(\mathbf{x}_0 - \mathbf{x}_0^b) \ \propto \ \exp[-\frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b)]$$

and

$$\rho(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)) \propto \exp[-\frac{1}{2}(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)]\,,$$

for $k = 0, 1, \ldots, N$. Taking the log of the posterior probability and using the assumptions that the observational errors are uncorrelated in time and uncorrelated with the background errors, we find that

$$\begin{aligned}\mathcal{J}(\mathbf{x}_0) \ &\equiv \ -\ln[\rho(\mathbf{x}_0 - \mathbf{x}_0^b \,|\, \mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k),\ k = 0, \ldots, N)] \\ &= \ -\ln[\rho(\mathbf{x}_0 - \mathbf{x}_0^b)] \ - \sum_{k=0}^{N} \ln[\rho(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))]\,. \qquad (15)\end{aligned}$$

(See Lorenc, 1986, 1988.) The solution $\mathbf{x}_0$ to the data assimilation problem, Problem 1, that minimizes $\mathcal{J}(\mathbf{x}_0)$ is therefore equivalent to the maximum Bayesian a posteriori likelihood estimate.

If the model and observation operators are linear and the errors are normally distributed (*i.e.*, Gaussian), then the *maximum a posteriori Bayesian estimate* and the *minimum variance estimate* are equivalent. The BLUE, given explicitly by (10)–(11), with zero mean and covariance (13), is thus the unique optimal in both senses.

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 filters 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 differences 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 modified dynamical system for the analysis states that can be written

$$\mathbf{x}_{k+1}^a = \mathcal{M}_{k,k+1}(\mathbf{x}_k^a) - \mathbf{K}_{k+1}\mathcal{H}_{k+1}(\mathcal{M}_{k,k+1}(\mathbf{x}_k^a)) + \mathbf{K}_{k+1}\mathbf{y}_{k+1}\,. \tag{18}$$

This system is driven by the observations and has different properties from the original discrete system model (1). The evolution of the analysed states from time t_k to time t_{k+1} is described by a modified nonlinear operator and the response of the system depends generally upon the spectrum of

its Jacobian, given by the matrix $(\mathbf{M}_{k,k+1} + \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{M}_{k,k+1})$, where $\mathbf{H}_k = \frac{\partial \mathcal{H}_k}{\partial \mathbf{x}}|_{\mathbf{x}_k^a}$ and $\mathbf{M}_{k,k+1} = \frac{\partial \mathcal{M}_{k,k+1}}{\partial \mathbf{x}}|_{\mathbf{x}_k^a}$. The choice of the gain matrices $\mathbf{K}_k$, $k = 0, 1, \ldots,$ 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 $\mathbf{x}_k^a$, given by (16), is taken to be the *best linear estimate* of the solution to the least squares assimilation problem

$$\min_{\mathbf{x}}[\frac{1}{2}(\mathbf{x} - \mathbf{x}_k^b)^T\mathbf{B}_k^{-1}(\mathbf{x} - \mathbf{x}_k^b) + \frac{1}{2}(\mathcal{H}_k(\mathbf{x}) - \mathbf{y}_k)^T\mathbf{R}_k^{-1}(\mathcal{H}_k(\mathbf{x}) - \mathbf{y}_k)] \quad (19)$$

at time t_k. The gain matrix $\mathbf{K}_k$ is then given by

$$\mathbf{K}_k = \mathbf{B}_k\mathbf{H}_k^T(\mathbf{H}_k\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (20)$$

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

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

$$\mathbf{B}_k = \mathcal{E}((\mathbf{x} - \mathbf{x}_k^b)(\mathbf{x} - \mathbf{x}_k^b)^T), \quad (21)$$

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

$$\mathbf{A}_k \equiv \mathcal{E}((\mathbf{x} - \mathbf{x}_k^a)(\mathbf{x} - \mathbf{x}_k^a)^T) = (\mathbf{I}_n - \mathbf{K}_k\mathbf{H}_k)\mathbf{B}_k. \quad (22)$$

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 first order approximation to the optimal due to the linearization $\mathbf{H}_k$ of the nonlinear observation operator that is used.

In evolving the 'optimal' BLUE analysis sequentially, two computational difficulties arise. The first is that the background covariance matrices $\mathbf{B}_k$ are required at each time step. These matrices can be propagated forward in time from the initial background error covariance matrix $\mathbf{B}_0$ using an extended Kalman filter (EKF) technique (Kalman, 1961). It is assumed that, at time t_0, *prior* background estimates $\mathbf{x}_0^b$ for the states are known and the errors between the true initial states and the background estimates are randomly distributed with mean zero and error covariance $\mathbf{B}_0$. The

steps of the extended Kalman filter assimilation scheme are then given as follows. For $k = 0, 1, \ldots$ find

$$\begin{aligned}
\mathbf{x}_k^a &= \mathbf{x}_k^b + \mathbf{K}_k(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^b))\,, && (23)\\
\text{where} \quad \mathbf{K}_k &= \mathbf{B}_k\mathbf{H}_k^T(\mathbf{H}_k\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k)^{-1}\,, && (24)\\
\mathbf{A}_k &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{B}_k\,, && (25)\\
\mathbf{x}_{k+1}^b &= \mathcal{M}_{k,k+1}(\mathbf{x}_k^a)\,, && (26)\\
\mathbf{B}_{k+1} &= \mathbf{M}_{k,k+1}\mathbf{A}_k\mathbf{M}_{k,k+1}^T\,. && (27)
\end{aligned}$$

For systems where the model and observation operators are linear, the analysis $\mathbf{x}_N^a$ produced by the Kalman filter at time t_N is exactly equal to the solution $\mathbf{x}_N^a = \mathcal{M}_{0,N}(\mathbf{x}_0^a)$ to the least-squares data assimilation problem, Problem 1, at the end of the time window. Furthermore, the analysis states produced by the Kalman filter 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-off errors (Bierman, 1977).

For large geophysical and environmental systems the extended Kalman filter 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 difficulty in implementing the optimal assimilation scheme (16),(20) sequentially is that in order to compute the analysis $\mathbf{x}_k^a$ at each time step, we must find $\mathbf{B}_k\mathbf{H}_k^T\mathbf{w}_k^a$, where $\mathbf{w}_k^a$ solves the linear equations

$$(\mathbf{H}_k\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k)\mathbf{w}_k^a = (\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^b))\,. \qquad (28)$$

This is a very large inverse problem with $\mathrm{O}(10^5 - 10^6)$ variables to find. Moreover, the solution may be sensitive to small errors in the data if the matrix $(\mathbf{H}_k\mathbf{B}_k\mathbf{H}_k^T + \mathbf{R}_k)$ is ill-conditioned.

In practice most operational sequential assimilation schemes avoid these two difficulties by using approximations that can be implemented efficiently. 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 differ 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 influence 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 effective 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 $\tilde{\mathbf{B}}$, which has a 'fixed' structure for all k. The gain matrix $\mathbf{K}_k$ in (16) is then taken to be

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

(See Ghil and Malanotte-Rizzoli, 1991.) The matrix $\tilde{\mathbf{B}}$ 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 matrix (20) by the gain matrix (29), as in the optimal interpolation scheme (Bratseth, 1986; Lorenc *et al.*, 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 (19) (Rabier *et al.*, 1993). The covariance matrix $\mathbf{B}_k$ is replaced by the approximation $\tilde{\mathbf{B}}$, as defined for optimal interpolation. The solution converges to the analysis obtained by replacing the optimal gain (20) by (29) in (16). 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 and 3D-Representer.*** In these schemes iterative minimization methods are applied to the dual variational problem

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

The iterates converge to the solution $\mathbf{w}_k^a$ of the system (28) with $\mathbf{B}_k$ replaced by $\tilde{\mathbf{B}}$. The resulting analysis states converge to $\mathbf{x}_k^a = \tilde{\mathbf{B}}\mathbf{H}_k^T\mathbf{w}_k^a$, 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 fixed over time and by simplifying the inversion problem and/or solving the inversion iteratively. Examples illustrating the application of these schemes to simplified 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 filter*** methods, based on classical Kalman or square-root filtering, have recently received much attention. These methods use reduced rank estimation techniques to approximate the classical filters 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 filter similar to the classical filters. 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 sufficient 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 *et al.*, (2005); Ott *et al.* (2004); Tippett *et al.* (2003); Zupanski (2005). Although the implementations may suffer from some difficulties (Livings *et al.*, 2008),

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 efficient for very large geophysical problems, these approaches are promising and provide new directions for research.

4. Four-Dimensional Variational Assimilation Schemes

The least-squares data assimilation problem, Problem 1, is currently treated in many operational centres using four-dimensional variational schemes (4DVAR) (Sasaki, 1970; Talagrand, 1981; Rabier *et al.*, 2000; Talagrand, this book). In these schemes the constrained minimization problem, Problem 1, is solved iteratively by a gradient optimization method where the gradients are determined using an adjoint method.

4.1. 4DVAR AND THE ADJOINT METHOD

To solve the least-squares assimilation problem iteratively, the constrained problem, Problem 1, is first written as 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 (1) must be satisfied. The adjoint equations are given by

$$\boldsymbol{\lambda}_{N+1} = 0\,, \tag{30}$$

$$\boldsymbol{\lambda}_k = \mathbf{M}_{k,k+1}^T \boldsymbol{\lambda}_{k+1} + \mathbf{H}_k^T \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k))\,, \quad k = N-1, \ldots, 0, \tag{31}$$

where $\boldsymbol{\lambda}_k \in \mathbb{R}^n$, $k = 0, \ldots, N$, are the adjoint variables and $\mathbf{M}_{k,k+1} \in \mathbb{R}^{n \times n}$ and $\mathbf{H}_k \in \mathbb{R}^{n \times p_k}$ are the Jacobians of $\mathcal{M}_{k,k+1}$ and $\mathcal{H}_k$ with respect to $\mathbf{x}_k$. The adjoint variables $\boldsymbol{\lambda}_k$ measure the sensitivity of the objective function (3) to changes in the solutions $\mathbf{x}_k$ of the state equations for each value of k.

The gradient of the objective function (3) with respect to the initial data $\mathbf{x}_0$ is then given by

$$\nabla_{\mathbf{x}_0} \mathcal{J} = \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) - \boldsymbol{\lambda}_0\,. \tag{32}$$

At the optimum, the gradient (32) is required to be equal to zero. Otherwise this gradient provides the local descent direction needed in the itera-

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).

To make the variational methods more efficient, 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 find an increment $\delta\mathbf{x}_0^{(i)}$ 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

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

subject to the tangent linear model (TLM) equations

$$\delta\mathbf{x}_{k+1}^{(i)} = \mathbf{M}_{k,k+1}\delta\mathbf{x}_k^{(i)}\,, \tag{34}$$

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 $\mathbf{x}_0^{(i+1)} = \mathbf{x}_0^{(i)} + \delta\mathbf{x}_0^{(i)}$ 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 efficiency 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 $\boldsymbol{\chi}_0$, where

$$(\mathbf{x}_0 - \mathbf{x}_0^b) = \mathbf{U}\boldsymbol{\chi}_0 \,. \tag{35}$$

The transformed linear variational problem (33) becomes

$$\min_{\boldsymbol{\chi}_0} [\, \frac{1}{2}\|\mathbf{B}_0^{-1/2}\mathbf{U}\boldsymbol{\chi}_0\|_2^2 \,+\, \frac{1}{2}\|\hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}}\mathbf{U}\boldsymbol{\chi}_0 - \hat{\mathbf{R}}^{-1/2}\hat{\mathbf{d}}\|_2^2 \,] \,. \tag{36}$$

where $\hat{\mathbf{H}}$, $\hat{\mathbf{R}}$ are defined as in (7) and $\hat{\mathbf{d}}$ 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, affect the convergence of the outer loop of the incremental process.

If we choose $\mathbf{U} = \mathbf{B}_0^{1/2}$, where $\mathbf{B}_0^{1/2}$ is the symmetric square root of $\mathbf{B}_0$, the transformed problem (36) takes the form of a classical Tikhonov regularized inverse problem. The Hessian is then given by

$$\mathbf{I} + \mathbf{B}_0^{1/2}\hat{\mathbf{H}}\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}}\mathbf{B}_0^{1/2} \,, \tag{37}$$

which is essentially a low-rank update of the identity matrix. The matrix $\hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}}\mathbf{B}_0^{1/2}$ 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 simplified and the inner iteration may be implemented more efficiently. 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 define the background weighting, or covariance, matrix $\mathbf{B}_0$ 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 $\mathbf{U}$ from these variables to the original variables $(\mathbf{x}_0 - \mathbf{x}_0^b)$ is then defined and the matrix $\mathbf{B}_0$ 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 matrices by this technique. The validity of this approach depends, however, on the assumption that the transformed control variables $\boldsymbol{\chi}_0$ are truly decoupled, or uncorrelated. (See, for example, Bannister *et al.*, 2008; Katz, 2007; Wlasak *et al.*, 2006; Cullen, 2003; Weaver and Courtier, 2001.) Good choices for the control variables and appropriate preconditioners for the gradient minimization algorithms continue to be major topics of research and development.

4.4. MODEL REDUCTION

To increase the efficiency of the incremental methods further, the inner linear minimization problem is often approximated using low dimensional models. The simplest approach is to obtain the increments using a low-resolution linearized model for the dynamical system on a coarse grid. A prolongation operator is then used to map the low-resolution increments to the high-resolution model. Different resolutions can be used at each outer iteration of the procedure, leading to a multi-level approach (Y. Tremolet, 2005; G. Radnoti *et al.*, 2005). These methods are now applied in practice, but theory to support their use is needed.

An alternative technique uses projection operators determined by methods from control theory to produce 'optimal' reduced order models that most accurately capture the response of the full dynamic system. This approach allows much smaller system models to be used for the same computational accuracy, but currently these are expensive to derive (Lawless

et al., 2008). More efficient 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). Efficient 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 efficiency 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 simplified models can be found in Griffith (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 for systems with model errors is introduced in this section. A statistical interpretation of the problem is presented and techniques for solving the assimilation problem for models with random forcing errors are discussed. In reality, model errors are comprised of both systematic and random components. A framework for treating both types of model error using the

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

$$\mathbf{x}_{k+1} = \mathcal{M}_{k,k+1}(\mathbf{x}_k) + \boldsymbol{\epsilon}_k, \quad k = 0, \ldots, N-1, \tag{38}$$

where $\boldsymbol{\epsilon}_k \in \mathbb{R}^n$ 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 $\mathbf{y}_k$, $k = 0, \ldots, N$, 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* $\boldsymbol{\epsilon}_k$, $k = 0, \ldots, N$, *the objective function*

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

subject to $\mathbf{x}_k$, $k = 1, \ldots, N$, *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. The weighting matrices $\mathbf{B}_0 \in \mathbb{R}^{n\times n}$ and $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$, $\mathbf{Q}_k \in \mathbb{R}^{n\times n}$, $k = 0, 1 \ldots, N$, are taken to be symmetric and positive definite and are chosen to give the problem a 'smooth' solution. The choices of the weights should reflect the relative confidence in the accuracy of the background, the observations and the model dynamics and also the structure of the model errors over the time window of the assimilation.

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 $(\mathbf{x}_0 - \mathbf{x}_0^b)$ between the true initial states and the prior background estimates is Gaussian with mean zero and covariance matrix $\mathbf{B}_0 \in \mathbb{R}^{n \times n}$. The observational errors $\boldsymbol{\delta}_k \in \mathbb{R}^{p_k}$, defined in (2), are assumed to be unbiased, serially uncorrelated, Gaussian random vectors with covariance matrices $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$. The model errors $\boldsymbol{\epsilon}_k$, defined in (38), are also assumed to be randomly distributed variables that are unbiased and serially uncorrelated, with zero means and covariance matrices given by $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$. The model errors, the observational errors and the errors in the prior estimates are assumed to be uncorrelated. Under these statistical assumptions, the optimal analysis $\mathbf{x}_0$ that solves the data assimilation problem, Problem 2, is equal to the *maximum a posteriori Bayesian estimate* of the system states, given the observations and the prior estimates of the initial states.

5.2. OPTIMAL SOLUTION OF THE ASSIMILATION PROBLEM

In order to find the 'optimal' analysis that solves Problem 2, either sequential schemes that use the extended Kalman filter (EKF) or variational schemes that solve the minimization problem iteratively can be applied. The EKF propagates the analysis and the covariance matrices forward together, taking into account the model error statistics, in order to produce the 'optimal' linear unbiased state estimate at each time step, conditional on the current and all previous observations. For linear models, the solution obtained using the EKF is the exact optimal and is equal to the solution to the assimilation problem at the end of the time period. For nonlinear systems, approximate linearizations of the model and observation operators are introduced in the extended filter, and the optimality property is not retained.

Variational techniques, in contrast, solve the optimal assimilation problem, Problem 2, for all the analysis states in the assimilation window simultaneously. A direct gradient iterative minimization procedure is applied to the objective function (39), where the descent directions are determined from the associated adjoint equations. The full set of adjoint equations provides gradients of the objective function with respect to the initial states and with respect to all of the model errors at each time step. A forward solve of the model equations, followed by a reverse solve of the adjoint equations is needed to determine the gradients. Alternatively, the optimal assimilation problem can be solved by a dual variational approach in which

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 filter 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 filter.

Promising practical approaches to solving the assimilation problem for models with stochastic forcing errors include the sequential ensemble filter 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-effective 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; Ménard, this book) and for treating time-correlated stochastic errors (Zupanski, 1997) have been investigated. A general formulation for the treatment of systematic model errors has also been derived (Griffith and Nichols, 1996).

We present here a framework for treating systematic, time-correlated model errors based on the formulation of Griffith and Nichols (1996, 2000). Simple assumptions about the evolution of the errors are made, enabling the systematic error to be estimated as part of the assimilation process. The model equations are augmented by the evolution equations for the error and standard data assimilation techniques can then be applied to the augmented state system.

To take into account the systematic components of the model errors, we assume that the evolution of the errors in the model equations (38) is

described by the equations

$$\boldsymbol{\epsilon}_k = T_k(\mathbf{e}_k) + \mathbf{q}_k \,, \tag{40}$$

$$\mathbf{e}_{k+1} = \mathcal{G}_{k,k+1}(\mathbf{x}_k, \mathbf{e}_k) \,, \tag{41}$$

where the vectors $\mathbf{e}_k \in \mathbb{R}^r$ represent time-varying systematic components of the model errors and $\mathbf{q}_k \in \mathbb{R}^n$ are random errors. The random errors are commonly assumed to be unbiased, serially uncorrelated, and normally distributed with known covariances. The effect of the systematic errors on the model equations is defined by the operators $T_k : \mathbb{R}^r \to \mathbb{R}^n$. The operators $\mathcal{G}_{k,k+1} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^r$, describing the systematic error dynamics, are to be specified. The evolution of the errors may depend on the current states of the system.

In practice little is known about the form of the model errors and a simple form for the error evolution that reflects any available knowledge needs to be prescribed. The most common assumption is that the errors constitute a constant bias in each of the model equations. In this case the evolution of the errors is given by $\mathbf{e}_{k+1} = \mathbf{e}_k$, with $T_k = I$. Other forms include linearly evolving error and spectral forms varying on a given time-scale (see Griffith, 1997; Griffith and Nichols, 2000). These forms are expected to be appropriate, respectively, for representing average errors in source terms or in boundary conditions, for representing discretization error in models that approximate continuous dynamical processes by discrete-time systems, and for approximating the first order terms in a Fourier or spherical harmonic expansion of the model error.

Together the system equations and the model error equations (38), (40) and (41) constitute an *augmented* state system model. The aim of the data assimilation problem for the augmented system is to estimate the values of the augmented states $(\mathbf{x}_k^T,\ \mathbf{e}_k^T)^T$, for $k = 0, \ldots, N-1$, that best fit the observations, subject to the augmented state equations. Assuming that the errors in the initial states, the observations and the random components of the model errors, are unbiased, normally distributed, serially uncorrelated and uncorrelated with each other, then the solution delivers the maximum *a posteriori* estimate of the augmented system states. Although this formulation takes into account the evolution of the systematic model errors, the data assimilation problem remains intractable for operational use. If, however, the augmented system is treated as a 'perfect' deterministic model, then solving the augmented data assimilation problem becomes feasible. The aim of the data assimilation, in this case, is to estimate the systematic components of the model error simultaneously with the model states.

The 'perfect' augmented system equations are written

$$\mathbf{x}_{k+1} = \mathcal{M}_{k,k+1}(\mathbf{x}_k) + T_k(\mathbf{e}_k)\,, \tag{42}$$
$$\mathbf{e}_{k+1} = \mathcal{G}_{k,k+1}(\mathbf{x}_k, \mathbf{e}_k)\,, \tag{43}$$

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*

$$\begin{aligned}\mathcal{J} = &\frac{1}{2}((\mathbf{x}_0 - \mathbf{x}_0^b)^T,\ (\mathbf{e}_0 - \mathbf{e}_0^b)^T)\mathbf{W}_0^{-1}((\mathbf{x}_0 - \mathbf{x}_0^b)^T,\ (\mathbf{e}_0 - \mathbf{e}_0^b)^T)^T \\ &+ \frac{1}{2}\sum_{k=0}^{N-1} \left(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\right)^T \mathbf{R}_k^{-1} \left(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k)\right)\,,\end{aligned} \tag{44}$$

subject to the augmented system equations (42)–(43).

The augmented system equations (42)–(43) are treated as strong constraints on the problem. The initial values $\mathbf{x}_0$ and $\mathbf{e}_0$ of the model states and model errors completely determine the response of the augmented system and are taken to be the control variables in the optimization. The weighting matrices $\mathbf{W}_0 \in \mathbb{R}^{(n+r)\times(n+r)}$ and $\mathbf{R}_k \in \mathbb{R}^{p_k \times p_k}$, $k = 0, 1, \ldots, N$, are assumed to be symmetric and positive definite. Since the matrix $\mathbf{W}_0$ is nonsingular, the problem is well-posed and may be solved by any of the standard data assimilation schemes described in this Chapter.

We remark that if a sequential method is used, then the initial weighting matrix $\mathbf{W}_0$ must contain cross-variable terms relating the states and the model errors or the observations may have no effect on the error estimates. In the variational methods, the weighting matrices (or covariance matrices) are implicitly propagated and this is not a problem. Furthermore, in the sequential methods, since the error estimates are updated at every observation point, the error estimates may not behave smoothly. The variational method tends to average the analysis updates over time, automatically smoothing the estimates. For the variational methods, however, an additional set of adjoint equations must be solved to determine the gradient of the objective function with respect to the initial model errors $\mathbf{e}_0$.

Various applications of this approach to model error estimation, using both sequential and four-dimensional assimilation methods, are described in the literature for simplified models (Griffith, 1997; Martin, 2001; Martin,

Nichols and Bell, 1999; Griffith 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 find 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 simplified 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 effective and can lead to significantly improved forecasts (see Andersson and Thépaut, this book). For different types of error, different forms for the model error evolution are appropriate. Efficient 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. Wea. Rev.*, **129**, 2884–2903.

Anderson, J. L., and S. L. Anderson. (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741–2758.

Apte, A., M. Hairer, A.M. Stuart and J. Voss. (2007) Sampling the posterior: an approach to non-Gaussian data assimilation, *Physica D*, **230**, 50–64.

Bannister, R.N., D. Katz, M.J.P. Cullen, A.S. Lawless and N.K. Nichols. (2008) Modelling of Forecast Errors in Geophysical Fluid Flows, *Int. J. for Num. Meth. in Fluids*, **56**, 1147–1153. (doi:10.1002/fld.1618.)

Barnett, S. and R.G. Cameron. (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.

Bell, M.J., M.J. Martin and N.K. Nichols. (2004) Assimilation of data into an ocean model with systematic errors near the equator, *Quart. J. of the Roy. Met. Soc.*, **130**, 873-894.

Bergthorsson, P. and B.R. Döös. (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.

Bishop, C.H., B.J. Etherton and S.J. Majumdar. (2001) Adaptive sampling with the ensemble trasform Kalman filter. Part I: Theoretical aspects, *Mon. Wea. Rev.*, **129**, 420–436.

Boess, C. (2008) Using model reduction techniques within the incremental 4D-Var method, PhD Thesis, Fachbereich 3 - Mathematik und Informatik, Universitaet Bremen.

Bratseth, A.M. (1986) Statistical interpolation by means of successive corrections, *Tellus*, **38A**, 439–447.

Bunse-Gerstner, A., D. Kubalinska, G. Vossen and D. Wilczek. (2007) h_2-norm optimal model reduction for large-scale discrete dynamical MIMO systems, Universitaet Bremen, Zentrum fur Technomathematik, Technical Report 07-04.

Burgers, G., P.J. van Leeuwen and G. Evensen. (1998) Analysis scheme in the ensemble Kalman filter, *Mon. Wea. Rev.*, **126**, 1719–1724.

Cao, Y., J. Zhu, I.M. Navon and Z. Luo. (2007) A reduced-order approach to four-dimensional variational data assimilatio using proper orthogonal decomposition, *Int.*

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 effects of data selection with the DAO physical-space statistical analysis system, *Mon. Wea. Rev.*, **126**, 2913–2926.
Courtier, P. (1997) Dual formulation of four-dimensional variational assimilation, *Quart. J. Roy. Met. Soc.*, **123**, 2449–2461.
Courtier, P., J-N. Thépaut and A. Hollingsworth. (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, *Mon. Wea. Rev.*, **87**, 367–374.
Cullen, M.J.P. (2003) Four-dimensional variational data assimilation: A new formulation of the background covariance matrix based on a potential vorticity representation, *Quart. J. Roy. Met. Soc.*, **129**, 2777–2796.
Daley, R. (1993) *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, UK.
Daley, R. and E. Barker. (2001) NAVDAS: Formulation and diagnostics, *Mon. Wea. Rev.*, **129**, 869–883.
Derber, J.C. (1989) A variational continuous assimilation technique, *Mon. Wea. Rev.*, **117**, 2437–2446.
Dee, D.P. and A.M. da Silva. (1998) Data assimilation in the presence of forecast bias, *Quart. J. Roy. Met. Soc.*, **117**, 269–295.
ECMWF. (2007) *Proceedings of the ECMWF Workshop on Flow Dependent Aspects of Data Assimilation* 11-13 June, 2007, ECMWF, UK.
Evensen, G. (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**, 343–367.
Ghil, M. and P. Malanotte-Rizzoli. (1991) Data assimilation in meteorology and oceanography, *Adv. Geophys.*, **33**, 141–266.
Giering, R., and T. Kaminski. (1998) Recipes for Adjoint Code Construction. *ACM Trans. Math. Software*, **24**, 437–474.
Gratton, S., A.S. Lawless and N.K. Nichols. (2007) Approximate Gauss-Newton methods for nonlinear least squares problems, *SIAM J. Optim.*, **18**, 106–132.
Griewank, A. and G.F. Corliss. (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. http://www.reading.ac.uk/maths/research/maths-phdtheses.asp#1997
Griffith, A.K. and N.K. Nichols. (1996) Accounting for model error in data assimilation using adjoint methods, in M. Berz, C. Bischof, G. Corliss and A. Greiwank (eds.), *Computational Differentiation: Techniques, Applications and Tools*, SIAM, Philadelphia, pp. 195–204.
Griffith, A.K. and N.K. Nichols. (2000) Adjoint techniques in data assimilation for estimating model error, *J. of Flow, Turbulence and Combustion*, **65**, 469–488.
Houtekamer, P.L. and H.L. Mitchell. (1998) Data assimilation using an ensemble Kalman filter technique, *Mon. Wea. Rev.*, **126**, 796–811.
Johnson, C., B.J. Hoskins and N.K. Nichols. (2005a) A singular vector perspective of 4-DVar: Filtering and interpolation, *Quart. J. Roy. Met. Soc.*, **131**, 1–20.
Johnson, C., N.K. Nichols and B.J. Hoskins. (2005b) Very large inverse problems in atmosphere and ocean modelling, *Int. J. for Num. Meth. in Fluids*, **47**, 759–771.
Kalman, R.E. (1961) A new approach to linear filtering and prediction problems, *Trans. ASME, Series D*, **83**, 35–44.
Katz, D. (2007) The application of PV-based control variable transformations in variational data assimilation, PhD Thesis, Department of Mathematics, University of Reading. http://www.reading.ac.uk/maths/research/maths-phdtheses.asp#2007
Kim, S., G. L. Eyink, J. M. Restrepo, F. J. Alexander and G. Johnson. (2003) Ensemble filtering for nonlinear dynamics. *Mon. Wea. Rev.*, **131**, 2586–2594.

Lawless, A.S., S. Gratton and N.K. Nichols. (2005) An investigation of incremental 4D-Var using non-tangent linear models, *Quart. J. Roy. Met. Soc.*, **131**, 459–476.

Lawless, A.S., and N.K. Nichols. (2006) Inner loop stopping criteria for incremental four-dimensional variational data assimilation, *Mon. Wea. Rev.*, **134**, 3425–3435.

Lawless, A.S., N.K. Nichols and S.P. Ballard. (2003) A comparison of two methods for developing the linearization of a shallow water model. *Quart. J. Roy. Met. Soc.*, **129**, 1237–1254.

Lawless, A.S., N.K. Nichols, C. Boess and A. Bunse-Gerstner. (2008) Using model reduction methods within incremental four-dimensional variational data assimilation, *Mon. Wea. Rev.*, **136**, 1511–1522.

Livings, D.M., S.L. Dance and N.K. Nichols. (2008) Unbiased Ensemble Square Root Filters, *Physica D: Nonlinear Phenomena*, **237**, 1021–1028.

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., R.S. Bell and B. Macpherson. (1991) The Met. Office analysis correction data assimilation scheme, *Quart. J. Roy. Met. Soc.*, **117**, 59–89.

Martin, M.J. (2001) *Data Assimilation in Ocean Circulation Models with Systematic Errors*, The University of Reading, Department of Mathematics, PhD Thesis. http://www.reading.ac.uk/maths/research/maths-phdtheses.asp#2001

Martin, M.J., M.J. Bell and N.K. Nichols. (2001) Estimation of systematic error in an equatorial ocean model using data assimilation, in M.J. Baines (ed.), *Numerical Methods for Fluid Dynamics VII*, ICFD, Oxford, pp. 423–430.

Martin, M.J., N.K. Nichols and M.J. Bell. (1999) *Treatment of Systematic Errors in Sequential Data Assimilation*, Meteorological Office, Ocean Applications Division, Tech. Note, No. 21.

Nerger, L., W. Hiller and J. Schroeter (2005) A comparison of error subspace Kalman filters, *Tellus*, **57A**, 715–735.

Nichols, N.K. (2003a) Data Assimilation: Aims and Basic Concepts, *Data Assimilation for the Earth System*, (eds. R. Swinbank, V. Shutyaev, W.A. Lahoz), Kluwer Academic, pp. 9–20.

Nichols, N.K. (2003b) Treating Model Error in 3-D and 4-D Data Assimilation, *Data Assimilation for the Earth System*, (eds. R.Swinbank, V. Shutyaev, W.A. Lahoz), Kluwer Academic, pp. 127–135.

Ott, E., B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil and J.A.l Yorke. (2004) A local ensemble Kalman filter for atmospheric data assimilation, *Tellus*, **56A**, 415–428.

Pham, D. T. (2001) Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194–1207.

Rabier, F., P. Courtier, J. Pailleux, O. Talalgrand and D. Vasiljevic. (1993) A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensionsal variational analysis, *Quart. J. Roy. Met. Soc.*, **119**, 845–880.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons. (2000) The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, *Quart. J. Roy. Met. Soc.*, **126**, 1143-1170.

Radnoti, G., Y. Tremolet, E. Andersson, L. Isaksen, E. Holm and M. Janiscova. (2005) Diagnostics of linear and incremental approximations in 4D-Var revisted for highter resolution analysis, ECMWF Technical Memorandum, No. 467, European Centre for Medium-Range Weather Forecasts, Reading UK.

Rosmond, T. and L. Xu. (2006) Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus*, **58A**, 45–58.

Sasaki, Y. (1970) Some basic formulisms on numerical variational analysis, *Mon. Wea. Rev.*, **98**, 875–883.

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 filters, *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 filter 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 diffusion equation, *Quart. J. Roy. Met. Soc.*, **127**, 1815–1846.

Willcox, K., and J. Peraire. (2002) Model reduction via the proper orthogonal decomposition, *AIAA J.*, **40**, 2323–2330.

Wlasak, M.A., N.K. Nichols and I. Roulstone. (2006) Use of potential vorticity for incremental data assimilation, *Quart. J. Roy. Met. Soc.*, **132**, 2867–2886.

Xu, L., T. Rosmond, and R. Daley. (2005) Development of NAVDAS-AR: Formulation and initial tests of the linear problem. *Tellus*, **57A**, 546–559.

Zupanski, D. (1997) A general weak constraint applicable to operational 4DVAR data assimilation systems, *Mon. Wea. Rev.*, **123**, 1112–1127.

Zupanski, M. (2005) Maximum likelihood ensemble filter: Theoretical aspects, *Mon. Wea. Rev.*, **133**, 1710–1726.