Department of Mathematics
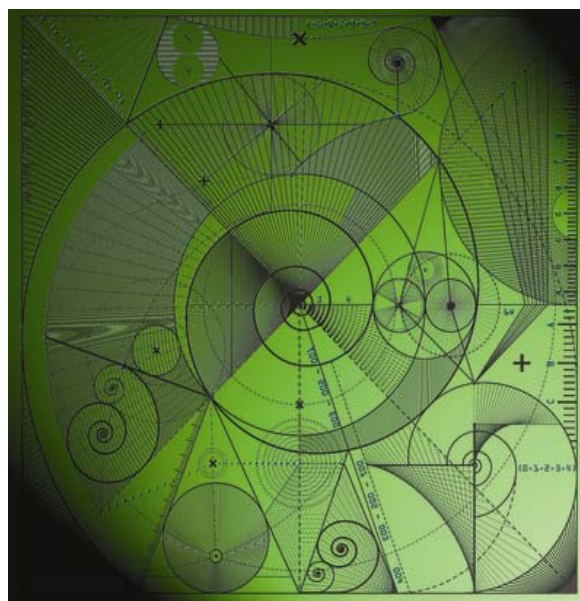
Conditioning of incremental variational data assimilation, with application to the Met Office system

by

S.A. Haben, A.S. Lawless and N.K. Nichols

Conditioning of incremental variational data assimilation, with application to the Met Office system

By S.A. Haben, A.S. Lawless* and N.K. Nichols
Department of Mathematics, University of Reading, Reading, U.K.

*Corresponding author. E-mail: a.s.lawless@reading.ac.uk

Abstract

Implementations of incremental variational data assimilation require the iterative minimization of a series of linear cost functions. The accuracy and speed with which these linear minimization problems can be solved is determined by the condition number of the Hessian of the problem. In this study we examine how different components of the assimilation system influence this condition number. Theoretical bounds on the condition number for a single parameter system are presented and used to predict how the condition number is affected by the observation distribution and accuracy and by the specified lengthscales in the background error covariance matrix. The theoretical results are verified in the Met Office variational data assimilation system, using both pseudo-observations and real data.

1   Introduction

An important component of numerical weather prediction (NWP) systems is the determination of an appropriate set of initial conditions from observations by means of techniques of data assimilation. In general observations of the atmosphere are only indirectly related to model variables and are many fewer in number than the number of model states that need to be initialized. Data assimilation techniques aim to combine these measurements with a previous forecast, or background field, in order to provide the best estimate of the model state given all the available information. One such technique is that of variational data assimilation, which defines the best estimate as that which minimizes a nonlinear least squares cost function representing the error between the solution and the observations and the error between the solution and the background. The weights in the cost function are defined by appropriate covariance matrices. Under certain assumptions this ensures the solution provides the maximum a posteriori Bayesian estimate of the system state (Lorenc, 1988).

Variational data assimilation is the method of choice for many current NWP systems, including that of the Met Office, the European Centre for Medium-range Weather Forecasting (ECMWF) and several other centres (Rawlins *et al.*, 2007; Rabier *et al.*, 2000; Gauthier and Thépaut, 2001; Gauthier *et al.*, 2007; Huang *et al.*, 2009). It is usually implemented using the incremental formulation, which approximates the minimization of the nonlinear cost function by a sequence of minimizations of linearized cost functions and is equivalent to a Gauss-Newton procedure (Courtier *et al.*, 1994; Lawless *et al.*, 2005). This has the advantage of allowing further approximations in the solution procedure in order to make the problem computationally feasible. The minimization of each linearized problem is performed using an iterative method, such as a conjugate gradient technique.

The speed and accuracy with which a solution to each linear minimization problem can be found is determined by the condition number of the Hessian of the linearized problem. A large condition number implies that the solution will be very sensitive to small changes in the data and iterative methods may need many iterations to reach the solution. Since in practice the number of iterations that can be performed is limited by operational time requirements, it is important that the system is well-conditioned, so that a good solution can be found in relatively few iterations.

Operationally the linear minimization problem is preconditioned by formulating the problem in variables whose errors are assumed to be uncorrelated. Bounds on the conditioning of the preconditioned and unpreconditioned systems found in Haben *et al.* (2009) for a one-parameter periodic system indicated that the preconditioned system has, in general, a signicantly reduced condition number compared to that of the unpreconditioned system. This has been supported with experiments using the operational systems at the Canadian Meteorological centre and the UK Met Office that have shown that preconditioning significantly improves the rate of convergence of the inner-loop minimisation (Lorenc, 1997; Gauthier *et al.*, 1999). Once the system has been transformed using this variable transformation, further preconditioning techniques may then be applied.

The main factors that influence the conditioning of the preconditioned system have only been partially studied. In Andersson *et al.* (2000) it was shown for a simple 2 grid point system with multiple observations that the condition number of the preconditioned system was proportional to the background error variance and the number of observations at the grid points, while being inversely proportional to the error variances on the observations. An experiment with the ECMWF minimisation scheme showed, after establishing that the dense, accurate surface observations over Europe dominated the conditioning, that doubling the observation error variance reduced the conditioning of the system, in rough agreement with the results from the simple model (Trémolet, 2007). In Haben *et al.* (2009) the effect of the observation error on the condition number was confirmed for a more general case and it was also shown that increasing the separation of the observations causes a reduction in the condition number.

In this paper we present theory for the conditioning of the variational data assimilation problem and use it to interpret the conditioning of the Met Of-

2

fice assimilation scheme. In particular we show, using both theory and the operational system, that the conditioning of variational data assimilation is dependent on the spacing and error variance of the observations. We begin in section 2 where we set out the variational assimilation problem. In section 3 we present theoretical results concerning the conditioning of the problem and illustrate them using a simple example. The conditioning of the Met Office system is investigated in section 4, using both pseudo-observations and real observations. We show how the condition number is dominated by the dense, surface observations over Europe and how this can be explained using the theory we have developed. Finally we make some concluding remarks in section 5.

2 Background

2.1 Incremental variational data assimilation

In four dimensional variational data assimilation (4DVar) we assume that we have observations $\mathbf{y}_i$ at times $t_i, i = 0, \ldots, n$ over a time window $[t_0, t_n]$, with observation error covariance matrices $\mathbf{R}_i$, and a background state $\mathbf{x}_0^b$ at the initial time t_0, with background error covariance matrix $\mathbf{B}$. Then the analysis is defined as the state which minimizes the nonlinear cost function

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

subject to the states $\mathbf{x}_i$ satisfying the nonlinear forecast model

$$
\mathbf{x}_i = \mathcal{M}(t_i, t_{i-1}, \mathbf{x}_{i-1}), \tag{2}
$$

for $i = 1, \ldots, n$. The operator $\mathcal{H}_i$ is the observation operator that maps the model state to observation space at time t_i. For the case where $n = 0$ and all the observations are at the initial time, the scheme is known as three-dimensional variational data assimilation (3DVar) and the model equation (2) plays no part in the minimization problem.

In many operational NWP centres variational data assimilation has been implemented using the incremental formulation, in which the minimization of (1) is replaced by the minimization of a sequence of linearized cost functions (Courtier *et al.*, 1994). The nonlinear cost function (1) is linearized about the model trajectory $\mathbf{x}_i, i = 1, \ldots, n$ produced from the current estimate of the initial state $\mathbf{x}_0$. This linearized cost function is minimized to find an increment $\delta\mathbf{x}_0$ at time t_0, in what is known as an inner loop step, and the increment is then used to improve the estimate of $\mathbf{x}_0$ in an outer loop step. This method is equivalent to applying an approximate Gauss-Newton method to solve the original nonlinear problem (Lawless *et al.*, 2005; Gratton *et al.*, 2007).

In the case of incremental 4DVar the linearized minimization problem takes the form

$$
\begin{aligned}
\tilde{J}(\delta\mathbf{x}_0) &= \frac{1}{2}[\delta\mathbf{x}_0 - (\mathbf{x}_0^b - \mathbf{x}_0)]^T \mathbf{B}^{-1}[\delta\mathbf{x}_0 - (\mathbf{x}_0^b - \mathbf{x}_0)] \\
&+ \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\delta\mathbf{x}_i - \mathbf{d}_i)^T \mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i - \mathbf{d}_i),
\end{aligned}
\tag{3}
$$

subject to the perturbations $\delta\mathbf{x}_i$ satisfying the linearized model equations

$$
\delta\mathbf{x}_i = \mathbf{M}(t_i, t_{i-1})\,\delta\mathbf{x}_{i-1},
\tag{4}
$$

where $\mathbf{d}_i = \mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)$ and the operators $\mathbf{H}_i$ and $\mathbf{M}(t_i, t_{i-1})$ are the Jacobians of the nonlinear operators $\mathcal{H}_i$ and $\mathcal{M}(t_i, t_{i-1}, \mathbf{x}_{i-1})$ respectively, calculated at the current estimate of the trajectory. For the case of 3DVar we have $n = 0$ and there is no evolution of the perturbation by the linear model (4).

In practice it is not possible to formulate and minimize the linear cost function (3) directly. The background error covariance matrix $\mathbf{B}$ cannot be represented explicitly and must be built implicitly. Furthermore, the minimization of the cost function as written has been shown to give slow convergence in practice (Lorenc, 1997; Gauthier *et al.*, 1999) and its ill-conditioning has also been demonstrated theoretically for common covariance structures (Haben *et al.*, 2009, 2010). For these reasons a variable transformation is defined of the form $\delta\mathbf{z} = \mathbf{U}\delta\mathbf{x}$, where the variables $\delta\mathbf{z}$ and the transformation $\mathbf{U}$ are chosen to ensure that the new variables are uncorrelated (Lorenc *et al.*, 2000; Cullen, 2003; Bannister, 2008; Katz *et al.*, 2010). Implicitly this is equivalent to preconditioning by $\mathbf{U}^{-1} = \mathbf{B}^{1/2}$ and the cost function in the transformed variables is given by

$$
\begin{aligned}
\hat{J}[\delta\mathbf{z}_0] &= \frac{1}{2}[\delta\mathbf{z}_0 - (\mathbf{z}_0^b - \mathbf{z}_0)]^T[\delta\mathbf{z}_0 - (\mathbf{z}_0^b - \mathbf{z}_0)] \\
&+ \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\mathbf{B}^{1/2}\delta\mathbf{z}_i - \mathbf{d}_i)^T \mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{B}^{1/2}\delta\mathbf{z}_i - \mathbf{d}_i),
\end{aligned}
\tag{5}
$$

where $\mathbf{z}_0^b = \mathbf{B}^{-1/2}\mathbf{x}_0^b$, $\mathbf{z}_0 = \mathbf{B}^{-1/2}\mathbf{x}_0$ and $\delta\mathbf{z}_i = \mathbf{B}^{-1/2}\delta\mathbf{x}_i$.

2.2 Conditioning of the variational problem

The accuracy and efficiency with which the inner minimization problem can be solved is determined by the condition number of the Hessian of the cost function. In this study we consider the condition number of a matrix defined in the matrix 2-norm, which for a symmetric matrix is equal to the ratio of the largest and smallest eigenvalues of the matrix. The condition number measures the sensitivity of the solution to small perturbations in the input data (Golub and Van Loan, 1996). If the condition number is large then small changes in the inputs can lead to large variations in the solutions. Moreover, a large condition number can lead to a slow rate of convergence of iterative minimization methods.

For example, the rate of convergence of the conjugate gradient method can be bounded in terms of the condition number of the Hessian (Golub and Van Loan, 1996, Theorem 10.2.6).

For the preconditioned variational cost function (5) the Hessian is given by the expression

$$\mathbf{S} = \mathbf{I} + \sum_{i=0}^{n} \mathbf{B}^{1/2}\mathbf{M}(t_i, t_0)^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{H}_i \mathbf{M}(t_i, t_0)\mathbf{B}^{1/2}, \qquad (6)$$

where $\mathbf{M}(t_i, t_0) = \mathbf{M}(t_i, t_{i-1})\mathbf{M}(t_{i-1}, t_{i-2})\ldots\mathbf{M}(t_1, t_0)$. In general we have fewer observations than variables we are trying to estimate and so the second term in the expression (6) is not full rank. In this case the smallest eigenvalue of $\mathbf{S}$ is one and the condition number is equal to the largest eigenvalue. We now investigate the different factors that affect the conditioning of the preconditioned Hessian. We first present the conditioning theory using a simple system.

3 Conditioning for a single parameter system

3.1 Theory

We consider the case of 3DVar applied to a single parameter system on a one-dimensional periodic domain, discretized using N grid points $\xi_i, i = 1, \ldots, N$. For this case the preconditioned Hessian (6) reduces to

$$\mathbf{S} = \mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T \mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}. \qquad (7)$$

We assume we have a set of $p < N$ direct observations of the parameter at grid points, so that $\mathbf{H}^T\mathbf{H}$ is a diagonal matrix, where the diagonal element is equal to one if that component of the state is observed and zero otherwise. We further assume the observation errors are uncorrelated with error variance σ_o^2, which implies that the observation error covariance matrix $\mathbf{R} = \sigma_o^2\mathbf{I}_p$. We write the background covariance matrix in the form $\mathbf{B} = \sigma_b^2\mathbf{C}$ where $\mathbf{C}$ is the background error correlation matrix with components $c_{i,j}$ and σ_b^2 denotes the background error variance. We assume that the correlation structures are homogeneous and isotropic, so that the coefficients $c_{i,j}$ depend only on the distance between points ξ_i and ξ_j. Then, under the assumptions given we can show that the condition number of the matrix (7) satisfies

$$
\begin{aligned}
1 + \frac{\sigma_b^2}{\sigma_o^2}\beta &\leq \kappa(\mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T \mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}) \\
&\leq 1 + \frac{\sigma_b^2}{\sigma_o^2}||\mathbf{H}\mathbf{C}\mathbf{H}^T||_\infty \\
&= 1 + \frac{\sigma_b^2}{\sigma_o^2}\left(\max_{i \in J}\sum_{j \in J}|c_{i,j}|\right),
\end{aligned}
\qquad (8)
$$

where $\beta = \frac{1}{p} \sum_{i,j \in J} c_{i,j}$, and J is the set of indices of the variables that are observed. A proof of this result is given in Haben *et al.* (2009).

From this bound we can understand the effect of the observation accuracy and spacing on the conditioning of the preconditioned 3DVar problem. We note first that, as the observations become more accurate and σ_o^2 decreases, both the upper and lower bounds in (8) increase. Hence an increase in accuracy of the observations results in a more poorly conditioned Hessian. Furthermore, the condition number of the Hessian is of the form

$$
\begin{aligned}
\kappa(\mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}) &\equiv \lambda_{\max}(\mathbf{I}_N + \mathbf{B}^{1/2}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}) \\
&= 1 + \left(\frac{\sigma_b^2}{\sigma_o^2}\right)\lambda_{\max}(\mathbf{C}^{1/2}\mathbf{H}^T\mathbf{H}\mathbf{C}^{1/2}), \quad (9)
\end{aligned}
$$

where $\lambda_{\max}$ indicates the largest eigenvalue of the matrix. In the case where the second term in (9) is much greater than one we see that the condition number itself is approximately proportional to the inverse of the observation error variance. Thus as the accuracy of the observations increases by a given factor, with the background error variance fixed, we would expect the condition number of the Hessian also to increase by the same factor. Similarly, for a set of observations with a given accuracy, we expect the condition number to decrease as the background becomes more accurate.

We also note that the upper bound on the condition number is dependent on the infinity norm of the matrix $\mathbf{H}\mathbf{C}\mathbf{H}^T$, which is defined as the maximum absolute row sum of the matrix. Since in this example we assume that observations are at the grid points, then the matrix $\mathbf{H}\mathbf{C}\mathbf{H}^T$ is a reduced version of the correlation matrix $\mathbf{C}$, which is formed by deleting those rows and columns that are not observed. If we assume that the correlations between grid points decrease with distance, then the coefficients of this reduced matrix will be smaller the farther apart the observations are. Hence, as the spacing between observations increases, we expect the condition number of the preconditioned Hessian to decrease. Similarly, when we have fewer observations then the infinity norm of the reduced correlation matrix will be smaller, since fewer coefficients appear in the sum, and the condition number will be smaller. On the other hand, as the observations become more dense or the number of observations increases, the upper bound on the condition number will become larger. For forms of the covariance matrix in which the coefficients are all positive, the lower bound will also increase. Thus in this case we would expect the problem to become more ill-conditioned. We note that these bounds can also be generalised to take into account more than one observation at each of the observed grid points. They then include as a special case the result of Andersson *et al.* (2000), who considered the effect of observation density when just two grid points are analysed.

From this analysis we can also expect the condition number of the Hessian to increase as the correlation lengthscales in the background error covariance matrix increase. The correlation lengthscales determine the distance over which errors in the background field are correlated. For larger lengthscales the coefficients of the background error covariance matrix are larger and so both the
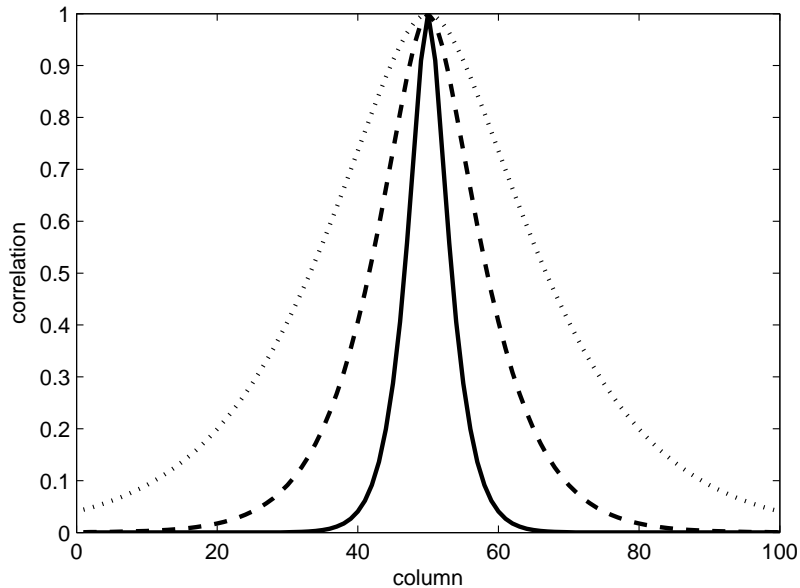
Figure 1: Row 50 of the correlation matrix generated using the SOAR function with a grid length of 0.1 for lengthscales $L = 0.2$ (Solid line), $L = 0.5$ (dashed line) and $L = 1$ (dotted line).

coefficient β and the infinity norm of $\mathbf{HCH}^T$ will be larger. Thus both the lower and upper bounds will increase and so the conditioning of the problem will worsen for larger correlation lengthscales.

3.2   Example using the SOAR function

In order to illustrate our theory using a specific example we consider the single parameter system described in section 3.1, where the background error correlation matrix $\mathbf{C}$ is given by the second-order auto-regressive (SOAR) correlation matrix. This correlation structure has been used in operational data assimilation systems, for example the Met Office 3DVar system (Ingleby, 2001). For a set of N discrete points on a 1-dimensional periodic domain with equal spacing Δx, the coefficients of the matrix are given by

$$c_{i,j} = \left(1 + \frac{r_{i,j}}{L}\right) \exp\left(-\frac{r_{i,j}}{L}\right), \tag{10}$$

where $r_{i,j} = \Delta x |i - j|$ is the distance between the grid points at position i and j with $|i - j| < N/2$. The remaining coefficients are determined by the periodicity of the domain. We illustrate the shape of the correlation function in
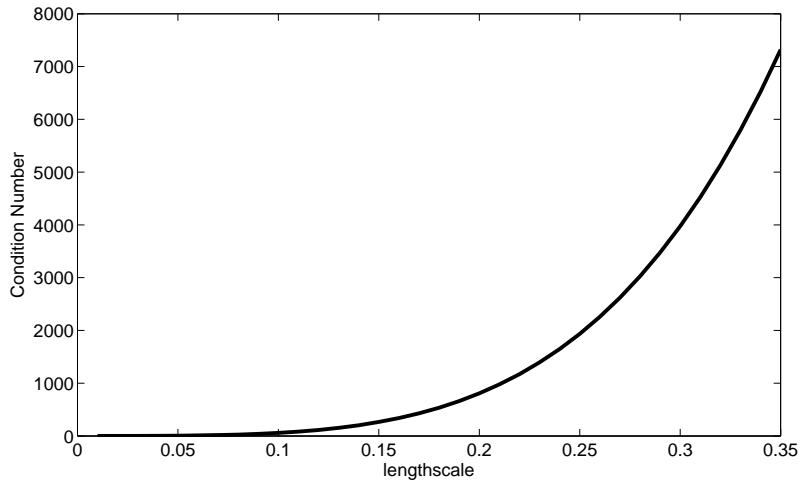
7

Figure 2: Variation of the condition number of the SOAR correlation matrix with lengthscale.

Figure 1 by plotting the 50th row of the matrix for a system of size 100 using different correlation lengthscales L. The grid length is fixed at $\Delta x = 0.1$. We see that the function has the form of a bell shape, whose width depends on the lengthscale. Furthermore, we see that this correlation function describes only positive correlations and so the absolute value of the matrix coefficients are a monotonically decreasing function of the distance between the points. In Figure 2 we show the condition number of this covariance matrix as a function of the correlation lengthscale. As the lengthscale is increased there is a sharp growth in the condition number of the matrix. This has also been found to be true for other commonly used correlation matrices (Haben *et al.*, 2009).

In order to illustrate the effect of the observation accuracy on the condition number of the Hessian we define a periodic domain of 500 grid points with 20 equally spaced observations made every 25 grid points. The lengthscale is fixed to be 0.2, the grid length to be $\Delta x = 0.1$ and the background error variance is fixed at $\sigma_b^2 = 1$. The condition number of the Hessian (7) for different values of the observation error variance is shown in table 1. We see that as predicted from the bounds (8), the condition number of the Hessian increases as the observation error variance decreases. We also find that for condition numbers much larger than one the condition number is approximately inversely proportional to the errors on the observations. For example, we see that when the observation error variance is halved from 0.1 to 0.05 then the condition number is approximately doubled.

We now consider the effect of the spacing of the observations on the condition number. We fix the observation and background error variances to be $\sigma_o^2 = \sigma_b^2 = $

8

Variance σ_o^2	Condition Number
0.01	101.0
0.05	21.0
0.1	11.0
0.5	3.0
1	2.0
2	1.5
5	1.2
10	1.1

Table 1: Condition number of the preconditioned Hessian of the single parameter system with changing observation variance. The first column shows the observation error variance and the second column shows the condition number.

1 and reduce the number of observations to be at only four equally spaced points. The observations are initially placed in the centre of the domain with a spacing of one grid point between them. In Figure 3 we plot the variation in the condition number of the Hessian as the spacing between the observations is increased from one to twelve grid points for three different values of the background error correlation lengthscale L. As expected from the discussion in section 3.1, the condition number decreases as the spacing between the observation increases. Furthermore, the graphs confirm that for a fixed value of the observation error variance, the condition number increases with the lengthscales specified in the correlation matrix of the background errors.

3.3 Extension to 4DVar

The theory of section 3.1 can be easily extended to the case of 4DVar. If we define the matrix

$$\hat{\mathbf{H}} = [\mathbf{H}_0^T, (\mathbf{H}_1 \mathbf{M}(t_1, t_0))^T, \ldots, (\mathbf{H}_n \mathbf{M}(t_n, t_0))^T]^T \qquad (11)$$

and define the matrix $\hat{\mathbf{R}}$ to be a block diagonal matrix with diagonal blocks equal to $\mathbf{R}_i$, then the Hessian (6) of the preconditioned 4DVar system can be written

$$\mathbf{S} = \mathbf{I} + \mathbf{B}^{1/2} \hat{\mathbf{H}} \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}}^T \mathbf{B}^{1/2}. \qquad (12)$$

We assume that we have the same single parameter system as described in section 3.1, but with observations taken at $n + 1$ time steps, from time t_0 to time t_n. We furthermore assume that the same observations are made at each step, each with observation error variance σ_o^2, so that we can write $\mathbf{H}_i = \mathbf{H}$ for $i = 0, \ldots, n$ and $\hat{\mathbf{R}} = \sigma_o^2 \mathbf{I}_{(n+1)p}$. Then the theory of Haben *et al.* (2009) can be applied to the 4DVar Hessian (12) to obtain the following bounds on the
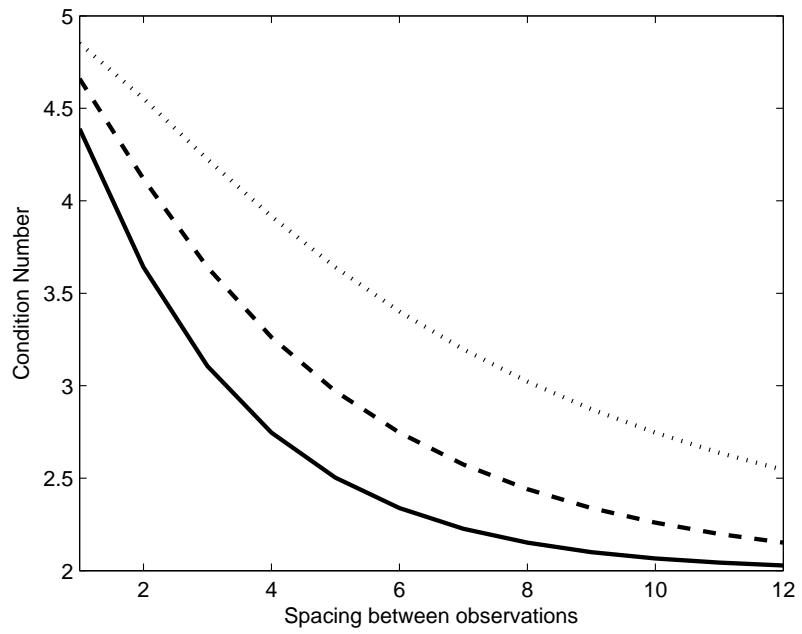
9

Figure 3: Condition number of the Hessian of the single parameter system against observation spacing using the SOAR background covariance matrix with correlation lengthscales $L = 0.2$ (Solid line), $L = 0.3$ (dashed line) and $L = 0.5$ (dotted line).

condition number,

$$1 + \frac{1}{(n+1)p} \frac{\sigma_b^2}{\sigma_o^2} \sum_{i,j=1}^{(n+1)p} (\hat{\mathbf{H}} \mathbf{C} \hat{\mathbf{H}}^T)_{i,j}$$

$$\leq \kappa(\mathbf{S}) \leq 1 + \frac{\sigma_b^2}{\sigma_o^2} ||\hat{\mathbf{H}} \mathbf{C} \hat{\mathbf{H}}^T||_\infty. \qquad (13)$$

A comparison of this expression with (8) shows that the bounds on the condition number of the 4DVar Hessian for this system are very similar to those of the 3DVar Hessian, with the matrix $\hat{\mathbf{H}}$ taking the place of the linear observation operator $\mathbf{H}$. Hence many of the qualitative conclusions discussed in section 3.1 for the 3DVar case also apply in the 4DVar case. In particular, as the accuracy of the observations increases then both the lower and upper bounds of (13) will increase and so the conditioning of the system will worsen.

Furthermore, we note that the matrix $\hat{\mathbf{H}} \mathbf{C} \hat{\mathbf{H}}^T$ which appears in both the lower and upper bounds can also be written in the form $\tilde{\mathbf{H}} \tilde{\mathbf{C}} \tilde{\mathbf{H}}^T$, where $\tilde{\mathbf{H}}$ is a block diagonal matrix of $n + 1$ blocks each equal to $\mathbf{H}$ and $\tilde{\mathbf{C}}$ is the four-dimensional covariance matrix associated with the four-dimensional background state $(\mathbf{x}_0^{b\,T}, \mathbf{x}_1^{b\,T}, \ldots, \mathbf{x}_n^{b\,T})^T$. Thus the matrix $\hat{\mathbf{H}} \mathbf{C} \hat{\mathbf{H}}^T$ is formed by deleting rows and columns from each block of the four-dimensional covariance matrix, where the deleted entries correspond to variables that are not observed. Hence we would expect the bound on the norm of this matrix to depend on the way the covariances evolve with a particular choice of the linear model. If the linear model dynamics act to ensure that the coefficients of the correlation matrix remain positive and remain a decreasing function of distance, then we would expect the bounds on the condition number to decrease as the observations are spaced further apart, as for the 3DVar case. However, as observations are taken at more time levels, the upper bound in (13) will increase, since coefficients at more observation times are being summed, and so the condition number of the problem may also increase. Thus the behaviour of the condition number is likely to be very dependent on the form of the linear model dynamics and on the number of observations. Having considered this scalar example we now turn our attention to the variational data assimilation scheme of the Met Office.

4 Met Office variational assimilation scheme

The variational data assimilation system of the Met Office is based the incremental formulation, preconditioned by means of the square root of the background error covariance matrix, as described in section 2.1 (Rawlins *et al.*, 2007). In the operational system only one outer loop update of the scheme is performed and hence the computational expense is dominated by solving the inner loop problem. Here the inner loop is minimized using a conjugate gradient scheme. The state variables are defined on a global grid of 216 latitudinal points and 163 longitudinal points, giving a horizontal resolution of $10/9 \times 5/3$ degrees.

	Condition number	
Error Variance	3DVar	4DVar
1	152422	180781
10	15243	18078
25	6098	7232
50	3050	3618
75	2033	2412
100	1525	1809

Table 2: Change of condition number with observation error variance in the Met Office 3DVar and 4DVar systems using pseudo-observations.

The largest eigenvalues of the Hessian are calculated using the Lanczos method and, since the Hessian is of the form (6), the condition number is simply equal to the largest eigenvalue. We begin by verifying the theory developed above using pseudo-observations before looking at the effect of real observations in the assimilation system.

4.1 Pseudo-observations

The performance of data assimilation experiments using pseudo-observations has the advantage that we can alter the observation accuracy and density in a controlled manner. A subset of experiments using pseudo-observations in the Met Office 3DVar system was reported in Haben *et al.* (2010). Here we present a more complete set of these results using both 3DVar and 4DVar, in order to link them to experiments using real observations. For the 3DVar case we define a set of 16 surface pressure pseudo-observations and position them in a 4-by-4 grid with one grid spacing between each row and column of 4 observations. The observations are positioned to be in a square approximately covering the U.K. For the 4D-Var case we have 6 surface pressure pseudo-observations at the start, middle and end of a 160-minute time window, with the observations arranged in a 2-by-3 grid over the U.K. The reduced number of observations at each time for the 4D-Var case is determined by the limitation on the total number of pseudo-observations that may be specified in the Met Office system.

We first consider the effect of changing the accuracy of the observations. The position of the observations is fixed and the observation error variances are varied from $1Pa$ to $100Pa$. In table 2 we show the variation in the condition number of the Hessian as the observation error variance is changed. We find that, as predicted by the theory and experiments of section 3, the condition number for both 3DVar and 4DVar increases as the observations become more accurate, with the relationship between them being approximately inversely proportional. For example, a doubling of the variance from $25Pa$ to $50Pa$ approximately halves the condition number of 3DVar from $6,098$ to $3,050$ and of 4DVar from $7,232$ to $3,618$.
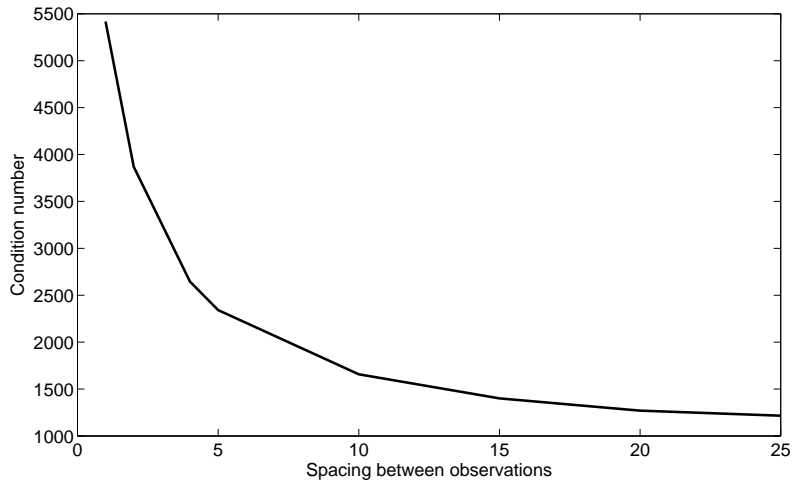
Figure 4: Condition number of the Met Office 3DVar assimilation scheme against spacing of 8 equatorial pseudo-observations.

We next examine the effect of changing the spacing of the observations. The observation error variance is fixed at $100Pa$ and the spacing between the observations is increased. As in the previous experiment we use a 4-by-4 grid of surface pressure pseudo-observations for 3DVar and a 2-by-3 grid of pseudo-observations at three different times for 4D-Var, with equal spacing between grid points. When the spacing between observations increases from one grid length to four grid lengths the condition number of the Hessian decreases from 1,525 to 690 for 3DVar and from 2,809 to 1,061 for 4DVar. A further spread of the observations, to a 16-grid-point spacing, results in even lower condition numbers of 163 for 3DVar and 344 for 4DVar. To illustrate further this effect we perform another 3DVar experiment using only 8 equally-spaced pseudo observations, positioned in a line along the equator, with an observation error variance of 1Pa. In Figure 4 we show the variation of the condition number as the spacing between the observations is increased from one to 25 grid lengths. We see clearly the reduction of the condition number with increased separation. Furthermore we note that the curve has a similar shape to those shown in Figure 3 for the simple example using the SOAR function. This reflects the use of the SOAR correlation function to model the horizontal background error correlations in the Met Office 3DVar scheme (Ingleby, 2001). Hence the results of experiments using pseudo-observations in the Met Office system confirm the theory presented and illustrated using the scalar system in section 3. We now consider how the conditioning of the system is affected by real observations. We begin by considering the effect of observation type.

13

4.2 Observation type

The Met Office, like most NWP centres, uses a variety of different observation types, including surface observations, upper-air measurements and satellite data (for example, ATOVS and SSMI). In Trémolet (2007) it was suggested that the conditioning of the ECMWF variational data assimilation is dominated by the accurate and densely distributed surface observations. In this section we look at the conditioning of the Met Office 3DVar and 4DVar schemes and investigate how the conditioning is affected by the different observation types. To do this we perform assimilations using all available observations and then separate assimilations in which single observations types are assimilated individually. We consider data from two different dates, 12Z on 27 October 2008 and 12Z on 14 July 2009.

In Figure 5 we show the condition number of the Hessian calculated from the 3DVar assimilations using both the full observations and single observation types. Some observation types, such as GPSRO and SSMIS, give a condition number of less than 50 when assimilated alone and so have not been shown in the plot. These are not considered to be major contributers to the conditioning of the inner loop minimisation at the Met Office. A similar plot for the conditioning of the Hessian of the 4DVar problem is shown in Figure 6, where the 4DVar system is run with a 6-hour assimilation window, from 9Z to 15Z.

We notice first that the values for the condition numbers of both the 3DVar and 4DVar systems are very similar for the two different dates. Experiments with the 4DVar system using data at different times of the day also produced similar condition numbers. It is clear from Figures 5 and 6 that in both 3DVar and 4DVar the observations types which give the largest condition number are the surface observations followed by ATOVS. The condition number of the variational cost function when only the surface observations are included is comparable in magnitude to the condition number when full observations are used within the assimilation. This is true for both 3DVar and 4DVar and suggests that the surface observations dominate the conditioning of the variational assimilation problem.

To further test this hypothesis we perform further experiments in which we use all observations except the surface observations in the assimilation and then all except the ATOVS observations. The condition numbers for these experiments are given in table 3 for the 3DVar assimilation and table 4 for the 4DVar assimilation. In these tables we show the condition numbers for both dates when we assimilate all observations, only surface or only ATOVS observations and all observations except surface observations. For 3DVar we also include an experiment in which we assimilate all observations except ATOVS. We see that the condition number when using all the observations is similar in magnitude to the condition number for the Hessian where only the surface observations are used. When only ATOVS observations are assimilated the condition number is about half of that where full observations are used. We find that the removal of the surface observation provides more than a 60 % reduction in the condition number of the assimilation problem. However, the conditioning of the 3DVar
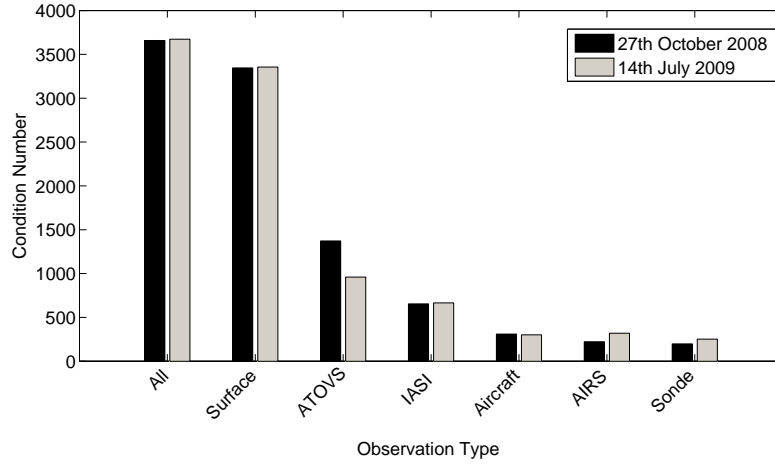
Figure 5: Conditioning of the Met Office 3DVar scheme using only selected observation types.
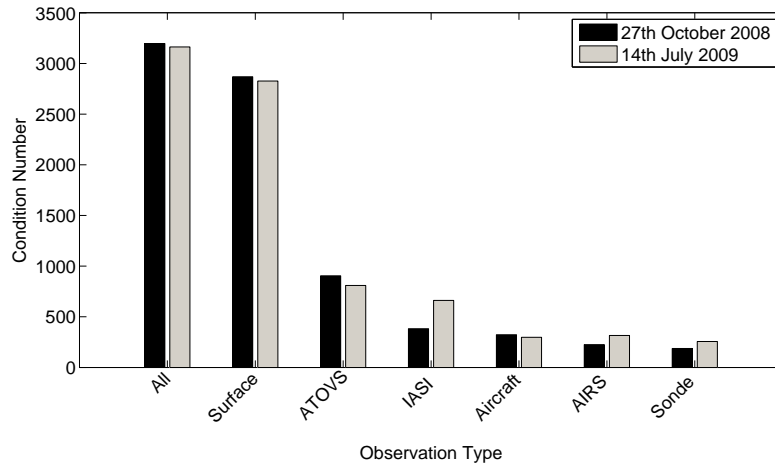

Figure 6: Conditioning of the Met Office 4DVar scheme using only selected observation types.

	Condition number	
Observation Type	27/10/2008	14/07/2009
All	3658	3673
Only Surface	3345	3355
Only ATOVS	1372	960
No Surface	1431	1215
No ATOVS	3667	3624

Table 3: Condition number of the Met Office 3DVar Scheme when assimilating different observations. Details of the experiments are given in the main text.

	Condition number	
Observation Type	27/10/2008	14/07/2009
All	3197	3163
Only Surface	2869	2827
Only ATOVS	905	809
No Surface	972	976

Table 4: As table 3, using the 4DVar scheme.

problem remains largely unaffected if ATOVS observations are removed. Hence we conclude that the surface observations dominate the conditioning of 3DVar and 4DVar.

From the theory and example of section 3 we expect the condition number to be dominated by areas of dense and accurate observations. An examination of the distribution of observations assimilated into the Met Office data assimilation system shows that there is a high density of surface observations over Europe. A calculation of the leading eigenvector of the Hessian produced when assimilating all observation types shows its largest components to be concentrated around Europe, giving a further indication that the observations in this area dominate the conditioning. A similar result was found by Trémolet (2007), who concluded that the dense and accurate network of surface pressure observations over Europe control the convergence of the ECMWF 4DVar system. Our experiments with the Met Office variational data assimilation scheme show that the same effect is seen in this system and that this is consistent with the new theory presented in section 3 and the pseudo-observation experiments of section 4.1.

4.3 Thinning of observations

A consequence of the high dependence of the condition number on the dense surface observations is that the conditioning should be better if a less dense set of observations is used. In practice dense observations are usually thinned, to ensure that no correlations between observation errors remain which are not accounted for in the assimilation scheme (Dando *et al.*, 2007). We now consider

Experiment		Condition number	
Scheme	Observations	No Thinning	Thinned Data
3DVar	Only Surface	3395	501
3DVar	All	3590	1406
4DVar	Only Surface	2778	456
4DVar	All	2975	928

Table 5: Variation of condition number using thinned data for 3DVar and 4DVar experiments, assimilating only surface data or all data.

the effect of observation thinning on the condition number of the Met Office system. We consider only the surface observations since as shown in section 4.2, the conditioning of the Met Office minimisation problem is dominated by these observations. For these experiments we use data from 12Z on 11 March 2009. The density of the surface observations is reduced by applying a 300km thinning. As a result of the thinning a large number of the land observations are removed. The Met Office variational system is then run first using the original observation data and then using the thinned surface observations. We calculate the condition number for experiments in which only the surface observations are assimilated and in which all observations are assimilated, using both the 3DVar and the 4DVar methods.

In table 5 we show the condition numbers calculated from the 3DVar and 4DVar experiments. We see that when only surface observations are assimilated then there is a large improvement in the conditioning when the data are thinned. The condition number falls by over 80% in both the 3DVar and 4DVar cases. When all other data are included in the assimilation then the thinning of surface observations has the effect of reducing the condition number to less than 40% of its original value, to 1406 for 3DVar and 928 for 4DVar. By comparison with tables 3 and 4 we see that these numbers are of the same order of magnitude as the experiments in which no surface data are assimilated. Hence we conclude that once the surface observations are thinned they have little effect on the conditioning of the system. As expected from the theory, by reducing the density of the observations used in the assimilation we are able to improve the conditioning of the problem. Of course the condition number cannot be the only criterion used to thin data, since any thinning of observations in this way also reduces the amount of information provided to the assimilation system and so may degrade the accuracy of the analysis. Nevertheless, when only a limited amount of computer time is available in which to perform the assimilation, it may be that more accuracy can be obtained by solving the more well conditioned problem, even with less data, than by solving the ill-conditioned problem with all the data.

5   Conclusions

The conditioning of the variational data assimilation problem plays an important role in determining how accurately the current state of the atmosphere can be determined in operational NWP. In most operational assimilation systems an initial preconditioning is performed by means of a variable transformation. In this work we have considered the conditioning of this preconditioned problem. Theoretical results have illustrated that the condition number of the problem is likely to increase when the observations are accurate and dense or when the background error correlation lengthscales are large. These results have been confirmed in a simple scalar example and illustrated using the variational data assimilation system of the Met Office.

With advances in observing technology and the move to higher resolution systems, it is clear that many more dense and accurate observations will be used in future variational data assimilation schemes of NWP centres. The results presented here imply that this will worsen the conditioning of the minimization problem. We have shown that thinning of the data can help to improve the conditioning, but a balance must be sought between the loss of accuracy due to solving an ill-conditioned problem and the loss of accuracy caused by removing observations. Further preconditioning techniques will also be needed in order to ensure that the data assimilation component of the forecasting system can be solved quickly and accurately. Some progress has already been made in this area (for example Tshimanga *et al.* 2008; Fisher *et al.* 2009), but further development of such techniques is required.

References

E. Andersson, M. Fisher, R. Munro and A. McNally, 2000: Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence, *Q. J. R. Met. Soc.*, **126**, 1455–1472.

R. Bannister, 2008: A Review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics, *Q. J. R. Met. Soc.*, **134**, 1971-1996.

P. Courtier, J.-N. Thépaut and A. Hollingsworth, 1994: A strategy for opera-

tional implementation of 4D-Var, using an incremental approach, *Q. J. R. Met. Soc.*, **120**, 1367–1387.

M. J. P. Cullen, 2003: Four-dimensional variational data assimilation: A new formulation of the background-error covariance matrix based on a potential vorticity representation, *Q. J. R. Met. Soc.*, **129**, 2777–2796.

M. L. Dando, A. J. Thorpe, J. R. Eyre, 2007: The optimal density of atmospheric sounder observations in the Met Office NWP system, *Q. J. R. Met. Soc.*, **133**, 1933–1943.

M. Fisher, J. Nocedal, Y. Trémolet and S.J. Wright, 2009: Data assimilation in weather forecasting: a case study in PDE-constrained optimization, *Optim. Eng.*, **10**, 409–426.

P. Gauthier, C. Charette, L. Fillion, P. Koclas and S. Laroche, 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part 1: The global analysis, *Atmosphere-Ocean*, **37**, 103–156.

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

P. Gauthier and J.-N. Thépaut, 2001: Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France, *Mon. Wea. Rev.*, **129**, 2089–2102.

G.H. Golub and C. F. Van Loan, 1996: *Matrix Computations, third edition,* Johns Hopkins University Press.

S. Gratton, A.S. Lawless and N.K. Nichols, 2007: Approximate Gauss-Newton methods for non-linear least-squares problems, *SIAM J. Optim.*, **18**, 106-132.

S. A. Haben, A.S. Lawless and N.K. Nichols, 2009: Conditioning of the 3DVAR data assimilation problem, *Mathematics Report 3/2009*, Department of Mathematics, University of Reading, U.K. Available from
http://www.reading.ac.uk/maths/research/maths-report-series.aspx

S. A. Haben, A.S. Lawless and N.K. Nichols, 2010: Conditioning and preconditioning of the variational data assimilation problem, *Preprint MPS_2010_16*, Department of Mathematics, University of Reading, U.K. *Submitted for publication.* Available from
http://www.reading.ac.uk/maths/research/maths-preprints.aspx

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, Y.-H. Kuo, 2009: Four-dimensional variational data assimilation for WRF: Formulation and preliminary results, *Mon. Wea. Rev.*, **137**, 299–314.

N.B. Ingleby, 2001: The statistical structure of forecast errors and its represen-

tation in The Met. Office Global 3-D Variational Data Assimilation Scheme, *Q. J. R. Met. Soc.*, **127**, 209-231.

D. Katz, A.S. Lawless, N.K. Nichols, M.J.P. Cullen and R.N. Bannister, 2010: Correlations of control variables in variational data assimilation, *Preprint MPS_2010_11*, Department of Mathematics, University of Reading, U.K. *Submitted for publication.* Available from http://www.reading.ac.uk/maths/research/maths-preprints.aspx

A.S. Lawless, S. Gratton and N.K. Nichols, 2005: An investigation of incremental 4D-Var using non-tangent linear models, *Q. J. R. Met. Soc.*, **131**, 459-476.

A.C. Lorenc, 1988: Optimal nonlinear objective analysis, *Q. J. R. Met. Soc.*, **114**, 205-240.

A.C. Lorenc, 1997: Development of an operational variational assimilation scheme, *J. Met. Soc. Japan*, **75**, 339–346.

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, D. Li, T. J. Payne, F. W. Saunders, 2000: The Met. Office global 3-dimensional variational data assimilation scheme, *Q. J. R. Met. Soc.*, **126**, 2991–3012.

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

F. Rawlins, S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc and T.J. Payne, 2007: The Met Office global four-dimensional variational data assimilation scheme, *Q. J. R. Met. Soc.*, **133**, 347–362.

Y. Trémolet, 2007: Incremental 4D-Var convergence study, *Tellus A*, **59**, 706–718.

J. Tshimanga, S. Gratton, A. T. Weaver, A. Sartenaer, 2008: Limited-memory preconditioners, with application to incremental four-dimensional variational data assimilation, *Q. J. R. Met. Soc.*, **134**, 751-769.