University of Reading

Variational data assimilation for very large environmental problems

Book or Report Section

Accepted Version

Lawless, A. ORCID: https://orcid.org/0000-0002-3016-6568 (2013) Variational data assimilation for very large environmental problems. In: Cullen, M., Freitag, M.A., Kindermann, S. and Scheichl, R. (eds.) Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences. Radon Series on Computational and Applied Mathematics, 2 (13). De Gruyter, Berlin, pp. 55-90. ISBN 9783110282269 Available at https://centaur.reading.ac.uk/37707/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.
Published version at: http://www.degruyter.com/view/product/182025

Publisher: De Gruyter

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Variational data assimilation for very large environmental problems

Amos S. Lawless

Abstract. Variational data assimilation is commonly used in environmental forecasting to estimate the current state of the system from a model forecast and observational data. The assimilation problem can be written simply in the form of a nonlinear least squares optimization problem. However the practical solution of the problem in large systems requires many careful choices to be made in the implementation. In this article we present the theory of variational data assimilation and then discuss in detail how it is implemented in practice. Current solutions and open questions are discussed.

Keywords. 3D-Var, 4D-Var, adjoint model, background errors, error covariance, incremental formulation, nested models, observation errors, optimization, reduced order models, tangent linear model, weak-constraint.

AMS classification. 49-02, 65-02, 86-02.

1 Introduction

Data assimilation is the process of combining a numerical model forecast with observational data in order to estimate the current state of a dynamical system. It has been an essential part of numerical weather prediction (NWP) since its beginnings in the 1940s, when it was recognized that errors in the initial model state could rapidly lead to large errors in the forecast. Early data assimilation schemes were based on a simple interpolation between the observations and the model state, with later schemes also taking account of the statistics of the errors in the data. Such schemes included smoothing splines, successive correction, optimal interpolation and analysis correction [82], [85]. The possible use of methods based on variational calculus was proposed by Sasaki [103], [104] in the late 1950s and 1960s, but at the time a practical implementation was not possible. A real breakthrough in the application of variational schemes to NWP came in the late 1980s with a series of papers demonstrating how the problem could be solved using techniques from the theory of optimal control, in particular the use of adjoint equations to calculate the gradient of an objective function, or cost

The author is supported in part by the U.K. Natural Environment Research Council, through the National Centre for Earth Observation.

To appear in *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences* (2013), Eds. Cullen, M.J.P., Freitag, M. A., Kindermann, S., Scheichl, R., Radon Series on Computational and Applied Mathematics 13.

function [77], [107]. This led to a series of papers in which the feasibility of variational data assimilation was studied on a series of different simplified atmospheric models [108], [98], [26], [93] (these experiments usually only included the large-scale atmospheric dynamics and not the subgrid-scale processes of full weather prediction models).

Despite the encouraging results of these experiments variational data assimilation remained impractical for operational use due to the high computational cost. The introduction of the incremental method of variational assimilation in 1994 [27], together with increasing computing power, opened up the possibility of an affordable implementation for operational weather prediction. Over the following decade many weather forecasting centres began to develop variational data assimilation for operational use [99], [84], [100], [42], [43], [61]. At the same time variational data assimilation began to be applied to other applications, such as ocean forecasting [116], [112] and atmospheric chemistry [38].

A common feature of many of these applications is that the size of the state variable being estimated is extremely large. Current numerical weather prediction models may require the initialization of the order of 10^8 variables in order to make a forecast. As computing power increases the spatial resolution of the models tends to increase and hence so does the number of variables being represented. Furthermore the real-time nature of environmental forecasting requires that the data assimilation problem be solved quickly. These two factors imply that when implementing variational data assimilation schemes in practice compromises must be made. Hence it is important to design the algorithms carefully to ensure that as accurate a solution as possible is obtained within the time available. Ideally such design should also include knowledge of the physics of the problem, so that the final solution is physically realistic. In the remainder of this article we will discuss some of the different choices that arise in the implementation of variational data assimilation for very large systems and the practical approaches that have been developed. First we briefly present the mathematical theory of variational data assimilation.

2 Theory of variational data assimilation

We consider a discrete nonlinear dynamical system given by the equation

$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i), \tag{2.1}$$

where $\mathbf{x}_i \in \mathbb{R}^n$ is the state vector at time t_i and $\mathcal{M}_i$ is the nonlinear model operator that propagates the state at time t_i to time t_{i+1} for $i = 0, 1, \ldots, N-1$. We assume that we have imperfect observations $\mathbf{y}_i \in \mathbb{R}^{p_i}$ at times $t_i, i = 0, \ldots, N$ that are related to the system state through the equation

$$\mathbf{y}_i = \mathcal{H}_i(\mathbf{x}_i) + \boldsymbol{\epsilon}_i, \tag{2.2}$$

where $\mathcal{H}_i : \mathbb{R}^n \rightarrow \mathbb{R}^{p_i}$ is known as the observation operator and maps the state vector to observation space. The observation errors $\boldsymbol{\epsilon}_i$ are usually assumed to be unbiased, serially uncorrelated, Gaussian errors with known covariance matrices $\mathbf{R}_i$. For the numerical weather prediction problem the vector $\mathbf{x}_i$ would contain several meteorological variables, such as pressure, temperature and the three-dimensional wind, at each grid point of the model domain. The observation operator $\mathcal{H}_i$ may just be a simple interpolation in space, if the state variable is observed directly. However, it could be a much more complicated nonlinear function of the state. For example, for a satellite radiance measurement the observation operator can include a complex radiative transfer model.

We assume that at the initial time t_0 we have an *a priori* estimate of the state, usually referred to as a *background* field, that we denote $\mathbf{x}^b$. This background field is assumed to have unbiased, Gaussian errors with known covariance matrix $\mathbf{B}$. In practice the background field is usually a short-term forecast of the state from a previous assimilation cycle. The problem of four-dimensional variational data assimilation (4D-Var)[1] is then to find the initial state that minimizes the weighted least squares distance to this background while minimizing the weighted least squares distance of the model trajectory to the observations over the time interval $[t_0, t_N]$. Mathematically we can formulate this as an optimization problem:
Find the state $\mathbf{x}_0^a$ at time t_0 that minimizes the function

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i) \quad (2.3)$$

subject to the states $\mathbf{x}_i$ satisfying the nonlinear dynamical system (2.1). In the case where $N = 0$ there is no model evolution and the scheme is referred to as three-dimensional variational data assimilation (3D-Var). The solution $\mathbf{x}_0^a$ is commonly referred to as the *analysis*. In environmental data assimilation the function $\mathcal{J}(\mathbf{x}_0)$ is usually called the *cost function*, but the terms *objective function* and *penalty function* are often used in other fields.

The minimization problem given by equation (2.3) can be interpreted in a statistical or deterministic sense. From Bayes' theorem it can be shown that $\mathbf{x}_0^a$ gives the maximimum *a posteriori* estimate of the state under the assumptions given [82]. This includes the assumption of Gaussianity of the error statistics for the background field and observations. In practice this assumption may not always hold. For example, for variables that are inherently non-negative, such as humidity in the atmosphere or concentrations in chemical models, Gaussian statistics may not be appropriate. In some cases these errors may be treated by assuming a lognormal distribution and using this to transform to variables whose statistics are Gaussian [13], [41]. Some allowance for non-Gaussian observation errors may also be made using the method of variational

[1] The scheme is referred to as four-dimensional since we usually fit three spatial dimensions in time, with time being the fourth dimension.

quality control, as discussed in section 3.3. Furthermore, nonlinearity in the dynamical model implies that the background errors are likely to be non-Gaussian if the background comes from a forecast whose length is beyond the linearity regime of the model. For this reason in numerical weather prediction the background field is usually from a forecast of only 6 or 12 hours. In some applications, such as the identification of the source of an atmospheric tracer, it may be more appropriate to specify other prior error distributions [12]. The alternative, deterministic interpretation of the minimization problem is to consider the term measuring the fit to the background state as a form of Tikhonov regularization in fitting the observations [65], [29], [90]. Each of these interpretations is able to provide different insights into the practical formulation of the problem.

It is instructive to consider the solution to the 3D-Var problem under the hypothesis that the observation operator $\mathcal{H}_0$ is approximately linear, such that

$$\mathcal{H}_0(\mathbf{x}^b) - \mathcal{H}_0(\mathbf{x}_0) \approx \mathbf{H}_0(\mathbf{x}^b)(\mathbf{x}^b - \mathbf{x}_0) \tag{2.4}$$

where $\mathbf{H}_0(\mathbf{x}^b)$ is the Jacobian of $\mathcal{H}_0$ evaluated at $\mathbf{x}^b$ (This assumption (2.4) is referred to as the tangent linear hypothesis). In this case the minimum value of (2.3) can be written explicitly as

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B}\mathbf{H}_0^T(\mathbf{H}_0\mathbf{B}\mathbf{H}_0^T + \mathbf{R}_0)^{-1}(\mathbf{y}_0 - \mathcal{H}_0(\mathbf{x}^b)). \tag{2.5}$$

This solution is equal to the best linear unbiased estimate (or BLUE). We see then that the analysis increment, defined as the difference between the analysis and the background $\mathbf{x}^a - \mathbf{x}^b$, lies in the range space of the background error covariance matrix $\mathbf{B}$. We return to the implications of this in section 3.2.

The covariance of the analysis error in this case is given by

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}_0^T\mathbf{R}_0^{-1}\mathbf{H}_0)^{-1}. \tag{2.6}$$

We find that for both 3D-Var and 4D-Var this is equal to the inverse of the Hessian of the cost function,

$$\mathbf{A} = (\nabla^2\mathcal{J})^{-1}. \tag{2.7}$$

In general an exact solution cannot be found and the cost function is minimized using iterative numerical methods, such as conjugate gradient or quasi-Newton methods. The use of these methods in data assimilation is discussed in more detail in section 3.4. On each iteration of such methods the value of the cost function and its gradient at the current iterate must be calculated. In order to calculate the gradient of (2.3) with respect to the initial state $\mathbf{x}_0$ we consider the discrete Euler-Lagrange equations. We introduce Lagrange multipliers $\boldsymbol{\lambda}_i$ at time t_i and define the Lagrangian

$$\mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i) = \mathcal{J}(\mathbf{x}_0) + \sum_{i=0}^{N-1} \boldsymbol{\lambda}_{i+1}^T(\mathbf{x}_{i+1} - \mathcal{M}_i(\mathbf{x}_i)). \tag{2.8}$$

Then necessary conditions for a minimum of (2.3) subject to the constraint are found by taking variations of $\mathcal{L}$ with respect to $\boldsymbol{\lambda}_i$ and $\mathbf{x}_i$. The first of these leads to the original nonlinear model equations (2.1), while the latter gives the discrete adjoint equations

$$\boldsymbol{\lambda}_i = \mathbf{M}_i^T \boldsymbol{\lambda}_{i+1} - \mathbf{H}_i^T \mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i) \tag{2.9}$$

for $i = 1, \ldots, N$ with boundary condition $\boldsymbol{\lambda}_{N+1} = 0$, where $\mathbf{H}_i$ and $\mathbf{M}_i$ are the Jacobians of the nonlinear operators $\mathcal{H}_i$ and $\mathcal{M}_i$ with respect to the state variable $\mathbf{x}_i$. In the data assimilation literature these Jacobians are referred to as the *tangent linear operator* and the *tangent linear model* (TLM) and the operators $\mathbf{H}_i^T$ and $\mathbf{M}_i^T$ are the adjoints of the observation operator and the nonlinear model operator. From (2.8) we then have that the gradient of the Lagrangian with respect to the initial state $\mathbf{x}_0$ is given by

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}_0} = -\mathbf{M}_0^T \boldsymbol{\lambda}_1 + \mathbf{H}_0^T \mathbf{R}_0^{-1}(\mathcal{H}_0(\mathbf{x}_0) - \mathbf{y}_0) + \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b). \tag{2.10}$$

From the theory of Lagrange multipliers this is equal to the gradient of function under the constraint, so we can write

$$\nabla \mathcal{J}(\mathbf{x}_0) = -\boldsymbol{\lambda}_0 + \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b), \tag{2.11}$$

where we have introduced the extra variable

$$\boldsymbol{\lambda}_0 = \mathbf{M}_0^T \boldsymbol{\lambda}_1 - \mathbf{H}_0^T \mathbf{R}_0^{-1}(\mathcal{H}_0(\mathbf{x}_0) - \mathbf{y}_0), \tag{2.12}$$

which can be calculated from the adjoint equations (2.9) with $i = 0$. Hence the adjoint equations provide an efficient method for calculating the gradient information needed for the minimization algorithm. Each iteration of a numerical optimization method therefore requires one run of the forward model (2.1) to calculate the value of the cost function and one run of the adjoint model (2.9) to calculate the gradient. This makes 4D-Var very expensive from a computational point of view.

We note that in this derivation we have implictly taken the adjoint with respect to the Euclidean inner product. For a general linear operator $\mathbf{L} : X1 \rightarrow X2$ and inner products $< ., . >_{X1}, < ., . >_{X2}$ in the spaces $X1, X2$ respectively, the adjoint of $\mathbf{L}$ is the operator $\mathbf{L}^* : X2 \rightarrow X1$ such that

$$< \mathbf{L}\mathbf{x}_1, \mathbf{x}_2 >_{X2} = < \mathbf{x}_1, \mathbf{L}^* \mathbf{x}_2 >_{X1} \tag{2.13}$$

for all $\mathbf{x}_1 \in X1, \mathbf{x}2 \in X2$. In the case where the Euclidean inner product is used in both spaces then the adjoint is equal to the transpose operator, which is why we define the transpose matrices $\mathbf{H}_i^T$ and $\mathbf{M}_i^T$ as the adjoint operators. In this case the Lagrange multipliers provide the correct gradient of the cost function with respect to the state vector, but it is difficult to interpret physically what these variables mean. For other applications of adjoint modelling, such as generating initial perturbations for ensembles of forecasts, it may desirable to give a physical interpretation to the

gradients calculated from the Lagrange multipliers. In these applications other inner products may be used, for example based on the energy or enstrophy[2] of the system [95].

2.1 Incremental variational data assimilation

The possibility of implementing variational data assimilation in an operational setting came with the proposal of incremental variational data assimilation [27]. In this formulation the solution to the nonlinear minimization problem (2.3) is approximated by a sequence of minimizations of linear quadratic cost functions. We define $\mathbf{x}_0^{(k)}$ to be the k^{th} estimate to the solution and linearize the cost function (2.3) around the model trajectory forecast from this estimate. The next estimate is then defined by

$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta\mathbf{x}_0^{(k)}, \tag{2.14}$$

where the perturbation $\delta\mathbf{x}_0^{(k)} \in \mathbb{R}^n$ is a solution of the linearised cost function

$$\begin{aligned}\tilde{\mathcal{J}}^{(k)}(\delta\mathbf{x}_0^{(k)}) &= \frac{1}{2}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}])^{\mathrm{T}}\mathbf{B}_0^{-1}(\delta\mathbf{x}_0^{(k)} - [\mathbf{x}^b - \mathbf{x}_0{}^{(k)}]) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)} - \mathbf{d}_i^{(k)}).\end{aligned} \tag{2.15}$$

Here $\mathbf{d}_i^{(k)} = \mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i^{(k)})$, where $\mathbf{x}_i^{(k)}$ is the nonlinear trajectory calculated from the current estimate at the initial time using the nonlinear model equation (2.1). The perturbation $\delta\mathbf{x}_i$ satisfies the linear dynamical equation

$$\delta\mathbf{x}_{i+1} = \mathbf{M}_i\delta\mathbf{x}_i. \tag{2.16}$$

The linearized observation operator $\mathbf{H}_i$ and the tangent linear model operator $\mathbf{M}_i$ are evaluated at the current estimate of the nonlinear trajectory, usually called the linearization state. The minimization (2.15) is referred to as the *inner loop*, while the update of the nonlinear model trajectory $\mathbf{x}_i^{(k)}$ is the *outer loop*. On each iteration of the inner loop the TLM is integrated to calculate the evolution of the perturbation, in order to calculate the cost function (2.15), and the adjoint model is integrated to provide the gradient.

The incremental method was later shown to be equivalent to an inexact Gauss-Newton method applied to the original nonlinear cost function (2.3) [72]. If we consider a general nonlinear least-squares cost function

$$\phi(\mathbf{x}) = \frac{1}{2}\mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{x}) \tag{2.17}$$

with $\mathbf{f}(\mathbf{x}) : \mathbb{R}^n \to \mathbb{R}^p$ and let $\mathbf{J}(\mathbf{x})$ be the Jacobian of $\mathbf{f}(\mathbf{x})$ with respect to $\mathbf{x}$, then the Gauss-Newton method for minimizing ϕ is

[2] In fluid dynamics the enstrophy is defined as the mean square vorticity of the fluid [58, Section 13.4]

Algorithm 2.1 (Gauss-Newton).

step 0 : choose $\mathbf{x}^{(0)}$

step 1 : repeat until convergence

step 1.1 : compute
$$\delta\mathbf{x} = -(\mathbf{J}(\mathbf{x}^{(k)})^T\mathbf{J}(\mathbf{x}^{(k)}))^{-1}\mathbf{J}(\mathbf{x}^{(k)})^T\mathbf{f}(\mathbf{x}^{(k)})$$

step 1.2 : update $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}$.

Sufficient conditions can be found such that the algorithm will converge to a local minimum of (2.17) [34]. Step 1.1 of the algorithm is equivalent to solving the minimization problem

$$\min_{\delta\mathbf{x}} ||\mathbf{J}(\mathbf{x})\delta\mathbf{x} + \mathbf{f}(\mathbf{x})||_2^2. \tag{2.18}$$

If we define

$$\mathbf{f}(\mathbf{x}_0) = -\begin{pmatrix} \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) \\ \mathbf{R}_0^{-1}(H_0[\mathbf{x}_0] - \mathbf{y}_0^o) \\ \vdots \\ \mathbf{R}_N^{-1}(H_N[\mathbf{x}_N] - \mathbf{y}_n^o) \end{pmatrix} \tag{2.19}$$

subject to (2.1), then the general cost function (2.17) is equal to the 4D-Var cost function (2.3). Applying the Gauss-Newton method to solve this problem we find that the inner minimization step (2.18) is equivalent to the linearized cost function (2.15).

An advantage of using this method to solve the nonlinear problem is that each inner minimization problem is linearly quadratic in $\delta\mathbf{x}$. Hence, whereas the nonlinear problem may have multiple minima, the inner problem has a unique solution that can be found efficiently using iterative minimization methods (we discuss these methods further in Section 3.4). Since these minimization methods are usually truncated according to some stopping criterion, the inner step of the Gauss-Newton method is not solved exactly. In this case the outer loop iterations can be shown to be locally convergent under certain conditions, provided that the inner loop minimization is solved to sufficient accuracy [71], [45].

In practice very few outer loop steps are performed. For example, the Met Office in the U.K. perform only one, while the European Centre for Medium-range Weather Forecasts (ECMWF) perform three [100], [39]. As for the fully nonlinear problem, the incremental method can be run as 3D-Var (no model evolution) or 4D-Var (including the model evolution). An alternative formulation that is often implemented is known as 3D-FGAT (First Guess at Appropriate Time). This includes the nonlinear model evolution in the calculation of the vectors $\mathbf{d}_i$, but no evolution is included for the perturbation and the TLM operator $\mathbf{M}_i$ in equation (2.16) is replaced by the identity. This ensures that the observations are compared with the nonlinear trajectory at the correct time, but approximates the perturbation in such a way that no TLM or adjoint model is needed. In this way some of the benefit of 4D-Var can be achieved without too much extra compuational cost [87], [70].

A major advantage of the incremental approach is that the inner loop minimization problem may be solved in a smaller dimensional space than the outer loop step, for example at a lower spatial resolution. In this way the TLM and adjoint model need only be run at the lower resolution on each inner loop iteration, while the linearization trajectory from the nonlinear model is still calculated at the higher resolution on each outer loop. This is discussed further in Section 3.5. The computational savings made by implementing the inner loop in this way made incremental 4D-Var feasible for operational weather and ocean forecasting.

Having presented the basic theory of variational data assimilation we now examine some of the issues that arise in its practical implementation. For the very large systems found in environmental modelling it is not always possible to apply the theory in an intuitive way. Many choices must be made in order to set up and solve the assimilation problem efficiently and compromises must often be made. It is the attention to detail in these choices that can determine the success or otherwise of the data assimilation scheme.

3 Practical implementation

3.1 Model development

The development of a 4D-Var scheme for the large models used in operational weather and ocean forecasting is a huge undertaking. In most cases the nonlinear model code already exists and has been developed over many years. These models are very large pieces of software, with maybe close to one million lines of code. In order to develop an incremental 4D-Var scheme the code for the TLM and adjoint model must first be written. The development of a TLM code and adjoint model code from the source code of a nonlinear model is a fairly automatic procedure. The correct code for the TLM can be found from a linearization of each statement of the nonlinear model source code, based on treating the nonlinear model as a series of arithmetic operations and applying the chain rule. The adjoint model is then found by a line-by-line transpose of the TLM source code in reverse order. This method is known as automatic differentiation. We do not go into details of its application here, but refer the reader to several good introductions in the literature [26], [10], [102], [44]. The automatic nature of this procedure has led to many software tools being developed that will produce a TLM and adjoint model code from a nonlinear mode source code. These automatic differentiation tools, or automatic adjoint compilers, are now available commercially for many different programming languages. [3]

In practice the TLM and adjoint models of many large environmental models have been developed by hand, rather than using the automatic compilers. There are several reasons for this. The first is that in many cases of operational weather and ocean forecasting the complexity of the already exisiting nonlinear model codes was such

[3] The term *automatic differentiation* refers to the approach itself, not just to the automatic tools.

that simple application of the automatic compilers was not possible. In many cases, particularly for large codes developed by many people, it is necessary to tidy the nonlinear model codes to make them suitable for use with the automatic compilers. Many centres felt that the effort to do this would have been greater than coding the TLM and adjoint model by hand.

The second reason for developing the TLM and adjoint codes by hand arises from the nature of the incremental approach to variational data assimilation. Since the TLM and adjoint are run at a lower resolution in the inner loop, the TLM is already an approximate linearization of the nonlinear model used in the outer loop. It is therefore justifiable to make further simplifications in the TLM, in order to reduce the computational cost. As long as the adjoint model is derived from the approximate TLM, then the inner loop minimization will contain the correct gradient information for convergence. In coding the models by hand it is easier to make such simplifications based on physical arguments. For example, many meteorological models contain parametrizations of sub-grid-scale processes (known as the *physics* in the meteorological literature), which include such things as clouds, precipitation and surface drag. The schemes used to represent these processes can be highly complex and often include non-differentiable functions, such as on-off switches. While it is possible for automatic differentiation to deal with such functions it is usually felt that this level of complexity is not necessary in the TLM and adjoint model. Hence a series of simpler parametrizations have been developed solely for use in incremental 4D-Var, that capture the main behaviour of the more complex schemes [118], [64], [99], [88].

An alternative approach, devised by the Met Office, is to start from the premise that the linear model must evolve finite and not infinitesimal perturbations and so there is no need for the linear model to be tangent to any nonlinear model. In this approach the linear model is designed with this in mind. In particular, the resolved dynamics is approximated by a discretization of the linearized continuous equations, with various simplifications in the equations and the discretization. Then simplified parametrizations can be used to represent sub-grid-scale processes [86], [74]. The adjoint model is derived from this approximate linear model by the process of automatic differentiation, ensuring that it provides the exact gradient of the discrete linear cost function.

An essential part of the development of the linear and adjoint models is their testing, as any small mistakes could lead to lack of convergence of the minimization algorithms. Robust tests exist to check the coding of a TLM and adjoint model. The test for the TLM is based on comparing the evolution of a perturbation in the TLM with the evolution of the same perturbation in the nonlinear model. A Taylor series expansion of the nonlinear model operator shows that the evolutions should be closer together as the perturbation size is reduced [98], [79]. Where an inexact TLM is used then this test is not able to differentiate between small coding errors and the desired inexactness. In this case other more subjective tests must be performed [74]. The adjoint model code can be tested by a verification of the adjoint identity (2.13). If we assume that the

spaces $X1$ and $X2$ are both equal to $\mathbb{R}^n$, then we must have

$$< \mathbf{M}_i \delta\mathbf{x}_i, \mathbf{M}_i \delta\mathbf{x}_i > = < \delta\mathbf{x}_i, \mathbf{M}_i^*(\mathbf{M}_i \delta\mathbf{x}_i) >, \tag{3.1}$$

which, in the Euclidean inner product, is equivalent to

$$(\mathbf{M}_i \delta\mathbf{x}_i)^T (\mathbf{M}_i \delta\mathbf{x}_i) = \delta\mathbf{x}_i^T (\mathbf{M}_i^T (\mathbf{M}_i \delta\mathbf{x}_i)). \tag{3.2}$$

This identity can be tested for random perturbations $\delta\mathbf{x}_i$. If the adjoint operator $\mathbf{M}_i^T$ has been correctly coded then this identity will hold to machine precision [93]. For large codes each of these tests should be available for each subroutine as well as at higher levels. A further test, also based on a Taylor expansion, is used to verify that the gradient of the cost function has been correctly coded [93].

3.2 Background error covariances

The background field $\mathbf{x}^b$ is a very important part of practical data assimilation systems in environmental forecasting. Since in many operational forecasting systems the background field is a forecast from a previous assimilation cycle, it contains information from observations assimilated at earlier times. In one of the early 4D-Var systems at ECMWF it was shown that, at any assimilation time, the background field has an approximately 85% influence on the analysis, with the new observations contributing only 15% [24]. The background error covariance matrix $\mathbf{B}$ determines the relative weight between the background field and observations and hence plays an essential role in the data assimilation algorithm. However, the calculation of these covariances for the assimilation system is a hugely complex task and very dependent on the specific system being modelled. Here we are only able to give an outline of the main steps involved. For further details in the context of atmospheric data assimilation the reader is referred to the comprehensive two-part review article of Bannister [6], [7].

As was seen from (2.5) in section 2, under certain simplified assumptions the analysis increment of 3D-Var can be shown to lie in the subspace spanned by the columns of the matrix $\mathbf{B}$. In order to understand the implications of this, we consider the case where we have a single observation y of the k^{th} component of the vector $\mathbf{x}$, with error variance σ_o^2. In this case the observation operator is linear and is given by the k^{th} unit vector $\mathbf{e}_k$ and the analysis equation (2.5) becomes

$$\mathbf{x}^a = \mathbf{x}^b + \begin{pmatrix} b_{1,k} \\ b_{2,k} \\ \vdots \\ b_{N,k} \end{pmatrix} \frac{y - \mathbf{x}^b(k)}{b_{k,k} + \sigma_o^2}, \tag{3.3}$$

where $b_{i,k}, i = 1, \ldots, N$ indicates the (i,k) element of the matrix $\mathbf{B}$ and $\mathbf{x}^b(k)$ is the k^{th} component of $\mathbf{x}^b$. Hence we see that the value of each entry $b_{i,k}$, which is

the covariance between the errors in the components of the background field $\mathbf{x}^b(i)$ and $\mathbf{x}^b(k)$, determines the analysis increment to the i^{th} component of the state given an observation of the k^{th} component. As a consequence the entries of this matrix determine how observations are used to infer information about unobserved parts of the state. Thus this matrix is fundamental in allowing information to be inferred about unobserved physical variables or unobserved regions of space. However, it is usually impossible to represent this matrix in matrix form. If the state vector is of size n then the matrix $\mathbf{B}$ is of size $n \times n$ and when n is of order 10^8 this matrix is impossible to calculate or store. Instead the action of this matrix is usually represented by a variable transform.

We consider the variable transform in the context of incremental variational data assimilation, since that is how it is usually implemented. We define a new variable $\delta\mathbf{z}_i \in \mathbb{R}^n$ and a transformation matrix $\mathbf{U}_i \in \mathbb{R}^{n\times n}$, such that

$$\delta\mathbf{x}_i = \mathbf{U}_i\delta\mathbf{z}_i, \qquad i = 0, \ldots, N. \tag{3.4}$$

In terms of this new variable the incremental cost function (2.15) can be written

$$\begin{aligned}\tilde{\mathcal{J}}^{(k)}(\delta\mathbf{z}_0^{(k)}) &= \frac{1}{2}(\delta\mathbf{z}_0^{(k)} - [\mathbf{z}^b - \mathbf{z}_0{}^{(k)}])^{\mathrm{T}}\mathbf{U}_0^T\mathbf{B}^{-1}\mathbf{U}_0(\delta\mathbf{z}_0^{(k)} - [\mathbf{z}^b - \mathbf{z}_0{}^{(k)}]) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i\mathbf{U}_i\delta\mathbf{z}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{U}_i\delta\mathbf{z}_i^{(k)} - \mathbf{d}_i^{(k)}),\end{aligned} \tag{3.5}$$

If the variables $\delta\mathbf{z}$ are chosen in such a way that they are uncorrelated then they have identity covariance matrix by definition and so $\mathbf{U}_0^T\mathbf{B}^{-1}\mathbf{U}_0$ can be replaced with the identity in the cost function (3.5). In this case the cost function no longer contains the original background error covariance matrix; instead it is implicitly defined through the variable transform, with $\mathbf{B} = \mathbf{U}_0\mathbf{U}_0^T$.

Furthermore, this variable transform is expected to lead to a better conditioned problem. To understand this we note that the Hessian of the original inner loop cost function (2.15) is given by

$$\mathbf{G} = \mathbf{B}^{-1} + \sum_{i=0}^{N}\mathbf{M}(t_i, t_0)^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\mathbf{H}_i\mathbf{M}(t_i, t_0), \tag{3.6}$$

where

$$\mathbf{M}(t_i, t_0) = \mathbf{M}_{i-1}\mathbf{M}_{i-2}\ldots\mathbf{M}_0 \tag{3.7}$$

is the tangent linear model solution operator from time t_0 to time t_i. Equivalently we can write this as

$$\mathbf{G} = \mathbf{B}^{-1} + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}}, \tag{3.8}$$

where

$$\hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}(t_1, t_0) \\ \vdots \\ \mathbf{H}_N\mathbf{M}(t_N, t_0) \end{pmatrix} \tag{3.9}$$

and $\hat{\mathbf{R}}$ is a block diagonal matrix with blocks equal to $\mathbf{R}_i, i = 0, \ldots, N$. If the background error covariance matrix is ill-conditioned, then we expect this to dominate the conditioning of the Hessian $\mathbf{G}$. We return to an examination of this in section 3.4. On the other hand the Hessian of the transformed problem (3.5) is given by

$$\tilde{\mathbf{G}} = \mathbf{I} + \sum_{i=0}^{N} \mathbf{U}_i^T \mathbf{M}(t_i, t_0)^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{M}(t_i, t_0) \mathbf{U}_i. \tag{3.10}$$

Usually the number of observations is less than the number of state variables being estimated and so the Hessian (3.10) is equal to the identity plus a low rank matrix. Then it has a minimum eigenvalue equal to one and the condition number (in the two-norm) is equal to the largest eigenvalue. Thus we would expect the transformed problem to be better conditioned.

Of course this theory all relies on being able to choose appropriate variables $\delta\mathbf{z}$ that are truly uncorrelated and it is here that a knowledge of the physical problem is necessary. In presenting how the transform is designed in practice it is easier to think about it in terms of the inverse transform, from model variables $\delta\mathbf{x}$ to uncorrelated variables $\delta\mathbf{z}$. A common approach in numerical weather prediction is to split the inverse transform into two parts. The first part, which we write $\mathbf{U}_p^{-1}$, is known as the parameter transform and transforms to physical variables $\delta\boldsymbol{\chi}$ whose errors are assumed to be uncorrelated between themselves, but still contain spatial correlations. The spatial transform, $\mathbf{U}_s^{-1}$, then removes spatial correlations between the physical variables $\delta\boldsymbol{\chi}$. We thus have the steps

$$\delta\boldsymbol{\chi} = \mathbf{U}_p^{-1}\delta\mathbf{x} \tag{3.11}$$

$$\delta\mathbf{z} = \mathbf{U}_s^{-1}\delta\boldsymbol{\chi}, \tag{3.12}$$

where for ease of notation we assume the transforms to be time-invariant. In practice the transforms $\mathbf{U}_p$ and $\mathbf{U}_s$ may not be square and a generalization of the inverse operator is needed. We now consider each of these transforms in turn.

Parameter transform

In designing a suitable transform of parameters $\mathbf{U}_p^{-1}$ it is necessary to have an understanding of the particular system being modelled, in order to decide which variables have errors that are likely to be uncorrelated. For atmospheric models the transform

is based on the concept of balanced variables. Balance relationships are diagnostic relationships that exist between certain atmospheric variables. For example, in mid-latitudes and at large horizontal length scales the horizontal wind is approximately in balance with the gradient of the pressure field, through the relationship of geostrophic balance. This relationship can be used in the parameter transform by assuming that errors in the the balanced part of the flow are uncorrelated with those in the unbalanced part [7]. This can be justified by an eigenanalysis of the linearized equation set, which shows that the balanced flow can be associated with one eigenvector and the unbalanced flow with the remaining eigenvectors. Hence under linear evolution these will evolve independently.

The variable that best represents the balanced flow in the atmosphere is potential vorticity (PV) [59] and so it would be natural to use this variable as the basis for the parameter transform. However the transform from PV to the original model variables requires the solution of a three-dimensional elliptic equation as part of the application of the operator $\mathbf{U}_p$. In dynamical regimes with small characteristic horizontal length scales, the PV is well-approximated by the vorticity, which only requires the solution of a two-dimensional equation [117]. Hence early work in this area proposed a transform based on this variable [97] and this is still the basis of the parameter transform in many operational weather forecasting systems [7]. It is recognized that this approximation is not valid in all parts of the atmosphere and it has been demonstrated on simple systems that significant correlations can remain between errors in the transformed variables [66]. For this reason attempts are being made to implement a transformation based on PV in large-scale systems [28], [8].

A similar approach may be followed in other applications, for example in ocean forecasting, though here there has been less work on the design of appropriate transforms than in the meteorological context. In many cases it may be assumed that errors in the model variables such as salinity and temperature are uncorrelated and only the spatial transform is needed [116], but work on defining balance relationships has allowed multi-variate covariances to be introduced [114].

Spatial transform

Once the parameter transform has been performed it is assumed that the errors in the resulting variables are uncorrelated between themselves. At this point it is necessary to specify the autocovariance information for each parameter through the spatial transform. In atmospheric models it is common to assume that the transforms in the horizontal and vertical planes are separable. In most systems a Fourier transform is used in the horizontal and for the vertical correlations a transformation to the eigenvectors of a vertical error covariance matrix is used. The order in which these transformations are performed varies between systems. If the horizontal transform is performed first, then the horizontal spectral modes are assumed to be uncorrelated and vertical correlations are specified separately for each mode. This assumption leads to correlations that are

homogeneous (independent of horizontal position) and isotropic (independent of orientation) in the transformed parameters [7]. The method allows vertical correlations that vary with horizontal scale, so that features with large horizontal scale have deeper vertical correlations [36]. However, it does not allow vertical correlations to vary with horizontal position [62]. The alternative is to first perform the vertical transform and then, assuming that these modes are independent, apply the Fourier transform to each vertical mode. This allows more variation of vertical correlations with horizontal position (for example, with latitude). However it is more difficult to obtain an appropriate variation in horizontal correlation length scales with height [84], [62]. In both cases a scaling transformation is also needed to ensure that the variance of the transformed variables is equal to one. In an ideal case we would like to obtain covariances that depend on both horizontal scale and horizontal position. This has led to the development of spatial transforms based on a wavelet basis [36], [5]. Such a transform has been implemented in the operational NWP system of ECMWF.

In ocean models the complex boundaries near the coast prohibit the simple use of a Fourier transform in the horizontal and so other methods must be used to represent spatial correlations. For example, the application of a correlation operator can be shown to be equivalent to the integration of an appropriately-constructed diffusion equation [113]. This can be used to design correlation models for use in data assimilation systems with irregular boundary conditions [116], [115].

The use of transforms for spatial covariances requires the specification of correlation lengthscales and variances for each of the transformed variables. Since the background field is usually a short-term forecast, these statistics must represent the structure of errors in the forecasting system being used and so be diagnosed from that. An early method for obtaining these statistical parameters used the difference between the observations and the background field (known as the *innovations*) [57]. However, a disadvantage of this method is that it relies on having a sufficient number of observations and is therefore biased towards data-dense areas. The most popular method in atmospheric data assimilation is that known as the 'NMC method' [97][4]. In this method the difference between two different forecasts valid at the same time is taken as a proxy for forecast errors and statistics are taken over a sample of many such forcasts. In atmospheric forecasting usually two forecasts starting 24 hours apart are used, with the earlier one run for 48 hours and the later one for 24 hours. By using an interval of 24 hours problems arising from modelling the diurnal variation of the atmosphere are avoided. However, this means that the differences are taken over a much longer time interval than the normal background forecast, which is usually 6 or 12 hours. As a result the covariance structures of the forecasts differences do not necessarily reflect those of the background error and often they need to be modified for use in the assimilation system [62], [36].

This has motivated the development of ensemble methods to generate statistics from

[4] So called because it was first introduced in the National Meteorological Center of the USA, now the National Center for Environmental Prediction.

shorter forecasts. Such a method for estimating background error statistics from an ensemble of short forecasts was developed for use at ECMWF in [36]. The basis of this method is that if the inputs to the assimilation system (for example, the background, observations and physical boundary conditions) are perturbed within the statistics of their errors, then the perturbation in the resulting analysis will be drawn from the distribution of analysis error. If a short forecast is produced from this analysis, then we expect the perturbation to the forecast to be drawn from the distribution of forecast error. This perturbed forecast can then be used as a background field for the next assimilation time and the process repeated to produce the next analysis and another forecast. Suppose that we run two such cycles in parallel for l cycles, starting from two different sets of perturbations at time t_0. Then at each assimilation time $t_i, i = 1, \ldots, l$ this will produce two perturbed short forecasts $\mathbf{x}_i^{b1}$ and $\mathbf{x}_i^{b2}$. It can be shown that the statistics of the true forecast error can then be calculated from the sample covariance of the differences between these pairs [6],

$$\mathbf{B} \approx \frac{1}{2(l-1)} \sum_{i=1}^{l} (\mathbf{x}_i^{b1} - \mathbf{x}_i^{b2})(\mathbf{x}_i^{b1} - \mathbf{x}_i^{b2})^T, \tag{3.13}$$

under the assumption that the errors in the two forecasts are uncorrelated. The factor of $1/2$ arises since the sample covariance itself is equal to the sum of the error covariances of the two different sets of forecasts. Since the forecasts used in this method are of the same length as the forecasts used to obtain the background field in the assimilation, the error statistics produced in this way are a more realistic representation of the true error statistics.

A key assumption in the methods presented so far is that the error covariance matrix represents a statistical average over time. The computational expense of calculating these statistics means that the matrix is kept constant from day to day, perhaps with different statistics being used with a change of season. More recently there has been interest in developing methods for estimating statistics that vary from day to day, since it is expected that the actual background errors will depend on the underlying flow. Such flow-dependent statistics arise naturally in ensemble methods of data assimilation, such as the ensemble Kalman filter. Methods are currently being designed to obtain some flow-dependent information in variational assimilation, by combining information from ensembles of forecasts with the statistically-averaged error covariance matrix, for example [18], [15].

3.3 Observation errors

As well as representing the errors in the background field it is important to treat properly the errors in the observations within a variational data assimilation system. Observational data received into operational weather and forecasting centres can contain errors from a variety of sources, including limitations in the measuring instrument, biases in the measurements and errors simply due to human error in recording the

measurement. Furthermore other errors arise from the way the data are used within the data assimilation system, both from inaccuracies in the operators used to map the model state to observation space and from the differences in spatial resolution between the model and the observations. The theory of variational data assimilation assumes that all observational errors are random, unbiased errors with a Gaussian distribution and known covariance. It is therefore important that as many of these sources of error as possible are accounted for in the data assimilation system.

A first essential step in an operational data assimilation system is to perform a quality control check on the data themselves. This may consist of several stages. First a check for obvious errors in the reporting of the data is made, such as errors in the reported position. For example, if a ship observation is reported over a land point it will be rejected from the assimilation. Then a so-called 'background check' may be made to see how close the observation is to the forecast background field. If the difference from the background is too large when compared with its expected error variance then the observation may be rejected and not used in the assimilation [2]. Once this check has been performed the next step is to identify observations that may have gross errors, that is errors that are unlikely to satisy the assumption of being random and normally distributed. This can be done either outside or within the assimilation process. Outside the assimilation each observation can be checked against nearby observations and any observations that largely disagree with others can be rejected [100]. Alternatively this check can be included in the assimilation process using the variational quality control method [63], [2]. In this method the probability density function of the observation errors are assumed to be a weighted combination of a standard Gaussian distribution and a flat probability distribution function, with the weights determined by the probability of gross error of the observation. Thus for each single observation y with weight α_y, the probability density function of the observation error is assumed to be of the form

$$\mathcal{P}_{QC} = (1 - \alpha_y)\mathcal{P}_N + \alpha_y \mathcal{P}_F, \tag{3.14}$$

where $\mathcal{P}_N$ indicates the appropriate Gaussian probability density function and $\mathcal{P}_F$ is a flat distribution over a finite interval centred at zero and is equal to zero outside this interval (the size of this interval is taken to be a multiple of the observation error standard deviation). The observation part of the cost function is then taken to be equal to the negative logarithm of $\mathcal{P}_{QC}$. In the case where $\alpha_y = 0$ this corresponds to the observation term in the original nonlinear cost function (2.3). In this method observations that have a high probability of gross error are given very little weight in the analysis. Initially these probabilities are assigned to each observation based on a study of historical data. The probabilities are then updated on each iteration of the minimization procedure by comparison with the current estimate of the state, to allow observations to be given more or less weight as the assimilation progresses. The introduction of non-Gaussianity means that variational quality control can introduce multiple minima into the cost function and so it is necessary to have a good starting point for the minimization. For this reason the minimization is first run for several iterations without the

quality control term before switching it on [1].

A second important aspect of observation errors is the treatment of systematic errors, or biases, in the observations. This is particularly important for satellite radiance data, where biases may occur from changes in the measuring instrument over time or from errors in the radiative transfer model needed as part of the observation operator [54]. Since the assimilation scheme assumes that the observations are unbiased, any biases in the observations can introduce biases into the analyses. As with the quality control these biases may be treated offline or within the assimilation scheme. For each satellite channel a bias model is assumed in such a way that we can define a new observation operator for the biased measurement

$$\tilde{\mathcal{H}}(\mathbf{x}, \boldsymbol{\beta}) = \mathcal{H}(\mathbf{x}) + \mathbf{b}(\boldsymbol{\beta}, \mathbf{x}), \tag{3.15}$$

with

$$\mathbf{b}(\boldsymbol{\beta}, \mathbf{x}) = \sum_{j=0}^{N_p} \beta_j \mathbf{p}_j(\mathbf{x}), \tag{3.16}$$

where $\mathbf{p}_j$ are predictors for $j = 0, \ldots, N_p$ and β_j are scalar coefficients [33]. A few predictor states are chosen that may be related to the state at the observation positions. The coefficients $\boldsymbol{\beta}$ can then be estimated in an offline regression using a few weeks of data [54] or a variational procedure can be used to estimate these coefficients. This can be included directly in the assimilation procedure by including (3.15) in the cost function in place of the standard observation operator and including a background estimate $\boldsymbol{\beta}^b$ of $\boldsymbol{\beta}$ with covariance $\mathbf{B}_{\boldsymbol{\beta}}$. The 4D-Var assimilation problem is then to minimize

$$\begin{aligned} \mathcal{J}_{\boldsymbol{\beta}}(\mathbf{x}_0, \boldsymbol{\beta}) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^b)^T \mathbf{B}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}^b) \\ + \quad & \frac{1}{2} \sum_{i=0}^{N} (\mathcal{H}_i(\mathbf{x}_i) + \mathbf{b}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}} \mathbf{R}_i^{-1} (\mathcal{H}_i(\mathbf{x}_i) + \mathbf{b}(\boldsymbol{\beta}, \mathbf{x}_i) - \mathbf{y}_i) \end{aligned} \tag{3.17}$$

subject to the dynamical equations, to estimate the state $\mathbf{x}_0$ and the coefficients $\boldsymbol{\beta}$ simulteneously [33]. Alternatively, a variational procedure can be used to estimate these coefficients offline at regular intervals, using the previous value as the background for the new estimate [3].

Finally we consider the specification of the observation error covariance matrix, which represents the covariance of the random components of the observation error. It is important to note that this error is defined by the difference between the actual measurement and the model representation of the true state $\mathbf{x}^t$ mapped into observation space by the observation operator, that is the error $\boldsymbol{\epsilon}_i^o$ at time t_i is given by

$$\boldsymbol{\epsilon}_i^o = \mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i^t). \tag{3.18}$$

This means that the error includes different components arising from the accuracy of the measuring instrument (instrument error), errors in the observation operator $\mathcal{H}_i$

and errors due to the difference in spatial resolution between the measurement and the model state (known as representativity error). The instrument error is the easiest to treat, since the variances of this error can usually be obtained from the instrument manufacturer and it is normally safe to assume that these errors are uncorrelated. However, this may not always be the case. For example, measurements derived by preprocessing satellite data may include spatial correlations [17]. Errors in the observation operator may include such things as errors in the radiative transfer models used to model satellite data, which can lead to error correlations between different satellite channels [16], [105].

Although it is recognized that observation error correlations exist, particularly with respect to satellite data, the correlations are not usually very well treated in current operational forecasting systems. Often the correlations are ignored and it is assumed that the observation error covariance matrix is diagonal. To balance this assumption either the error variances are inflated [56] or the data are thinned so that many fewer of them are used [32]. The reasons for this are the difficulty in calculating what the error correlations should be and the difficulty in then representing these correlations within an assimilation scheme in a way that the inverse correlation matrix can easily be applied. To estimate the correlations in satellite data the methods that have mainly been used are a comparison with independent measurements from radiosondes, based on the method of [57], and the use of diagnostics calculated from the data assimilation system itself, based on [35]. Various ways of then representing these correlations within the data assimilation system have been proposed, including the use of a circulant matrix [55], an eigenvalue decomposition [37] and a Markov matrix [105]. However there is so far little use of these methods in operational practice.

3.4 Optimization methods

The minimization of the inner loop cost function (2.15) requires the use of a suitable optimization algorithm. For the large problems of environmental modelling there are two particularly important constraints. The first is that because of the number of variables in the system it is not possible to obtain second derivative information. The Hessian or second derivative matrix would contain of the order 10^{16} elements, which is impossible to calculate or to store. Hence only methods that require first derivative information can be used. The second constraint is that often these problems must be solved within a real-time forecasting system and hence the computer time that can be used to solve the problem is very limited. Hence the methods much use as few function evaluations as possible. This means that usually the problem is not allowed to run to full convergence and the use of any line search algorithms is prohibitively expensive. Traditionally the algorithms that have most been used within data assimilation systems are quasi-Newton algorithms and conjugate gradient or related Lanczos algorithms, which require only first derivative information to be provided. The mathematical details of these algorithms are well explained elsewhere (e.g. [94]) and so here

we limit discussion to their implementation in data assimilation systems.

An essential aspect of the minimization procedure for variational data assimilation is an appropriate preconditioning. Experimental evidence indicates that the Hessian of the inner loop cost function (2.15) is badly conditioned and that this arises from the ill-conditioning of the background error covariance matrix [83]. This has been further confirmed by theoretical results that bound the condition number of the Hessian of the cost function in terms of the condition number of this covariance matrix [50], [51]. The first level of preconditioning that is applied is therefore to transform the problem to new variables, as described in section 3.2. The transformed problem (3.5) can be shown in general to be better conditioned both in theory and in practice [83], [42], [50], [51]. However, even after this transformation the problem is not very well-conditioned and can have a condition number of order $10^3 - 10^4$ [39], [52]. Experiments in the ECMWF system showed that the ill-conditioning that remains is related to the inclusion of dense, accurate surface observations over Europe [110] and this has also been shown to be true for the system of the Met Office [52]. This can be explained by theoretical bounds obtained by [50], [52] that show that the condition number of the transformed problem increases as the spacing between observations decreases and as observations become more accurate. Hence ideally a second level of preconditioning is required after the variable transformation has been performed.

In order to implement a further preconditioning it is necessary to find a preconditioning matrix $\mathbf{K}$ that is inexpensive to compute and such that the eigenvalues of $\mathbf{K}\tilde{\mathbf{G}}$ are more clustered than those of the Hessian $\tilde{\mathbf{G}}$ of the transformed problem. Often the preconditioning matrix may be represented in the factored form $\mathbf{K} = \mathbf{P}\mathbf{P}^T$ and the preconditioning matrix $\mathbf{P}$ is then used directly, for example in the preconditioned conjugate gradient method [111]. In order to design such a preconditioner some knowledge of the Hessian (3.10) of the transformed cost function is required. One way that this can be obtained is by using a Lanczos algorithm to perform the inner loop minimization. The Lanczos method produces estimates of the leading eigenvectors and eigenvalues of the Hessian of the function being minimized. If the first m eigenvalues λ_j and eigenvectors $\mathbf{u}_j, j = 1, \ldots, m$ have sufficiently converged then the inverse of the Hessian (3.10) can be approximated by the expression

$$\mathbf{K} = \mathbf{I} + \sum_{j=1}^{m} (\lambda_j - 1)\mathbf{u}_j \mathbf{u}_j^T. \tag{3.19}$$

This expression can then be used for the preconditioning of subsequent minimizations, under the assumption that the Hessian does not change greatly between one minimization and another [39], [111]. This method, known as spectral preconditioning, is used in the operational forecast system of ECMWF, where three outer loops are performed for each assimilation. During the first inner loop minimization the Lanczos vectors are stored and these are then used to precondition the mininimization of the second and third inner loop cost functions [39]. It has been shown that this preconditioner belongs to a larger class of limited memory preconditioners [111]. In order to define this class

we let $\mathbf{s}_i \in \mathbb{R}^n, i = 1, \ldots, l$, with $l < n$, be a set of $\tilde{\mathbf{G}}-$conjugate vectors. Then the limited-memory preconditioning matrix is given by

$$\mathbf{K}_l = \left(\mathbf{I}_n - \sum_{i=1}^{l} \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \tilde{\mathbf{G}} \mathbf{s}_i} \tilde{\mathbf{G}}\right) \left(\mathbf{I}_n - \sum_{i=1}^{l} \tilde{\mathbf{G}} \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \tilde{\mathbf{G}} \mathbf{s}_i}\right) + \sum_{i=1}^{l} \frac{\mathbf{s}_i \mathbf{s}_i^T}{\mathbf{s}_i^T \tilde{\mathbf{G}} \mathbf{s}_i}. \tag{3.20}$$

If the vectors $\mathbf{s}_i$ are chosen to be the eigenvectors of $\tilde{\mathbf{G}}$ then this formula results in the spectral preconditioning matrix (3.19).

The authors of [111] propose an alternative preconditioner from the same class, based on the Ritz pairs of the Hessian. Ritz pairs are approximate eigenpairs $(\theta_i, \mathbf{v}_i)$ defined in an appropriately chosen subspace. By choosing the subspace to be that spanned by the Lanczos vectors, the authors obtain the Ritz limited memory preconditioner

$$\mathbf{K}_l^{Ritz} = \left(\mathbf{I}_n - \sum_{i=1}^{l} \frac{\mathbf{v}_i \mathbf{v}_i^T}{\theta_i} \tilde{\mathbf{G}}\right) \left(\mathbf{I}_n - \sum_{i=1}^{l} \tilde{\mathbf{G}} \frac{\mathbf{v}_i \mathbf{v}_i^T}{\theta_i}\right) + \sum_{i=1}^{l} \frac{\mathbf{v}_i \mathbf{v}_i^T}{\theta_i}. \tag{3.21}$$

They found that the use of this preconditioner can provide an improvement over spectral preconditioning when the estimates of the Hessian eigenpairs are inaccurate. A similar result was also found in the Regional Ocean Modelling System (ROMS), in which both of these preconditioners are implemented [91]. One drawback of both of these methods is that, in order to generate the required information, the first minimization must be performed in order to generate the vectors $\mathbf{s}_i$ before any preconditioning can be applied. So far little attention has been paid to preconditioning of this first minimization.

With any minimization method it is important to specify appropriate stopping criteria and this is also the case in variational data assimilation. As discussed in section 2.1 it has been proved that the inner-loop step of the Gauss-Newton method (step 1.1 of Algorithm 2.1) needs to be solved to sufficient accuracy in order to ensure convergence of the outer loops [45]. The theory has been used to show how it is natural to use an inner-loop stopping criterion based on the relative change in the norm of the gradient, of the form

$$\frac{||\nabla \tilde{\mathcal{J}}_{(l)}^{(k)}||_2}{||\nabla \tilde{\mathcal{J}}_{(0)}^{(k)}||_2} < \epsilon, \tag{3.22}$$

where the subscript indicates the inner-loop iteration index and ϵ is a specified tolerance [73]. The tolerance used to stop the iterations must therefore be chosen carefully. If it is too high then there is no guarantee that the outer loop steps will converge. However the convergence should not be pushed below the level of noise on the observations, as then small spatial scales are adjusted to fit the observational noise [68]. In many practical forecasting problems such care is not always taken and other criteria are introduced. There are two main reasons for this. One is that in a time-critical forecasting system it may considered more important to solve each minimization problem

using approximately the same amount of wall-clock time rather than to the same accuracy. The second reason is that the preconditioning techniques decribed in this section require a minimum number of iterations to be performed on the first inner-loop minimization in order to acquire sufficiently accurate information about the Hessian. Hence criteria that have been introduced include stopping the iterations when the value of the cost function is close to its expected minimum value [84] or using a fixed number of iterations, particularly for the first minimization [110].

3.5 Reduced order approaches

As was mentioned in section 2.1 a major advantage of the incremental approach is that the inner loop problem may be solved in a smaller dimensional space than the outer loop update of the linearization trajectory. Within environmental prediction lower spatial resolution systems have often been used in the inner loop step, with the full resolution nonlinear model being used in the outer loop. Further simplifications may also be made to the linear dynamical model used in the inner loop, such as using simplified parametrizations of sub-grid scale processes as described in section 3.1. While a change in resolution is certainly the simplest way to achieve a more computationally tractable inner loop problem, it does not necessarily provide the most accurate low order representation of the linearized cost function and its constraint. In order to improve on this other reduced order approaches have been investigated in the context of incremental 4D-Var. These essentially fall into two categories, methods based on principal component analysis and methods based on near-optimal reduction of dynamical systems.

Principal component analysis, which is often referred to as principal orthogonal decomposition (POD) or the method of empirical orthogonal functions (EOFs), aims to represent the solution of the assimilation problem as a linear combination of basis vectors. The basis vectors are chosen to represent the leading directions of variability in the model and are calculated using a series of model states, or 'snapshots', from an integration of the nonlinear model. Such a method was used in an ocean model assimilation by [101]. From the sample of model states the authors generate the matrix $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_l)$, where $\mathbf{X}_i$ is the difference between the model state at time t_i and the mean state. The covariance matrix $\mathbf{X}\mathbf{X}^T/(l-1)$ is then diagonalised to find a set of orthonormal eigenvectors $\mathbf{v}_i$ (EOFs) and associated eigenvalues $\lambda_i, i = 1, \ldots, l$.[5] The solution $\delta\mathbf{x}_0$ to the inner loop minimization problem (2.15) is then defined by an expansion of the leading r eigenvectors

$$\delta\mathbf{x}_0 = \sum_{i=0}^{r} w_i \mathbf{v}_i = \mathbf{V}\mathbf{w} \tag{3.23}$$

[5] In practice the eigenvalues can be found by diagonalising the much smaller matrix $\mathbf{X}^T\mathbf{X}/(l-1)$ [11].

where $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_r)$ is the matrix of the leading r eigenvectors and $\mathbf{w} = (w_1, \ldots, w_r)^T$ are the weights to be determined. In this case the matrix $\mathbf{V}$ acts as a variable transformation in a similar way to the parameter transform (3.4) and so the background term can be written in the form

$$\mathcal{J}_b(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{B}_w^{-1}\mathbf{w}, \tag{3.24}$$

where the covariance matrix $\mathbf{B}_w$ is taken to be the diagonal matrix of eigenvalues. The number of vectors r that are used in the expansion is chosen in order to ensure that a large fraction of the total variance is retained, where this fraction is calculated from the eigenvalues as

$$\frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{l}\lambda_i}. \tag{3.25}$$

This method has been applied to assimilation in ocean models in an idealized setting [101] and using real data [60]. It is noted that the assumption behind this method is that the variability of the system can be well described by a low dimensional space. Although the approach reduces the size of the space in which the minimization is performed, the tangent linear model (2.16) must still be integrated at full resolution on each iteration.

An alternative approach, based on POD, was put forward by [22], [23]. In that work the solution to the full nonlinear 4D-Var problem is expressed as a perturbation from the sample mean that is expanded in terms of basis functions $\mathbf{\Phi}_i$, such that

$$\delta\mathbf{x}_0 = \sum_{i=0}^{r} w_i\mathbf{\Phi}_i, \tag{3.26}$$

where w_i are again weights to be determined. The basis functions are derived in a similar way to the EOFs, but by then projecting the perturbation fields $\mathbf{X}$ onto the eigenvectors $\mathbf{v}_i$, so that

$$\mathbf{\Phi} = \{\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_l\} = \mathbf{XV}. \tag{3.27}$$

The number of basis functions that are used in the expansion is again determined using the fractional variance (3.25). In this work the authors solve the nonlinear 4D-Var cost function (2.3) in the reduced space. As well as expressing the background term in terms of coefficients of the basis functions they also derive a Galerkin projection of the dynamical model onto the basis functions for use in the observation term. Thus this formulation has the advantage that the dynamical model and its adjoint are also expressed in the reduced space. Again this method relies on the snapshots being able to capture a low-dimensional subspace that adequately describes the full system.

A disadvantage with both the EOF and POD methods is that they do not use any information about the data assimilation problem itself within the reduction procedure. There have been two approaches proposed to improve on this. The first is an adaption of the POD method, called dual-weighted POD. In this method the snapshot perturbations $\mathbf{X}$ are weighted according the sensitivity of the cost function at the time of the

snapshot, where the weights are calculated using the adjoint model [30]. The other approach, put forward in the series of papers [76], [75], [14], is to use near-optimal model order reduction methods for linear dynamical systems to derive a reduced order model and observation operator. The inner loop problem of incremental 4D-Var (2.15) is subject to the dynamical system described by the evolution equation (2.16) and the output equation

$$\mathbf{d}_i = \mathbf{H}_i \delta \mathbf{x}_i. \tag{3.28}$$

Model reduction seeks linear restriction operators $\mathbf{S}_i^T$ and prolongation operators $\mathbf{T}_i$ that map the perturbation $\delta \mathbf{x}_i \in \mathbb{R}^n$ to $\delta \hat{\mathbf{x}}_i \in \mathbb{R}^r$ with $r << n$. These operators are chosen such that the output of the projected system

$$\begin{aligned} \delta \hat{\mathbf{x}}_{i+1} &= \mathbf{S}_i^T \mathbf{M}_i \mathbf{T}_i \delta \hat{\mathbf{x}}_i & (3.29) \\ \hat{\mathbf{d}}_i &= \mathbf{H}_i \mathbf{T}_i \delta \hat{\mathbf{x}}_i & (3.30) \end{aligned}$$

approximates well the output of the full dynamical system $\mathbf{d}_i$. The inner loop problem can then be defined in the reduced space as the minimization of

$$\begin{aligned} \min \hat{\mathcal{J}}^{(k)}[\delta \hat{\mathbf{x}}_0^{(k)}] &= \frac{1}{2}(\delta \hat{\mathbf{x}}_0^{(k)} - \mathbf{S}_0^T[\mathbf{x}^b - \mathbf{x}_0{}^{(k)}])^{\mathrm{T}} \\ &\times (\mathbf{S}_0^T \mathbf{B}_0 \mathbf{S}_0)^{-1}(\delta \hat{\mathbf{x}}_0^{(k)} - \mathbf{S}_0^T[\mathbf{x}^b - \mathbf{x}_0{}^{(k)}]) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathbf{H}_i \mathbf{T}_i \delta \hat{\mathbf{x}}_i^{(k)} - \mathbf{d}_i^{(k)})^{\mathrm{T}} \mathbf{R}^{-1}(\mathbf{H}_i \mathbf{T}_i \delta \hat{\mathbf{x}}_i^{(k)} - \mathbf{d}_i^{(k)}), \end{aligned}$$

subject to the reduced dynamical model (3.29). The linearization state is then updated with the perturbation

$$\delta \mathbf{x}_0^{(k)} = \mathbf{T}_0 \delta \hat{\mathbf{x}}_0^{(k)}. \tag{3.31}$$

The authors of these papers use the method of balanced truncation [92] to demonstrate this method in the case where the operators $\mathbf{M}$ and $\mathbf{H}$ are time-invariant. The aim of balanced truncation is to truncate the states of the system that are least affected by the inputs and have least effect on the outputs. Since these are not generally the same, the first step in the method is to transform the system into one in which these states coincide, the 'balancing' step. It is first necessary to find the state covariance matrices $\mathbf{P}$ and $\mathbf{Q}$ associated with the inputs and outputs respectively. These are found by solving the Stein equations

$$\begin{aligned} \mathbf{P} &= \mathbf{M}\mathbf{P}\mathbf{M}^T + \mathbf{B} & (3.32) \\ \text{and} \quad \mathbf{Q} &= \mathbf{M}^T \mathbf{Q} \mathbf{M} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}. & (3.33) \end{aligned}$$

The balancing transformation $\mathbf{\Psi}$ is then given by the matrix of eigenvectors of $\mathbf{PQ}$, while the eigenvalues of $\mathbf{PQ}$ are equal to the Hankel singular values of the full system.

The reduction step then calculates the restriction and prolongation operators from

$$\mathbf{S}^T = [\mathbf{I}_r, \mathbf{0}]\mathbf{\Psi}^{-1} \tag{3.34}$$

$$\mathbf{T} = \mathbf{\Psi}\begin{bmatrix} \mathbf{I}_r \\ \mathbf{0} \end{bmatrix}, \tag{3.35}$$

where the decay of the Hankel singular values is used to choose the model reduction order r. In idealised models the studies [76], [75], [14] show how this method improves the solution with respect to using low resolution models and how it is important to use information about the assimilation problem in the reduction procedure, including information about the background and observation error covariance matrices. However, whereas reduction methods based on POD can be implemented in large systems, the method of balanced truncation cannot. Although efficient numerical methods are available to apply balanced truncation to systems of moderately large size (e.g. [53], [69], [25]), these are not suitable for the very large systems found in environmental prediction. Efforts are being made to design near-optimal reduction methods for such systems based on Krylov methods [21], but these methods have not yet been tried out in data assimilation for large systems.

3.6 Issues for nested models

For very high resolution weather and ocean forecasting operational centres often use models covering only the domain of interest that are nested in a larger model, often of lower resolution, which we refer to here as the parent model. In most of the systems the nesting is a one-way nesting, whereby lateral boundary conditions for the nested model are provided by the parent model, but there is no feedback from the high resolution nested model to the parent model. This presents particular challenges for the application of variational data assimilation. For problems specific to high resolution weather forecasting we refer the reader to the review articles [96] and [31]. Here we consider only more general problems arising from using a high resolution nested grid, in particular treatment of the lateral boundary conditions and of the difference in representation of spatial scales between the parent and nested models.

With respect to the lateral boundary conditions, a decision must be made as to whether to estimate them as part of the assimilation procedure or to assume that they do not change. Both approaches have been used in practice. In the operational weather forecasting system of the Met Office the lateral boundary conditions are not updated, but are fixed by the parent model. Hence the increment $\delta\mathbf{x}$ on the boundary is set to zero. This has advantages for the practical implementation of the scheme. In particular it allows a simple sine transform to be used in the definition of the spatial background error covariances described in section 3.2, which then enforces zero boundary increments [83]. However, observational information close to the boundaries can be difficult to use, since the nested model cannot use observations lying outside the domain

and the analysis inside the domain may not be consistent with the boundary conditions provided [4], [47]. This can lead to features being artificially cut-off close to the boundaries.

The alternative approach is to estimate the boundary variables within the assimilation procedure [48], [67], [49]. This means that the state vector $\mathbf{x}$ is defined to include both the variables in the interior of the domain and on the lateral boundaries. In this way observations inside the nested domain can update the boundary values and so it is possible to ensure that the analysis is consistent throughout the domain. However in this case it is no longer possible to apply a sine transform to impose the spatial background error covariances. In order to be able to apply a spectral transformation an extension zone is created around the domain to obtain fields that are horizontally periodic. A Fourier transform can then be applied. One difficulty in analysing the boundaries in this way is that the lateral boundary conditions are only updated during the assimilation period. During the subsequent forecast no updates are available and the values from the parent model must be used, so there is some inconsistency between the boundary conditions of the analysis and those of the forecast. However, some consistency over the assimilation window can be ensured by estimating the boundary conditions at the beginning and end of the assimilation window, with both constrained by background values from the parent model. In this case the cost function to be solved is of the form

$$\begin{aligned} \mathcal{J}(\mathbf{x}_0, \mathbf{x}_{lbc}) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\mathbf{x}_{lbc} - \mathbf{x}^b_{lbc})^{\mathrm{T}}\mathbf{B}^{-1}_{lbc}(\mathbf{x}_{lbc} - \mathbf{x}^b_{lbc}) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i), \end{aligned} \tag{3.36}$$

where $\mathbf{x}_0$ represents the model variables in the interior of the domain and the lateral boundary conditions at initial time t_0, $\mathbf{x}_{lbc}$ is the lateral boundary condition at final time t_N, $\mathbf{x}^b$ is the background estimate of $\mathbf{x}_0$, with error covariance matrix $\mathbf{B}$ and $\mathbf{x}^b_{lbc}$ is the background estimate of $\mathbf{x}_{lbc}$, with error covariance matrix $\mathbf{B}_{lbc}$ [67].

The second challenge we consider is the difference in the spatial scales that can be represented in the nested and parent models. In particular, since the nested model often covers only a small domain, the assimilation scheme is not able to analyse adequately scales of the size of the domain and larger. In applications such as weather prediction it is important to capture these larger scales, since the physical system is inherently multiscale, with strong feedbacks between large and small scales. Hence attempts have been made to improve the large-scale information in nested model data assimilation by providing information on these scales from a parent model analysis. For example, the Met Office experimented with a system that combined large scale increments from a parent model analysis with the small scale increments from the nested model analysis [4]. In this method the large scales of the nested model analysis are forced to be equal to those of the parent model.

An alternative, proposed by [47], is to use the large scales of the parent analysis over the nested model domain as a weak constraint on the variational problem. We let $\mathbf{x}_p^a$ be the analysis from the parent model and define operators $\mathcal{H}_p$ and $\mathcal{H}_n$ such that $\mathcal{H}_p(\mathbf{x}_p^a)$ represents some large scales of the parent analysis on the nested domain and $\mathcal{H}_n(\mathbf{x})$ represents the same large scales from the nested model field $\mathbf{x}$. Then the difference between the large scales of the global analysis and those forecast by the nested model can be constrained by adding an extra term to the cost function (2.3) of the form

$$\frac{1}{2}(\mathcal{H}_p(\mathbf{x}_p^a) - \mathcal{H}_n(\mathbf{x}))^T \mathbf{B}_p^{-1}(\mathcal{H}_p(\mathbf{x}_p^a) - \mathcal{H}_n(\mathbf{x})), \tag{3.37}$$

where $\mathbf{B}_p$ is the error covariance matrix of the parent model large scales. This means that the analysis is constrained by large scales from the parent model, through this additional term, and by large scales from the nested model, through the background term. In theory this should introduce another term including the cross-correlation between these two sources of information. However, in their demonstration of the method in a 3D-Var scheme of the ALADIN model at Météo-France the authors of [47] concluded that this cross-correlation could be neglected, though at the cost of some inaccuracy.

A more theoretical study of this problem was carried out by [9]. They used a spectral analysis to show how information from waves longer than the domain size is projected onto different scales in the nested model domain, corresponding to the lowest wave numbers that can be represented on this domain. They demonstrated that by giving more weight to these scales in the background term of the cost function it was possible to retain more of the large scale information from a parent model background. In this method the large spatial scales from only the parent model are used as a constraint in the assimilation, as in [4], but they are not imposed exactly and may be altered by the assimilation process. The authors of [9] demonstrated benefit from this in an idealised system, but the method has not been tested in a realistic model.

3.7 Weak constraint variational assimilation

The formulation of variational data assimilation presented in section 2 assumes that the discrete dynamical model (2.1) is an exact representation of the physical system being observed. In practice we know that the models contain errors, caused by limitations in our knowledge of the physical equations and limitations in the numerical modelling, such as the need for sub-grid scale parametrizations. In theory it is possible to account for and estimate such errors in variational data assimilation, though implementation in practice is more complicated. We assume an additive error to the model equations, so that the true dynamical system can be written

$$\mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i) + \boldsymbol{\eta}_i, \tag{3.38}$$

where $\boldsymbol{\eta}_i$ are the unknown model errors at times t_i, which are assumed to be random, serially uncorrelated, Gaussian errors with covariance matrix $\mathbf{Q}_i$. Then we can define

a weak constraint 4D-Var problem, in which the model equations do not have to be exactly satisfied over the assimilation window. We define a cost function of the form

$$\mathcal{J}(\mathbf{x}_0, \boldsymbol{\eta}_0, \ldots, \boldsymbol{\eta}_{N-1}) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b)$$
$$+ \quad \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i) + \frac{1}{2}\sum_{i=0}^{N-1}\boldsymbol{\eta}_i^T\mathbf{Q}_i^{-1}\boldsymbol{\eta}_i \qquad (3.39)$$

subject to (3.38). The weak constraint problem is then to minimize (3.39) with respect to the initial state $\mathbf{x}_0$ and all the model errors $\boldsymbol{\eta}_i$.

An alternative formulation of the weak constraint problem (3.39) is to write it in terms of the model state $\mathbf{x}_i$ at each time t_i rather than in terms of the model errors. This leads to the cost function

$$\begin{aligned}\mathcal{J}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_N) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) \\ &+ \frac{1}{2}\sum_{i=0}^{N}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i) \\ &+ \frac{1}{2}\sum_{i=0}^{N-1}(\mathbf{x}_{i+1} - \mathcal{M}_i(\mathbf{x}_i))^T\mathbf{Q}_i^{-1}(\mathbf{x}_{i+1} - \mathcal{M}_i(\mathbf{x}_i)). \qquad (3.40)\end{aligned}$$

In [109] both formulations were presented in the incremental version of 4D-Var as possibilities for inclusion in the ECMWF system.

The inclusion of the model errors at each observation time increases the size of the argument of $\mathcal{J}$ by a factor of $N + 1$, the number of observation times. One way to reduce this cost is by assuming a relationship in time between the model errors $\boldsymbol{\eta}_i$. Theoretical work by [46] used an augmented state approach to solve for the state and the model error, with a dynamical equation used to explain the evolution of the error. The authors introduced a general form for the error evolution, including both a systematic and random component of the error. Various options for the systematic evolution were proposed, including a contant bias error and simple dynamical evolutions, and the methods were illustrated on simple systems. In the context of a regional atmospheric model [119] demonstrated a weak-constraint 4D-Var system under the assumption that the model error was serially correlated and obeyed a first order Markov process.

Since this early work there have been several idealised studies with weak-constraint 4D-Var, but the move towards operational implementations in large-scale systems has been slow. One of the biggest challenges remains the specification of the model error covariance matrix $\mathbf{Q}_i$ for real systems. An initial idea was to take this matrix to be a scalar multiple of the background error covariance matrix $\mathbf{B}$. However, in experiments with the ECMWF atmospheric forecasting system using formulation (3.39), [110] showed that this choice implies that corrections to the model error lie in the same

space as those to the background. This leads to estimates of model error that are very similar to the increments to the initial conditions. An alternative method, proposed in the same paper, is based on the use of model tendency fields, that is fields of the change in model variables over a model time step. The statistics of $\mathbf{Q}_i$ are estimated from an ensemble of differences between model tendency fields using the NMC method, in a similar way that differences between the model fields themselves are used in the estimation of the background error covariances (as explained in section 3.2). [110] interprets differences between these tendencies as a proxy for the uncertainty in the model forcing. The statistics from this sample are then fit to the same statistical model as is used for the matrix $\mathbf{B}$. The use of a covariance matrix estimated in this way was tested in weak-constraint 4D-Var experiments that assumed a constant error over the assimilation window. This was shown to give an improvement over the use of a covariance matrix defined by a scalar multiple of $\mathbf{B}$.

The work of [80] illustrated the implementation of weak-constraint 4D-Var using such a matrix, again in the ECMWF system, to estimate a constant bias error in the stratosphere, where the model is known to have biases. A similar scheme has been introduced into the operational assimilation system of ECMWF [40]. In this implementation the deviation of the error from its mean value is minimized, so that the last term of (3.39) becomes

$$\frac{1}{2}(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}})^T \mathbf{Q}^{-1}(\boldsymbol{\eta} - \bar{\boldsymbol{\eta}}), \tag{3.41}$$

where $\bar{\boldsymbol{\eta}}$ is the estimate of the model bias from the previous analysis cycle. In this way the assimilation ensures that the estimated error does not vary too quickly from one analysis cycle to the next.

Despite these initial successes much more work is needed. One particular difficulty is that it is not clear how to differentiate between model bias and observation bias, since the assimilation only measures the difference between the model and the observations. [110] showed a case study of observation bias being interpreted as a model error by weak-constraint 4D-Var. This problem was discussed further by [78] in the context of ocean data assimilation. They suggested that to estimate both model and observation bias it is necessary to include information on the spatial and temporal structure of these biases in the covariance matrices.

In order to then move away from the assumption of a constant bias and treat time-varying systematic and random model errors, more sophisticated methods for describing the evolution of errors must be developed. This evolution is likely to be dependent on the specific model being used, yet general methods for representing this are also needed. At the same time efficient and accurate representations of the covariances of these model errors must be found. The use of the weak-constraint formulation of 4D-Var holds much promise to counteract the inadequacies of models, but many challenges remain open to be able to implement this in very large environmental models.

4 Summary and future perspectives

Variational data assimilation is now a well-established method for combining observational data with very large environmental models. However, as has been illustrated in this article, its successful implementation requires careful and judicious choices in each aspect of the assimilation scheme. In some cases these choices are determined by the physical system being modelled or the observational data available, such as the specification of the error covariances in the system. In other cases the choices may be determined by the size of the problem and the need to solve it in an efficient manner, often for real-time forecasting, or by features of the numerical model itself, such as lateral boundary conditions. In each instance the choices to be made will inevitably be a compromise between the ideal solution and what is practically feasible in a given system. We have presented some of the solutions that have been found that have allowed variational data assimilation to be implemented in large environmental forecasting systems. Nevertheless much research continues to improve on these solutions so as to find better estimates of the state and so produce better forecasts.

One particularly active area in numerical weather prediction is the desire to use more information from ensembles of forecasts to provide time-varying covariances for the background errors, combining the advantages of ensemble filtering methods with the advantages of 4D-Var. ECMWF have implemented a system in which an ensemble of 4D-Var assimilations are run and the statistics from this ensemble are used to update the variances of the background errors [15]. Extensions to this method to calculate also the covariance information are being sought. An alternative approach is to use information from ensembles of forecasts to calculate covariance information throughout the whole assimilation window. This method was proposed by [81] and tested in a global weather prediction model by [19], [20]. An advantage of this method is that the tangent linear and adjoint models are not required in the 4D-Var, since all the evolution information comes through the ensemble of nonlinear model forecasts. Hence this makes development of the system much easier.

Besides the many great challenges that we have discussed in this article, new challenges are arising for the future evolution of variational data assimilation systems. The advent of massively parallel computers means that the algorithms used currently to solve the assimilation problem may no longer be efficient on future computer architectures. Hence work is needed to develop new algorithms to solve the problem, particularly with respect to efficient minimization and preconditioning methods. This may be easier as systems move to a weak-constraint form of 4D-Var but, as discussed above, that introduces its own difficulties [40]. Another challenge comes from the move towards more integrated Earth-system models, with different environmental models coupled to each other. For example, for seasonal to decadal prediction it is now common to use coupled atmosphere-ocean models, but the initialization of these models with data assimilation is still in its infancy. Particular problems arise from the very different time scales in the atmosphere and ocean system and from the model biases in atmo-

sphere and ocean models. Some work has been done to implement 4D-Var in such systems in order to estimate the ocean state and coupling parameters [106], [89], but the estimation of the complete state in coupled atmosphere-ocean models remains an open problem for the coming years.

Acknowledgments. I would like to thank two anonymous reviewers whose detailed reading of the manuscript helped lead to many improvements in the final version.

Bibliography

[1] W.K. Anderson and V. Venkatakrishan, Aerodynamic design optimization on unstructured grids with a continuous adjoint formulation, *Computers and Fluids* 28 (1999), 443–480.

[2] E. Andersson and H. Järvinen, Variational quality control, *Quart. J. Roy. Meteor. Soc.* 125 (1999), 697–722.

[3] T. Auligné, A.P. McNally and D.P. Dee, Adaptive bias correction for satellite data in a numerical weather prediction system, *Quart. J. Roy. Meteor. Soc.* 133 (2007), 631–642.

[4] S. Ballard, Z. Li, M. Dixon, S. Swarbrick, O. Stiller and H. Lean, Development of 1-4km resolution data assimilation for nowcasting at the Met Office, in: *World Weather Research Program Symposium on Nowcasting and Very Short Range Forecasting (WSN05)*, p. Paper 3.02, 2005.

[5] R.N. Bannister, Can wavelets improve the representation of forecast error covariances in variational data assimilation?, *Mon. Wea. Rev.* 135 (2007), 387–408.

[6] ———, A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, *Quart. J. Roy. Meteor. Soc.* 134 (2008), 1951–1970.

[7] ———, A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, *Quart. J. Roy. Meteor. Soc.* 134 (2008), 1971–1996.

[8] R.N. Bannister and M.J.P. Cullen, A regime-dependent balanced control variable based on potential vorticity, in: *Proceedings of ECMWF workshop in flow-dependent aspects of data assimilation*, pp. 1–13, 2007.

[9] G.M. Baxter, S.L. Dance, A.S. Lawless and N.K. Nichols, Four-dimensional variational data assimilation for high resolution nested models, *Computers & Fluids* 46 (2011), 137–141.

[10] C. Bischof, A. Carle, G. Corliss, A. Griewank and P. Hovland, ADIFOR: Generating derivative codes from Fortran programs, *Scientific Programming* 1 (1992), 11–29.

[11] J. Blum, F.-X. Le Dimet and I. M. Navon, Data assimilation for geophysical fluids, *Handbook of Numerical Analysis: Computational Methods for the Atmosphere and the Oceans. Elsevier, Amsterdam* XIV (2008), 377–433.

[12] M. Bocquet, Reconstruction of an atmospheric tracer source using the principle of maximum entropy. I: Theory, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 2191–2208.

[13] M. Bocquet, C.A. Pires and L. Wu, Beyond Gaussian statistical modeling in geophysical data assimilation, *Mon. Wea. Rev.* 138 (2010), 2997–3023.

[14] C. Boess, A.S. Lawless, N.K. Nichols and A. Bunse-Gerstner, State estimation using model order reduction for unstable systems, *Computers & Fluids* 46 (2011), 155–160.

[15] M. Bonavita, L. Isaksen and E. Hólm, On the use of EDA background error variances in the ECMWF 4D-Var, *Quart. J. Roy. Meteor. Soc.* 138 (2012), 1540–1559.

[16] N. Bormann and P. Bauer, Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data, *Quart. J. Roy. Meteor. Soc.* 136 (2010), 1036–1050.

[17] N. Bormann, S. Saarinen, G. Kelly and J.N. Thepaut, The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data, *Mon. Wea. Rev.* 131 (2003), 706–718.

[18] M. Buehner, Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 1013–1043.

[19] M. Buehner, P.L. Houtekamer, C. Charette, H.L. Mitchell and B. He, Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments, *Mon. Wea. Rev.* 138 (2010), 1550–1566.

[20] ———, Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations, *Mon. Wea. Rev.* 138 (2010), 1567–1586.

[21] A. Bunse-Gerstner, D. Kubalinska, G. Vossen and D. Wilczek, h2-norm optimal model reduction for large scale discrete dynamical MIMO systems, *Journal of computational and applied mathematics* 233 (2010), 1202–1216.

[22] Y. Cao, J. Zhu, Z. Luo and I.M. Navon, Reduced-order modeling of the upper tropical Pacific ocean model using proper orthogonal decomposition, *Computers and Mathematics with Applications* 52 (2006), 1373–1386.

[23] Y. Cao, J. Zhu, I.M. Navon and Z. Luo, A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition, *International Journal for Numerical Methods in Fluids* 53 (2007), 1571–1583.

[24] C. Cardinali, S. Pezzulli and E. Andersson, Influence-matrix diagnostic of a data assimilation system, *Quart. J. Roy. Meteor. Soc.* 130 (2004), 2767–2786.

[25] Y. Chahlaoui and P. Van Dooren, *Model reduction of time-varying systems*, Dimension reduction of large-scale systems (Mehrmann-V. Benner, P. and D. Sorensen, eds.), Springer-Verlag, 2005, pp. 131–142.

[26] W.C. Chao and L-P. Chang, Development of a four-dimensional variational analysis system using the adjoint method at GLA. Part I: Dynamics, *Mon. Wea. Rev.* 120 (1992), 1661–1673.

[27] P. Courtier, J-N. Thepaut and A. Hollingsworth, A strategy for operational implementation of 4D-Var, using an incremental approach, *Quart. J. Roy. Meteor. Soc.* 120 (1994), 1367–1387.

[28] M.J.P. Cullen, Four-dimensional variational data assimilation: A new formulation of the background-error covariance matrix based on a potential-vorticity representation, *Quart. J. Roy. Meteor. Soc.* 129 (2003), 2777–2796.

[29] _____, A demonstration of 4D-Var using a time-distributed background term, *Quart. J. Roy. Meteor. Soc.* 136 (2010), 1301–1315.

[30] D.N. Daescu and I.M. Navon, A dual-weighted approach to order reduction in 4DVAR data assimilation, *Mon. Wea. Rev.* 136 (2008), 1026–1041.

[31] S.L. Dance, Issues in high resolution limited area data assimilation for quantitative precipitation forecasting, *Physica D: Nonlinear Phenomena* 196 (2004), 1–27.

[32] M.L. Dando, A.J. Thorpe and J.R. Eyre, The optimal density of atmospheric sounder observations in the Met Office NWP system, *Quart. J. Roy. Meteor. Soc.* 133 (2007), 1933–1943.

[33] D.P. Dee, Bias and data assimilation, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 3323–3343.

[34] J.E. Dennis, Jr and R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Society for Industrial and Applied Mathematics, 1996.

[35] G. Desroziers, L. Berre, B. Chapnik and P. Poli, Diagnosis of observation, background and analysis-error statistics in observation space, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 3385–3396.

[36] M. Fisher, Background error covariance modelling, in: *ECMWF seminar on recent developments in data assimilation for atmosphere and ocean*, pp. 45–63, 2003.

[37] _____, *Accounting for correlated observation error in the ECMWF analysis*, ECMWF, Technical memorandum MF/05106, 2005.

[38] M. Fisher and D.J. Lary, Lagrangian four-dimensional variational data assimilation of chemical species, *Quart. J. Roy. Meteor. Soc.* 121 (1995), 1681–1704.

[39] M. Fisher, J. Nocedal, Y. Trémolet and S.J. Wright, Data assimilation in weather forecasting: a case study in PDE-constrained optimization, *Optimization and Engineering* 10 (2009), 409–426.

[40] M. Fisher, Y. Trémolet, H. Auvinen, D. Tan and P. Poli, *Weak-constraint and long-window 4D-Var*, ECMWF, Technical memorandum 655, 2011.

[41] S.J. Fletcher and M. Zupanski, A data assimilation method for log-normally distributed observational errors, *Quart. J. Roy. Meteor. Soc.* 132 (2007), 2505–2519.

[42] P. Gauthier, C. Charette, L. Fillion, P. Koclas and S. Laroche, Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis, *Atmosphere-Ocean* 37 (1999), 103–156.

[43] P. Gauthier, M. Tanguay, S. Laroche, S. Pellerin and J. Morneau, Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada, *Mon. Wea. Rev.* 135 (2007), 2339–2354.

[44] R. Giering and T. Kaminski, Recipes for Adjoint Code Construction, *ACM Trans. On Math. Software* 24 (1998), 437–474.

[45] S. Gratton, A.S. Lawless and N.K. Nichols, Approximate Gauss-Newton methods for nonlinear least squares problems, *SIAM J. Optim.* 18 (2007), 106–132.

[46] A.K. Griffith and N.K Nichols, Adjoint methods in data assimilation for estimating model error, *Flow, Turbulence and Combustion* 65 (2000), 469–488.

[47] V. Guidard and C. Fischer, Introducing the coupling information in a limited-area variational assimilation, *Quart. J. Roy. Meteor. Soc.* 134 (2008), 723–735.

[48] N. Gustafsson, L. Berre, S. Hörnquist, X.Y. Huang, M. Lindskog, B. Navascués, K.S. Mogensen and S. Thorsteinsson, Three-dimensional variational data assimilation for a limited area model, *Tellus A* 53 (2001), 425–446.

[49] N. Gustafsson, X.Y. Huang, X. Yang, K. Mogensen, M. Lindskog, O. Vignes, T. Wilhelmsson and S. Thorsteinsson, Four-dimensional variational data assimilation for a limited area model, *Tellus A* 64 (2012), 14985.

[50] S.A. Haben, *Conditioning and preconditioning of the minimisation problem in variational data assimilation*, Ph.D. thesis, Department of Mathematics and Statistics, University of Reading, 2011.

[51] S.A. Haben, A.S. Lawless and N.K. Nichols, Conditioning and preconditioning of the variational data assimilation problem, *Computers & Fluids* 46 (2011), 252–256.

[52] ———, Conditioning of incremental variational data assimilation, with application to the Met Office system, *Tellus A* 63 (2011), 782–792.

[53] S.J. Hammarling, Numerical Solution of the Stable, Non-negative Definite Lyapunov Equation Lyapunov Equation, *IMA Journal of Numerical Analysis* 2 (1982), 303–323.

[54] B.A. Harris and G. Kelly, A satellite radiance-bias correction scheme for data assimilation, *Quart. J. Roy. Meteor. Soc.* 127 (2001), 1453–1468.

[55] S.B. Healy and A.A. White, Use of discrete Fourier transforms in the 1D-Var retrieval problem, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 63–72.

[56] F. Hilton, N.C. Atkinson, S.J. English and J.R. Eyre, Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments, *Quart. J. Roy. Meteor. Soc.* 135 (2009), 495–505.

[57] A. Hollingsworth and P. Lönnberg, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus A* 38 (1986), 111–136.

[58] J.R. Holton, *An introduction to dynamic meteorology*, Elsevier Academic Press, 2004.

[59] B.J. Hoskins, M.E. McIntyre and A.W. Robertson, On the use and significance of isentropic potential vorticity maps, *Quart. J. Roy. Meteor. Soc.* 111 (1995), 877–946.

[60] I. Hoteit and A. Köhl, Efficiency of reduced-order, time-dependent adjoint data assimilation approaches, *Journal of oceanography* 62 (2006), 539–550.

[61] X.Y. Huang, Q. Xiao, D.M. Barker, X. Zhang, J. Michalakes, W. Huang, T. Henderson, J. Bray, Y. Chen, Z. Ma, J. Dudhia, Y. Guo, X. Zhang, D-J. Won, H-C. Lin and Y-H. Kuo, Four-dimensional variational data assimilation for WRF: Formulation and preliminary results, *Mon. Wea. Rev.* 137 (2009), 299–314.

[62] N.B. Ingleby, The statistical structure of forecast errors and its representation in the Met Office Global 3-Dimensional variational data assimilation scheme, *Quart. J. Roy. Meteor. Soc.* 127 (2001), 209–231.

[63] N.B. Ingleby and A.C. Lorenc, Bayesian quality control using multivariate normal distributions, *Quart. J. Roy. Meteor. Soc.* 119 (1993), 1195–1225.

[64] M. Janiskova, J-N. Thepaut and J-F. Geleyn, Simplified and regular physical parametrizations for incremental four-dimensional variational assimilation, *Mon. Wea. Rev.* 127 (1999), 26–45.

[65] C. Johnson, B.J. Hoskins and N.K. Nichols, A singular vector perspective of 4D-Var: Filtering and interpolation, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 1–19.

[66] D. Katz, A.S. Lawless, N.K. Nichols, M.J.P. Cullen and R.N. Bannister, Correlations of control variables in variational data assimilation, *Quart. J. Roy. Meteor. Soc.* 137 (2011), 620–630.

[67] T. Kawabata, H. Seko, K. Saito, T. Kuroda, K. Tamiya, T. Tsuyuki, Y. Honda and Y. Wakazuki, An Assimilation and Forecasting Experiment of the Nerima Heavy Rainfall with a Cloud-Resolving Nonhydrostatic 4-Dimensional Variational Data Assimilation System, *J. Met. Soc. Japan. Ser. II* 85 (2007), 255–276.

[68] S. Laroche and P. Gauthier, A validation of the incremental formulation of 4D variational data assimilation in a nonlinear barotropic flow, *Tellus* 50A (1998), 557–572.

[69] A. Laub, M. Heath, C. Paige and R. Ward, Computation of system balancing transformations and other applications of simultaneous diagonalization algorithms, *Automatic Control, IEEE Transactions on* 32 (1987), 115–122.

[70] A.S. Lawless, A note on the analysis error associated with 3D-FGAT, *Quart. J. Roy. Meteor. Soc.* 136 (2010), 1094–1098.

[71] A.S. Lawless, S. Gratton and N.K. Nichols, Approximate iterative methods for variational data assimilation, *International journal for numerical methods in fluids* 47 (2005), 1129–1135.

[72] _____, An investigation of incremental 4D-Var using non-tangent linear models, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 459–476.

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

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

[75] A.S. Lawless, N.K. Nichols, C. Boess and A. Bunse-Gerstner, Approximate Gauss–Newton methods for optimal state estimation using reduced-order models, *International journal for numerical methods in fluids* 56 (2008), 1367–1373.

[76] _____, Using model reduction methods within incremental four-dimensional variational data assimilation, *Mon. Wea. Rev.* 136 (2008), 1511–1522.

[77] F.-X. Le Dimet and O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus* 38A (1986), 97–110.

[78] D.J. Lea, J.P. Drecourt, K. Haines and M.J. Martin, Ocean altimeter assimilation with observational- and model-bias correction, *Quart. J. Roy. Meteor. Soc.* 134 (2008), 1761–1774.

[79] Y. Li, I.M. Navon, W. Yang, X. Zou, J.R. Bates, S. Moorthi and R.W. Higgins, Four-dimensional variational data assimilation experiments with a multilevel semi-Lagrangian semi-implicit general circulation model, *Mon. Wea. Rev.* 122 (1994), 966–983.

[80] M. Lindskog, D. Dee, Y. Trémolet, E. Andersson, G. Radnóti and M. Fisher, A weak-constraint four-dimensional variational analysis system in the stratosphere, *Quart. J. Roy. Meteor. Soc.* 135 (2009), 695–706.

[81] C. Liu, Q. Xiao and B. Wang, An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test, *Mon. Wea. Rev.* 136 (2008), 3363–3373.

[82] A.C. Lorenc, Analysis methods for numerical weather prediction, *Quart. J. Roy. Meteor. Soc.* 112 (1986), 1177–1194.

[83] ———, Development of an operational variational assimilation scheme, *J. Met. Soc. Japan* 75 (1997), 339–346.

[84] A.C. Lorenc, S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Barker, J.R. Bray, A.M. Clayton, T. Dalby and D. Li, The Met. Office global 3-dimensional variational data assimilation scheme, *Quart. J. Roy. Meteor. Soc* 126 (2000), 2991–3012.

[85] A.C. Lorenc, R.B. Bell and B. Macpherson, The Meteorological Office analysis correction data assimilation scheme, *Quart. J. Roy. Meteor. Soc.* 117 (1991), 59–89.

[86] A.C. Lorenc and T. Payne, 4D-Var and the butterfly effect: Statistical four-dimensional data assimilation for a wide range of scales, *Quart. J. Roy. Meteor. Soc.* 133 (2007), 607–614.

[87] A.C. Lorenc and F. Rawlins, Why does 4D-Var beat 3D-Var?, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 3247–3257.

[88] J.F. Mahfouf and F. Rabier, The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics, *Quart. J. Roy. Meteor. Soc.* 126 (2000), 1171–1190.

[89] T. Mochizuki, N. Sugiura, T. Awaji and T. Toyoda, Seasonal climate modeling over the Indian Ocean by employing a 4D-VAR coupled data assimilation approach, *Journal of Geophysical Research* 114 (2009), C11003.

[90] A.J.F. Moodey, A.S. Lawless, R.W.E. Potthast and P.J. van Leeuwen, Nonlinear error dynamics for cycled data assimilation methods, *Inverse Problems* 29 (2013), 025002.

[91] A.M. Moore, H.G. Arango, G. Broquet, B.S. Powell, A.T. Weaver and J. Zavala-Garay, The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems Part I–System overview and formulation, *Progress in Oceanography* 91 (2011), 34–49.

[92] B. Moore, Principal component analysis in linear systems: Controllability, observability, and model reduction, *Automatic Control, IEEE Transactions on* 26 (1981), 17–32.

[93] I.M. Navon, X. Zou, J. Derber and J. Sela, Variational data assimilation with an adiabatic version of the NMC spectral model, *Mon. Wea. Rev.* 120 (1992), 1433–1446.

[94] J. Nocedal and S.J. Wright, *Numerical optimization*, Springer Verlag, 2006.

[95] T.N. Palmer, R. Gelaro, J. Barkmeijer and R. Buizza, Singular vectors, metrics and adaptive observations, *J. Atmos. Sci.* 55 (1998), 633–653.

[96] S.K. Park and D. Zupanski, Four-dimensional variational data assimilation for mesoscale and storm-scale applications, *Meteorology and Atmospheric Physics* 82 (2003), 173–208.

[97] D.F. Parrish and J.C. Derber, The National Meteorological Center's spectral statistical-interpolation analysis system, *Mon. Wea. Rev.* 120 (1992), 1747–1763.

[98] F. Rabier and P. Courtier, Four-dimensional assimilation in the presence of baroclinic instability, *Quart. J. Roy. Meteor. Soc.* 118 (1992), 649–672.

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

[100] F. Rawlins, S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc and T.J. Payne, The Met Office global four-dimensional variational data assimilation scheme, *Quart. J. Roy. Meteor. Soc.* 133 (2007), 347–362.

[101] C. Robert, S. Durbiano, E. Blayo, J. Verron, J. Blum and F.X. Le Dimet, A reduced-order strategy for 4D-Var data assimilation, *Journal of Marine Systems* 57 (2005), 70–82.

[102] N. Rostaing, S. Dalmas and A. Galligo, Automatic Differentiation in Odyssée, *Tellus* 45A (1993), 558–568.

[103] Y. Sasaki, An objective analysis based on the variational method, *J. Met. Soc. Japan* 36 (1958), 77–88.

[104] _____, Some basic formalisms in numerical variational analysis, *Mon. Wea. Rev.* 98 (1970), 875–883.

[105] L.M. Stewart, *Correlated observation errors in data assimilation*, Ph.D. thesis, Department of Mathematics, University of Reading, 2010.

[106] N. Sugiura, T. Awaji, S. Masuda, T. Mochizuki, T. Toyoda, T. Miyama, H. Igarashi and Y. Ishikawa, Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations, *Journal of Geophysical Research* 113 (2008), C10017.

[107] O. Talagrand and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Quart. J. Roy. Meteor. Soc.* 113 (1987), 1311–1328.

[108] J-N. Thepaut and P. Courtier, Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model, *Quart. J. Roy. Meteor. Soc.* 117 (1991), 1225–1254.

[109] Y. Trémolet, Accounting for an imperfect model in 4D-Var, *Quart. J. Roy. Meteor. Soc.* 132 (2006), 2483–2504.

[110] _____, Model-error estimation in 4D-Var, *Quart. J. Roy. Meteor. Soc.* 133 (2007), 1267–1280.

[111] J. Tshimanga, S. Gratton, A.T. Weaver and A. Sartenaer, Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation, *Quart. J. Roy. Meteor. Soc.* 134 (2008), 751–769.

[112] J. Vialard, A.T. Weaver, D.L.T. Anderson and P. Delecluse, Three-and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part II: Physical validation, *Mon. Wea. Rev.* 131 (2003), 1379–1395.

[113] A. Weaver and P. Courtier, Correlation modelling on the sphere using a generalized diffusion equation, *Quart. J. Roy. Meteor. Soc.* 127 (2001), 1815–1846.

[114] A.T. Weaver, C. Deltel, E. Machu, S. Ricci and N. Daget, A multivariate balance operator for variational ocean data assimilation, *Quart. J. Roy. Meteor. Soc.* 131 (2005), 3605–3625.

[115] A.T. Weaver and I. Mirouze, On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation, *Quart. J. Roy. Meteor. Soc.* (2012), doi: 10.1002/qj.1955.

[116] A.T. Weaver, J. Vialard and D.L.T. Anderson, Three-and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part I: Formulation, internal diagnostics, and consistency checks, *Mon. Wea. Rev.* 131 (2003), 1360–1378.

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

[118] Q. Xu, Generalized adjoint for physical processes with parameterized discontinuities. Part I: Basic issues and heuristic examples, *J. Atmos. Sci.* 53 (1996), 1123–1155.

[119] D. Zupanski, A general weak constraint applicable to operational 4DVAR data assimilation systems, *Mon. Wea. Rev.* 125 (1997), 2274–2292.

Author information

Amos S. Lawless, School of Mathematical and Physical Sciences, University of Reading, PO Box 220, Whiteknights, Reading, RG6 6AX, U.K..
E-mail: a.s.lawless@reading.ac.uk