School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

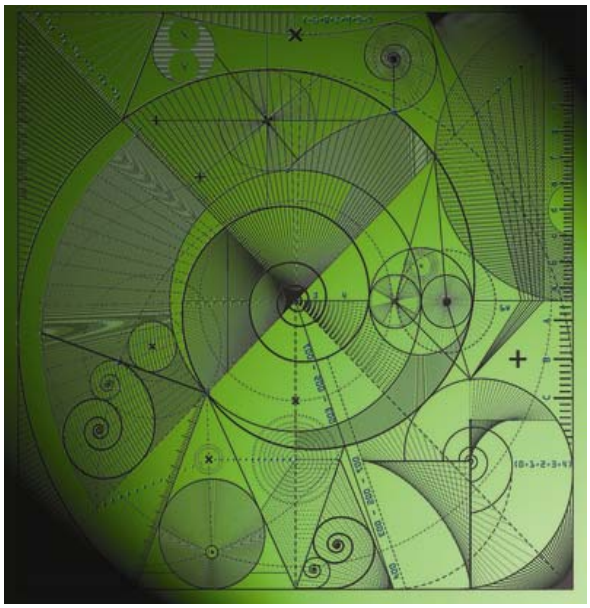
Preprint MPS-2012-17

15 August 2012

Data assimilation with correlated observation errors: analysis accuracy with approximate error covariance matrices

by

L.M. Stewart, S.L. Dance and N.K. Nichols



Data assimilation with correlated observation errors: analysis accuracy with approximate error covariance matrices

L.M. Stewart*, S.L. Dance†, N.K. Nichols
School of Mathematical and Physical Sciences,
University of Reading,
Reading RG6 6BB, UK

August 2012

Abstract

Remote sensing observations often have correlated errors, but the correlations are typically ignored in data assimilation for numerical weather prediction. The assumption of zero correlations is often used with data thinning methods, resulting in a loss of information. As operational centres move towards higher-resolution forecasting, there is a requirement to retain data providing detail on appropriate scales. Thus an alternative approach to dealing with observation error correlations is needed. In this paper, we consider several approaches to approximating observation error correlation matrices: diagonal approximations, eigendecomposition approximations, and Markov matrices. These approximations are applied in incremental 4D-Var experiments with a one-dimensional shallow water model using synthetic observations. Our experiments quantify analysis accuracy in comparison with a reference or 'truth' trajectory, as well as with analyses using the 'true' observation error covariance matrix. We show that it is often better to include an approximate correlation structure in the observation error covariance matrix than to incorrectly assume error independence. Furthermore, by choosing a suitable matrix approximation, it is feasible and computationally cheap to include error correlation structure in a variational data assimilation algorithm.

*Current address: Met Office, Meteorology Building, University of Reading, Reading, RG6 6BB.
†Corresponding author email: s.l.dance@reading.ac.uk

1 Introduction

Data assimilation provides techniques for combining observations of atmospheric variables with a priori knowledge of the atmosphere to obtain a consistent representation known as the analysis. The weighted importance of each contribution is determined by the size of its associated errors; hence it is crucial to the accuracy of the analysis that these errors be correctly specified.

Observation errors can generally be attributed to four main sources: (1) Instrument noise (2) Observation operator, or forward model error - For satellite observations this includes errors associated with the discretisation of the radiative transfer equation and errors in the mis-representation of gaseous contributors. (3) Representativity error - This is present when the observations can resolve spatial scales or features in the horizontal or vertical that the model cannot. For example, a sharp temperature inversion can be well-observed using radiosondes but cannot be represented precisely with the current vertical resolution of atmospheric models. (4)Pre-processing - For example, if we eliminate all satellite observations affected by clouds and some cloud-affected observations pass through the quality control, then one of the assimilation assumptions is violated and the cloudy observations will contaminate all satellite channels that are influenced by the cloud. It is reasonable to assume that instrument errors are independent and uncorrelated. However the other three sources of error result in observation error correlations.

Quantifying observation error correlations is not a straightforward problem because they can only be estimated in a statistical sense, not observed directly. Nevertheless, attempts have been made to quantify error correlation structure for a few different observation types such as Atmospheric Motion Vectors (Bormann et al., 2003) and satellite radiances (Sherlock et al., 2003; Stewart et al., 2009; Stewart, 2010; Stewart et al., 2012; Bormann and Bauer, 2010; Bormann et al., 2010). Using diagnosed correlations such as these in an operational assimilation system is far from straightforward: early attempts by the UK Met Office using IASI and AIRS data have resulted in conditioning problems with the 4D-Var minimization (Weston, 2011).

Due to the large number of observations, the computational demands of using observation error correlations are significant. If the observation vector is of size m then the observation error covariance matrix contains $(m^2 + m)/2$ independent elements. When observations have independent errors, i.e, the errors are uncorrelated, $(m^2 - m)/2$ of these elements are zero, and we only need represent m elements. Hence in current operations, observation errors are usually assumed uncorrelated. In most cases, to compensate for the omission of error correlation, the observation error variances are inflated so that the observations have a more appropriate lower weighting in the analysis (e.g., Collard, 2004). The assumptions of zero correlations and variance inflation are often used in conjunction with data thinning methods such as superobbing (Berger and Forsythe, 2004). Superobbing reduces the density of the data by averaging the properties of observations in a region and assigning this average as a single observation value. Under these assumptions, increasing the observation density beyond some threshold value has been shown to yield little or no improvement in analysis accuracy (Berger and Forsythe, 2004; Liu and Rabier, 2003; Dando et al., 2007). Stewart et al. (2008) and Stewart (2010) showed that the observation information content in the analysis is severely degraded under the incorrect assumption of independent observation errors. Such studies, combined with examples demonstrating that ignoring correlation structure hinders the use of satellite data (e.g., constraining channel selection algorithms (Collard, 2007)), suggest that error correlations for certain observation types have an important role to play in improving numerical weather forecasting. Indeed, the inclusion of observation error correlations has been shown to increase the accuracy of gradients of observed fields represented in the analysis (Seaman, 1977). Furthermore, retaining even an approximate error correlation structure shows clear benefits in terms of analysis information content (Stewart et al., 2008; Stewart, 2010).

Approximating observation error correlations in numerical weather prediction is a relatively new direction of research but progress has been made. Healy and White (2005) used circulant matrices to approximate symmetric Toeplitz observation error covariance matrices. Results showed that assuming uncorrelated observation errors gave misleading estimates of information content, but using an approximate circulant correlation structure was preferable to using no correlations. Fisher (2005) proposed giving the observation error covariance matrix a block-diagonal structure, with (uncorrelated) blocks corresponding to different instruments or channels. Individual block matrices were approximated by a truncated eigendecomposition. The method was shown to be successful in representing the true error correlation structure using a subset of the available eigenpairs. However, spurious long-range correlations were observed when too few eigenpairs were used.

In this paper we carry out numerical experiments using incremental 4D-Var (section 2) to address the following questions: *Is it better to model observation error correlation structure approximately than not at all?* and *Is it computationally feasible to model observation error correlations?* We use identical twin experiments so that a 'truth' trajectory is available and we are able to consider analysis errors explicitly. We specify two 'true' correlated observation error covariance matrices that we use to simulate synthetic observation errors, described in section 3. However, in the 4D-Var minimization, we use approximate observation error correlation structures to compute the cost function, with the aim that these approximations will provide more accurate analyses than incorrectly assuming uncorrelated errors, without a large increase in computational cost. The approximations chosen include diagonal matrices with inflated variances (Collard, 2004), a truncated eigendecomposition (Fisher, 2005), and a Markov matrix (sections 3.2-3.4). The Markov matrix has not been applied in this context before, and has the advantage of a tri-diagonal inverse.

The experiments are carried out using a one dimensional, irrotational, nonlinear shallow water model described in section 4 and the experimental design is given in section 5. The results (sections 7 and 8) show that including some correlation structure, even a basic approximation, often produces a smaller analysis error than using a diagonal approximation. We conclude (section 9) that it is computationally feasible and advantageous for analysis accuracy to include approximate observation error correlations in data assimilation. These encouraging results in a simple model should be investigated further for the potential improvement of operational assimilation systems.

2 4D-Var

Consider a discretised representation of the true state of the atmosphere $x_i^t \in \mathbb{R}^n$, at time t_i, where n is the total number of state variables. The analysis used in NWP will consist of the same model variables as this discretisation and must be consistent with the first guess or background field and the actual observations. The background field, $x^b \in \mathbb{R}^n$, is valid at initial time t_0 and usually given by a previous forecast. Observations are available at a sequence of times t_i, and are denoted $y_i \in \mathbb{R}^{m_i}$, where m_i is the total number of measurements available at time t_i. The background state and observations will be approximations to the true state of the atmosphere,

$$x^b = x_0^t + \epsilon^b, \tag{1}$$

$$y_i = h_i(x_i^t) + \epsilon_i^o, \tag{2}$$

where $\epsilon^b \in \mathbb{R}^n$ are the background errors valid at initial time, $\epsilon_i^o \in \mathbb{R}^{m_i}$ are the observation errors at time t_i, and h_i is the, possibly nonlinear, observation operator mapping from state space to measurement space at time t_i; for example, a radiative transfer model which simulates radiances from an input atmospheric profile. The errors are assumed unbiased and mutually independent, and also to have covariances $B = \mathbb{E}[\epsilon^b(\epsilon^b)^T]$ and $R_i = \mathbb{E}[\epsilon_i^o(\epsilon_i^o)^T]$.

The objective of 4D-Var is to minimise the cost function,

$$J(x_0) = \frac{1}{2}(x_0 - x^b)^T B^{-1}(x_0 - x^b) + \frac{1}{2}\sum_{i=0}^{n}(h_i(x_i) - y_i)^T R_i^{-1}(h_i(x_i) - y_i), \tag{3}$$

subject to the strong constraint that the sequence of model states must also be a solution to the model equations,

$$x_{i+1} = m(t_i, t_{i+1}, x_i), \tag{4}$$

where x_i is the model state at time t_i and $m(t_0, t_i, x_0)$ is the nonlinear model evolving x_0 from time t_0 to time t_i. The strong constraint given by equation (4) implies the model is assumed to be perfect.

The cost function (3) measures the weighted sum of the distances between the model state x_0 and the background at the start of the time interval t_0 and the sum of the observation innovations $(h_i(x_i) - y_i)$ computed with respect to the time of the observation. 4D-Var therefore provides an initial condition such that the forecast best fits the observations within the whole assimilation interval.

Incremental 4D-Var (Courtier et al., 1994), reduces the cost of the 4D-Var algorithm by approximating the full nonlinear cost function (3) by a series of convex quadratic cost functions. The minimisation of

these cost functions is constrained by a linear approximation M to the nonlinear model m (4). Each cost function minimisation is performed iteratively and the resultant solution is used to update the nonlinear model trajectory. Full details of the procedure are described by Lawless et al. (2005) and Stewart (2010). We summarize the algorithm here, denoting k as iteration number and $x_i^{(k)}$ as the full nonlinear solution valid at time t_i and the k-th iteration.

1. At the first timestep ($k = 0$) define the current guess $x_0^{(0)} = x^b$.

Loop over k:

2. Run the nonlinear model to calculate $x_i^{(k)}$ at each time step i.
3. Calculate the innovation vector for each observation
$$d_i^{(k)} = y_i - h(x_i^{(k)}).$$
4. Start the inner loop minimisation. Find the value of $\delta x_0^{(k)}$ that minimises the incremental cost function
$$\begin{aligned} J^{(k)}(\delta x_0^{(k)}) &= \frac{1}{2}(\delta x_0^{(k)} - (x^b - x_0^{(k)}))^T B^{-1}(\delta x_0^{(k)} - (x^b - x_0^{(k)})) \\ &\quad + \frac{1}{2}\sum_{i=0}^{n}(H_i \delta x_i^{(k)} - d_i^{(k)})^T R_i^{-1}(H_i \delta x_i^{(k)} - d_i^{(k)}) \end{aligned} \tag{5}$$
subject to
$$\delta x_{i+1}^{(k)} = M(t_i, t_{i+1}, x^{(k)})\delta x_i^{(k)},$$
where H_i is the linearisation of the observation operator h_i and $M(t_i, t_{i+1}, x^{(k)})$ is the linearisation of the model $m(t_i, t_{i+1}, x^{(k)})$. Both of these linearisations are around the current state estimate $x_i^{(k)}$.
5. Update the guess field using
$$x_0^{(k+1)} = x_0^{(k)} + \delta x_0^{(k)}.$$
6. Repeat outer loop (steps 2 - 5) until the desired convergence is reached.

In our implementation, the conjugate gradient method (Golub and van Loan, 1996) is used to carry out the inner loop minimisation to solve (5). The maximum number of outer and inner loops performed are 20 and 200, respectively. The outer and inner loops are terminated once the following criteria are satisified. For the outer loop, we require that (Lawless and Nichols, 2006)
$$\frac{|J^{(k+1)} - J^{(k)}|}{1 + |J^{(k)}|} < 0.01, \tag{6}$$
where the superscripts indicate the outer loop iteration index . For the inner loop, we require that
$$\frac{\left\|\nabla J_m^{(k)}\right\|_2}{\left\|\nabla J_0^{(k)}\right\|_2} < 0.1, \tag{7}$$
where the subscripts $0, m$ indicate the inner loop iteration index and k indicates the outer loop iteration index. With these stopping criteria and tolerance values, we expect the converged solution of the minimization to be accurate to approximately 2 decimal places (Stewart, 2010; Gill et al., 1981).

3 Observation error covariance matrices

This paper describes incremental 4D-Var experiments using approximate forms of observation error covariance matrices to take account of correlated observation errors in the minimization. In section 3.1 we describe the 'true' error covariance structures we use to simulate synthetic observation errors. In sections 3.2-3.4 we explain the choice of the various approximate forms of observation error covariance matrix we employ in the incremental 4D-Var cost-function. The goal is that these choices will improve the accuracy of the analysis and have only a modest computational burden.

3.1 True error covariance structures

For our experiments we use two different forms of true observation error covariance structure, R_t. We make use of the general decomposition of an observation error covariance matrix R into a diagonal variance matrix $D \in \mathbb{R}^{m\times m}$ and a correlation matrix $C \in \mathbb{R}^{m\times m}$ such that

$$R = D^{1/2} C D^{1/2}, \tag{8}$$

so that we may vary the observation error variance and correlation structure separately. Note that if $C = I$ then R is diagonal and D is the diagonal matrix of error variances.

In Experiment 1 we use a true error correlation matrix with a Markov distribution, C_M, given by

$$C_M(i,j) = \exp\left\{\frac{-|i-j|\Delta x_D}{L_R}\right\} \tag{9}$$

where $\Delta x_D = 0.01$m is the spatial separation and $L_R = 0.1$m is the length scale. The Markov matrix is the resultant covariance matrix from a first-order autoregression process, (Wilks, 1995), and is discussed in more detail in section 3.3.

In the second experiment the true error correlation structure follows a SOAR (second-order autoregressive) distribution. The SOAR error covariance matrix is given by

$$C_S(i,j) = \left(1 + \frac{|i-j|\Delta x_D}{L_R}\right)\exp\left\{\frac{-|i-j|\Delta x_D}{L_R}\right\}. \tag{10}$$

The SOAR matrix is an idealised correlation structure that is often used to model background error correlation structures in the horizontal (e.g., Ingleby, 2001). It is commonly employed in preference to a Gaussian structure because its distribution has longer tails, better at matching empirical estimates, and is better conditioned for inversion (Haben, 2011; Haben et al., 2011).

3.2 Diagonal approximation

The diagonal approximation that we use in our experiments is commonly employed operationally. Diagonal matrices are simple and numerical efficient: In incremental 4D-Var the inverse observation error covariance matrix is required for $2N$ matrix-vector products in evaluating the cost function and gradient, where N is the number of assimilation timesteps. When the observation error covariance matrix is diagonal, its inverse will also be diagonal, resulting in very inexpensive matrix-vector products.

The simplest diagonal approximation of an error covariance matrix is taking the diagonal equal to the true variances. However, by ignoring entirely the correlated component of the observation error, the observations will be overweighted in the analysis because they will appear more informative than they truly are. Therefore in order to compensate for the lack of correlation, a diagonal approximation given by the diagonal of the true matrix scaled by an inflation factor is used (Hilton et al., 2009). This reduces the weighting of the observations in the analysis. The diagonal approximation is now in the form

$$\hat{D} = \begin{pmatrix} d_1\sigma_1^2 & 0 & \dots & 0 \\ 0 & d_2\sigma_2^2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & & 0 \\ 0 & \dots & 0 & d_m\sigma_m^2 \end{pmatrix}, \tag{11}$$

where d_i is the real, positive inflation factor for variance σ_i^2.

In our experiments, the diagonal matrix representations are a diagonal matrix of the true error variances, and scalar multiples of this matrix. The scalar multiples are chosen to be between two and four, in line with our earlier 3D-Var information content results (Stewart et al., 2008; Stewart, 2010) and results given in Collard (2004). These showed that a 2-4 times variance inflation was preferable to a simple diagonal approximation when observation and background error correlations were both present; but when there were correlated observation errors and uncorrelated background errors, a simple diagonal approximation performed better than variance inflation.

3.3 Markov matrix

The second approximate form of matrix that we employ is a Markov matrix. This is a novel choice that has not previously been reported in the literature for observation error covariance approximation. The $(i,j)th$ element of a general Markov matrix, R, is given by

$$R(i,j) = \sigma^2 \rho^{|i-j|}, \tag{12}$$

where σ^2 is the observation error variance, and $0 \leq \rho \leq 1$ is a parameter describing the strength of the correlations. This matrix has a tri-diagonal inverse (Rodgers, 2000),

$$R^{-1} = \frac{1}{\sigma^2 (1-\rho^2)} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \tag{13}$$

The storage needed for reconstructing matrix (13) is limited to the value of ρ, and the number of operations involved in a matrix-vector product using a tri-diagonal matrix is the same order as that using a diagonal matrix. Therefore calculating the cost function using the Markov matrix approximation is a possibility for operational assimilation.

In our experiments, we let $\rho = \exp(-\Delta x_D / L_R)$, as in (9). The Markov matrix representations are Markov structured matrices with length scales $L_R = 0.2$m, $L_R = 0.1$m, $L_R = 0.05$m, and $L_R = 0.01$m, i.e, double, the same as, half, and a tenth of the true length scale. These values are chosen to represent different levels of error dependence. In Figure 1, the central row of each Markov matrix is plotted. Note that as the length scale decreases, the thickness of the central correlation band decreases. We also test the Markov matrix representation for the case where L_R is small enough that $C_M(i,j) = 0$ for $i \neq j$; this should produce the same result as using the diagonal approximation with the true error variances, and is a continuity test on our system.

3.4 Eigendecomposition (ED) matrix

Starting from the general covariance decomposition (8), Fisher (2005) proposed that the observation error covariance matrix be approximated using a truncated eigendecomposition $\hat{C}$ of the error correlation matrix C,

$$R = D^{1/2}(\alpha I + \sum_{k=1}^{K} (\lambda_k - \alpha) v_k v_k^T) D^{1/2} \equiv D^{1/2} \hat{C} D^{1/2}, \tag{14}$$

where (λ_k, v_k) is an (eigenvalue, eigenvector) pair of C, K is the number of leading eigenpairs used in the approximation, and α is chosen such that trace(R)=trace(D), i.e, so that there is no mis-approximation of the total error variance. The inverse of (14) is easily obtainable and is given by

$$R^{-1} = D^{-1/2}(\alpha^{-1} I + \sum_{k=1}^{K} (\lambda_k^{-1} - \alpha^{-1}) v_k v_k^T) D^{-1/2} = D^{-1/2} \hat{C}^{-1} D^{-1/2}. \tag{15}$$

The representation (14) allows the retention of some of the true correlation structure, with the user choosing how accurately to represent the inverse error covariance matrix (15) through the choice of K. Care must be taken to ensure numerical stability (for example choosing K such that $\lambda_k^{-1} - \alpha^{-1}$ is never too small).

In Fisher (2005) the leading eigenpairs of C are found using the Lanczos algorithm. In our experiments, the leading eigenpairs needed for the representation are pre-computed using the MATLAB function *eigs()* (MathWorks, 2009) which uses an implicitly restarted Arnoldi method (Lehoucq and Sorensen, 1996; Sorensen, 1992).

By studying the eigenspectra of the true error correlation matrices we can estimate how many eigenpairs are needed for a good representation. The eigenspectra of a Markov matrix and a SOAR matrix, both of size 1001×1001, with length scale $L_R = 0.1$m and spatial separation $\Delta x_D = 0.01$m, are plotted

in Figure 2. The plots show that the eigenvalue size declines sharply as the eigenvalue number increases. The condition number (ratio of largest to smallest eigenvalue) for the Markov matrix is 400 and for the SOAR matrix is 4.8×10^5. After 100 eigenvalues 80% and 99% of the overall uncertainty is represented for the Markov and SOAR matrix respectively. (Uncertainty percentages are calculated using (sum of eigenvalues used)/(sum of all eigenvalues or trace of matrix) $\times$ 100%.) Therefore we use 100 eigenpairs as an empirical upper limit to the number of eigenpairs used in the assimilation.

The number of eigenpairs we use in our approximations are $k = 10$, $k = 20$, $k = 50$, and $k = 100$. This represents 1%, 2%, 5% and 10% of the total number of eigenpairs. An ED approximation using the full number of eigenpairs $k = 1001$ is equivalent to using the true error correlation matrix in the system. Obviously using all the eigenpairs is an expensive procedure and would not be attempted operationally. However in these smaller dimensioned experiments, knowing the performance of the assimilation under the true error correlation matrix allows us to quantify the success of an assimilation using an approximated correlation matrix relative to the truth. We therefore also run the assimilation using the ED approximation with the full number of eigenpairs.

4 Shallow water model (SWM)

In this section we describe the forecast model used in our experiments. This a one-dimensional, nonlinear, shallow-water model system (SWM) describing the irrotational flow of a single-layer, inviscid fluid over an object. The SWM has been used for a variety of assimilation experiments e.g., Lawless et al. (2005, 2008); Katz et al. (2011); Steward et al. (2012). As an idealized system, it allows clearer understanding of the results without the obfuscating complexity of a more realistic system.

The continuous equations describing 1D shallow water flow are given by

$$\frac{Du}{Dt} + \frac{\partial \phi}{\partial x_D} = -g\frac{\partial h_o}{\partial x_D}, \tag{16}$$

$$\frac{D(\ln \phi)}{Dt} + \frac{\partial u}{\partial x_D} = 0, \tag{17}$$

on the domain $x_D \in [0, L]$, $t \in [0, T]$, where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x_D},$$

and $h_o = h_o(x_D)$ is the height of the bottom orography, u is the fluid velocity, $\phi = gh$ is the geopotential where g is the gravitational acceleration and $h > 0$ is the depth of fluid above the orography. The spatial boundary conditions are taken to be periodic, so that at any time $t \in [0, T]$,

$$\begin{aligned} u(0,t) &= u(L,t), \\ \phi(0,t) &= \phi(L,t), \\ h_o(0) &= h_o(L). \end{aligned}$$

The nonlinear shallow water equations are discretised using a two-time-level semi-implicit, semi-Lagrangian scheme, described by Lawless (2001); Lawless et al. (2003).

The experiments in this paper model a flow field described in Houghton and Kasahara (1968) in which shallow water motion is forced by some orography. Using the shallow water equations (16)-(17), we consider a fluid at rest when $t < 0$, with the geopotential equal to $\phi_0 - h_o(x_D)$, where ϕ_0 is a constant. At $t = 0$ the fluid is set in motion with a constant velocity u_0 at all grid points, causing a wave motion to develop outwards from the obstacle in the fluid. The solution close to the object becomes a steady state solution Lawless et al. (2003). We restrict the fluid motions to be not too highly nonlinear so as to keep our assumptions of linearity as valid as possible. We use a periodic domain where the boundaries are at a sufficient distance from the obstacle to ensure any propagating wave motions in the vicinity of the obstacle respect the asymptotic conditions.

The data used in the experiment is based on Case A in Houghton and Kasahara (1968). We consider a 1D spatial domain between $[0, 10\text{m}]$ equally divided into 1001 grid points with spatial step $\Delta x_D = 0.01\text{m}$.

The height of the obstacle in the fluid is given by

$$h_o(x_D) = \begin{cases} h_C\left(1 - \frac{x_D^2}{a^2}\right) & 0 \leq |x_D| \leq a \\ 0 & |x_D| < 0 \text{ or } |x_D| > a \end{cases}$$

where h_C is the maximum height of the obstacle and a is half the length over which the base of the obstacle extends. The values of a and h_C are set as: $a = 40\Delta x_D = 0.4$m, $h_C = 0.05$m. The temporal domain is 100 timesteps with step size $t = 9.2 \times 10^{-3}$s. At $t = 0$ the initial velocity is $u_0 = 0.1\text{ms}^{-1}$, and the geopotential is $\phi(x_D) = g(0.2 - h_o(x_D))$ where $g = 10\text{ms}^{-2}$.

5 Twin Experiments

Our numerical experiments are performed using an assimilation-forecast system based on an incremental 4D-Var system implemented for the SWM of section 4. In order to assess the impact of modelling correlated observation error structure we use different approximations to the observation error covariance matrix in the incremental 4D-Var cost function, as discussed in section 3: diagonal approximations, Markov approximations, and ED approximations. By keeping all other variables the same, any changes in the analysis trajectory can be attributed to the specification of the observation errors. In practice such an approach is not always possible since the true error covariance matrix is rarely known explicitly. The one-dimensional construction of the SWM means we are considering error correlations between observations in the horizontal. However the techniques we are using could easily be translated to a one-dimensional vertical profile, such as the radiance profiles used in 1D-Var. Therefore our assimilation tests will remain independent of any discussion on issues of spatial resolution or horizontal thinning.

An identical twin experiment is performed by running the nonlinear SWM forward in time from the initial conditions described in section 4, to generate a true model solution at each assimilation timestep, $t_0, \ldots t_{100}$. This is known as the truth trajectory. Observations of fluid velocity u and geopotential ϕ are sampled from the truth trajectory. It is assumed that there is an observation at each grid point, and after every 10 timesteps, i.e, 10 sets of 1001 observations in total; this density was chosen to represent a very well-observed system. Random noise with multivariate normal distribution $N(\mathbf{0}, R_t)$ is added as observation error. The errors in the u and ϕ observations are assumed mutually uncorrelated so that the observation error vectors for the u and ϕ fields may be generated independently. The correlation structures for the errors are given in section 3.1. For most of the experiments the standard deviation of the noise is set at 20% of the mean field value. This is 0.02ms^{-1} for u observations and $0.2\text{m}^2\text{s}^{-2}$ for ϕ observations.

The initial background is taken as the initial truth trajectory plus random noise from a normal distribution with mean zero and covariance matrix B_t. All experiments are run with uncorrelated background errors where the background error variances are half those of the observation errors, i.e., $B_t = \frac{1}{2}\text{diag}\,(R_t)$. We have chosen a diagonal background error covariance here for simplicity. In these preliminary experiments observations are available at every grid point, so the role of B_t in observation spreading is not critical to the analysis, and the effects of correlations in R will dominate our results.

An incremental 4D-Var data assimilation algorithm is then run using these observation and background data, with the covariances used in the calculation of the cost function taken to be B_t for the background error covariance and R_f (an approximate covariance structure) for the observation error covariance. The time window for the assimilation is 100 model timesteps. The assimilation finds an analysis valid at initial time $t = 0$, for each model grid point. We then integrate the analysis forward in time to give an updated forecast. The accuracy of the resulting analysis and forecast can be evaluated using the error measures described in the next section.

6 Analysis error measures

We now describe the diagnostics used to evaluate the success of each approximation. The assimilation is run using different approximations R_f to the true error covariance matrix R_t. We illustrate the comparative behaviour of the assimilation under different approximations by comparing:

(a) Error 1 (E1): The norm of the analysis error in the true solution

$$\left\|\overline{x}_{R_f} - x^*\right\|_2 \tag{18}$$

where x^* is the true solution of the original model run from which the observations are sampled, and $\overline{x}_{R_f}$ is the converged solution to the assimilation problem when the approximation R_f is used in calculating the cost function, but the observation errors themselves are sampled using the true error covariance R_t;

(b) Error 2 (E2): The percentage norm of the analysis error in the converged solution relative to the norm of the true converged solution

$$\frac{\left\|\overline{x}_{R_f} - \overline{x}_{R_t}\right\|_2}{\left\|\overline{x}_{R_t}\right\|_2} \times 100 \tag{19}$$

where $\overline{x}_{R_t}$ is the true converged solution to the assimilation problem when the true error covariance matrix R_t is used both in calculating the cost function and in sampling the observation errors.

Errors E1 and E2 provide us with information on the closeness of different analyses. Since the magnitude of the ϕ field is an order larger than that of the u field, we produce separate error norms (18) and (19) for u and ϕ to avoid changes in the u field being overshadowed by changes in the ϕ field.

7 Experiment 1: Markov error correlation structure

In our first experiment we investigate the impact on analysis accuracy of using a diagonal matrix, a Markov matrix, and an eigendecomposition (ED) matrix to represent a Markov error correlation structure. The analysis errors E1 and E2 at the start of the assimilation window ($t = 0$) for different approximations to a Markov error correlation structure are given in Tables 1 and 2.

The error in the background field is $\left\|x^b - x^*\right\|_2 = 0.32$ for the u field and $\left\|x^b - x^*\right\|_2 = 6.32$ for the ϕ field. We can see in Tables 1 and 2 that in all cases the approximation results in an improvement to the background field. Using the true error covariance matrix, i.e, a Markov matrix with length scale $L_R = 0.1$m, produces the smallest analysis errors; the percentage error E2 is zero for this matrix because $R_t = R_f$. Using a diagonal matrix approximation results in the largest analysis errors.

Using a Markov approximation with double ($L_R = 0.2$m) or half ($L_R = 0.05$m) the true length scale results in a small E2 error of less than 2% for the u and ϕ fields. This implies that choosing the exact length scale is not essential to producing accurate results. Also, using a Markov matrix approximation with length scale between $L_R = 0.2$m and $L_R = 0.05$m results in a smaller E2 error than that of an ED approximation using 100 eigenpairs. Using more eigenpairs in the ED approximation produces a more accurate analysis, but at greater computational expense because additional eigenpairs must be stored and used in cost function computations. We can therefore infer that although using more eigenpairs is beneficial, a Markov approximation using an approximate length scale is cheaper and more effective. In the next section we will see if the same conclusions are drawn when the true error covariance matrix follows a non-Markov distribution.

It is worth noting that using an ED approximation with a small number of eigenpairs can generate a smaller analysis error than when a diagonal approximation is used, and is comparable with a weakly correlated Markov approximation. For example, a diagonal matrix approximation results in an E2 error of 7.2% in the u field compared to a 5.9% error under an ED matrix with 10 eigenpairs and a 5.6% error under a Markov matrix with length scale $L_R = 0.01$m. Combined with the results for a Markov matrix approximation, this implies that it is often better to include some correlation structure, even if it is a weak approximation, than none at all.

Similar tests were performed for different observation frequencies. We found that using more frequent observations resulted in a small improvement in E2 for the three matrix approximations tested: a diagonal matrix, a Markov matrix with $L_R = 0.05$m, and an ED matrix with $k = 50$. Increasing the frequency of observations had the biggest impact on the diagonal approximation. Nevertheless, even when there were observations at every timestep (100 observation sets) the error E2 under a diagonal approximation was still significantly larger (5.7%) than when a Markov (1.3%) and an ED approximation (3.4%) were used.

In the next section we will extend the experiments performed here to a different choice of true error correlation structure.

8 Experiment 2: SOAR error correlation structure

In this section we consider the effect of our choice of the true observation error correlation structure. In Experiment 1 the true error correlation matrix was generated from a Markov distribution. We now change the correlation matrix to represent a SOAR distribution with length scale $L_R = 0.1$m. The matrix representations used to approximate this correlation structure are the same as those used in Experiment 1. Using a SOAR matrix will allow us to determine whether the Markov approximation also minimises analysis error when the true correlation structure is not in Markov form, and how well the ED and diagonal approximations perform in comparison.

The analysis errors E1 and E2 at $t = 0$ for the different approximations to the SOAR error covariance matrix are given in Tables 3 and 4. Comparing the results to Table 1 and 2, we observe that the qualitative nature of the errors is very similar. For example, using the true error covariance matrix structure results in the smallest errors and diagonal approximations result in the largest errors. The approximations resulting in the smallest analysis errors are a Markov matrix with length scale $L_R = 0.2$m and an ED matrix using 100 eigenpairs. It is intuitive that a Markov matrix with a longer length scale is preferable, because of the longer tails in a SOAR function. The E2 error in the u field is also small for Markov approximations with length scale between $L_R = 0.2$m and $L_R = 0.05$m, compared to a 9.4% error when a $4\times$ diagonal approximation is used. Inflated diagonal approximations perform slightly worse than a simple diagonal approximation; this is in line with the information content results in Stewart (2010), when the background errors were uncorrelated.

It is also expected that an ED matrix using 100 eigenpairs results in a very small analysis error relative to the converged solution, because as we observed in section 3.4, 100 eigenpairs represent 99% of the overall uncertainty in the matrix. It is encouraging that an ED approximation using even fewer eigenpairs also results in an improved E2 error relative to a diagonal approximation; using 5% of the available eigenpairs results in an E2 error in the ϕ field of 2.3% compared to 5.3% when a diagonal approximation is used. The E1 errors in using an ED approximation to model a SOAR error covariance structure are smaller than those generated when an ED approximation was used to model a Markov error covariance structure in Experiment 1. This is because, for a SOAR error covariance matrix, more uncertainty is represented using the same number of eigenpairs; as demonstrated in the steeper gradient in Figure 2.

It is also interesting to look at individual analysis errors over the domain. At each grid point the analysis error is given by the difference between the true analysis and the analysis resulting from the assimilation. Figures 3 and 4 show the analysis errors in the u and ϕ fields at $t = 0$ and $t = 50$, respectively. By looking at the spread of analysis errors for the diagonal and Markov approximations we see that the difference between the two is not uniform over the domain, i.e, in some regions, a diagonal approximation is much worse than a Markov approximation compared to the average. Such differences can be important operationally. For example, if a temperature error was reduced by 0.2K on average, and is reduced by 2K on one occasion. This 2K change can result in a modification of the wind forecast from 20 knots to 40 knots.

Comparing Figure 3 to 4 we observe that as the forecast evolves the analysis errors become smoother. For the u field the overall magnitude of the errors remains the same, but there is a reduction in error for the ϕ field. At the start of the time window, the diagonal approximation performs worse that the Markov approximation for both u and ϕ fields. At the centre of the time window, the errors in the u field for a Markov and a diagonal approximation are very similar, but for the ϕ field, the Markov approximation is still noticeably better. We can explain this by considering the assumptions on the shallow water model. The model in this assimilation is assumed perfect, and by construction is well-behaved, meaning that small errors in the analysis at $t = 0$ will be smoothed out over time. However, for a more complex operational system, a slight error in the true analysis field at $t = 0$ may propagate and grow with time, resulting in a modified forecast. It would therefore be interesting to extend these results to an imperfect and more poorly behaved model system.

Finally in this section we study how the error in the assimilation depends on the level of noise on the observations. Previous experiments were run with the standard deviation of the noise at 20% of the mean field value, here we vary this value between 1% and 30%. The error in the assimilation is described by E2, as defined in Section 7.1.2 (19). A plot of this error measure versus the percentage observation error in the u and ϕ field is shown in Figures 5 and 6, respectively. We see that for all three approximations studied, the E2 error increases with the percentage observation error. In the u field,

E2 increases close to linearly with noise level for the Markov and ED approximation; similarly for the ϕ field below 20% noise level. However, the diagonal approximation increases more rapidly with noise level in both fields, although the gradient becomes more linear as the observation errors increase. We can conclude that using a correlated matrix approximation is preferable to a diagonal one regardless of the level of observation error noise.

9 Summary and Discussion

The correct treatment of observation errors is a double problem for operational weather centres. Firstly the statistical properties of the errors are relatively unknown. Observations taken by different instruments are likely to have independent errors, but pre-processing techniques, mis-representation in the forward model, and contrasting observation and model resolutions can create error correlations. Secondly, even when good estimates of the errors can be made, the number of observations is of order 10^6 for a global assimilation run, and so the storage and subsequent computation using observation error correlations is ostensibly infeasible.

In this paper we developed an incremental 4D-Var data assimilation algorithm that used correlated approximations to model a simulated error correlation structure. This was applied to a one-dimensional shallow water model, and the impact of each approximation on analysis accuracy was determined. These results were encouraging but of course suffer from some limitations. The assumption that every model variable is observed directly prohibits a direct comparison with satellite data assimilation, in which the desired atmospheric fields are nonlinear combinations of the observed quantities. The assumption of uncorrelated background errors was useful here in allowing us to clearly see the effect of correlated observation errors, but it is unrealistic and could be relaxed in future experiments.

We concluded from the experiments that by choosing a suitable matrix approximation it is feasible to cheaply include some level of error correlation structure in a variational data assimilation algorithm. For different simulated observation error distributions and levels of error noise, we showed it is better to include some level of correlation structure in the observation error covariance matrix approximation than to assume incorrectly error independence. For example, an eigendecomposition approximation with 5% of the available eigenpairs results in a smaller analysis error than a diagonal approximation. However, the best results were achieved using a Markov matrix approximation, and this was found to be robust to changes in true correlation form and lengthscale. The next step will be to develop and test these promising ideas for real observations in an operational system.

10 Acknowledgments

Laura Stewart was supported by a UK NERC PhD studentship with CASE sponsorship from the UK Met Office.

References

H. Berger and M. Forsythe. Satellite wind superobbing. *Met Office Forecasting Research Technical Report*, 451, 2004.

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. *Q. J. R. Meteorol. Soc.*, 136(649):1036–1050, 2010. doi: 10.1002/qj.616.

N. Bormann, S. Saarinen, G. Kelly, and J.-N. Thépaut. The spatial structure of observation errors in Atmospheric Motion Vectors for geostationary satellite data. *Monthly Weather Review*, 131:706–718, 2003.

N. Bormann, A. Collard, and P. Bauer. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *Q. J. R. Meteorol. Soc*, 136(649):1051–1063, 2010. doi: 10.1002/qj.615. URL `http://dx.doi.org/10.1002/qj.615`.

A.D. Collard. On the choice of observation errors for the assimilation of AIRS brightness temperatures: A theoretical study. *ECMWF Technical Memoranda*, AC/90, 2004.

A.D. Collard. Selection of IASI channels for use in Numerical Weather Prediction. *Q.J.R.Meteorol.Soc.*, 133:1977–1991, 2007.

P. Courtier, J.-N. Thépaut, and A.Holingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q.J.R.Meteorol.Soc.*, 120:1367–1387, 1994.

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

M. Fisher. Accounting for correlated observation error in the ECMWF analysis. *ECMWF Technical Memoranda*, MF/05106, 2005.

P. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, San Diego, 1981.

G. Golub and C.F. van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, third edition, 1996.

S. Haben, A.S. Lawless, and N.K. Nichols. Conditioning of incremental variational data assimilation, with application to the met office system. *Tellus A*, 63:782–792, 2011.

S.A. Haben. *Conditioning and Preconditioning of the Minimisation Problem in Variational Data Assimilation*. PhD thesis, University of Reading, 2011. Available from http://www.reading.ac.uk/maths-and-stats/research/maths-phdtheses.aspx.

S.B. Healy and A.A. White. Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Q.J.R.Meteorol.Soc.*, 131:63–72, 2005.

F. Hilton, A. Collard, V. Guidard, R. Randriamampianina, and M. Schwaerz. Assimilation of IASI radiances at European NWP centres. In *Proceedings of Workshop on the assimilation of IASI data in NWP, ECMWF, Reading, UK, 6-8 May 2009*, 2009.

D.D. Houghton and K. Kasahara. Nonlinear shallow fluid flow over an isolated ridge. *Commun. Pure Appl. Math.*, 21:1–23, 1968.

N. B. Ingleby. The statistical structure of forecast errors and its representation in the Met Office global 3-D variational data assimilation scheme. *Q.J. R. Meteorol. Soc.*, 127:209–231, 2001.

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. Royal Met. Soc.*, 137:620–630, 2011.

A.S. Lawless. *Development of linear models for data assimilation in numerical weather prediction*. PhD thesis, University of Reading, 2001.

A.S. Lawless and N.K. Nichols. Inner-loop stopping criteria for incremental four-dimensional variational data assimilation. *Monthly Weather Review*, 134:3425–3435, 2006.

A.S. Lawless, N.K. Nichols, and S.P.Ballard. A comparison of two methods for developing the linearization of a shallow-water model. *Q.J.R.Meteorol.Soc.*, 129:1237–1254, 2003.

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

A.S. Lawless, N.K. Nichols, C. Boess, and A. Bunse-Gerstner. Using model reduction methods within incremental 4d-var. *Monthly Weather Review*, 136:1511–1522, 2008.

R.B. Lehoucq and D.C. Sorensen. Deflation techniques for an implicitly re-started arnoldi iteration. *SIAM J. Matrix Analysis and Applications*, 17:789–821, 1996.

Z.-Q. Liu and F. Rabier. The potential of high-density observations for numerical weather prediction: A study with simulated observations. *Q.J.R.Meteorol.Soc.*, 129:3013–3035, 2003.

The MathWorks. MATLAB documentation. In *http://www.mathworks.com/access/helpdesk/help/techdoc*, 2009.

C.D. Rodgers. *Inverse Methods for Atmopsheric Sounding: Theory and Practice*. World Scientific, Singapore, 2000.

R. Seaman. Absolute and differential accuracy of analyses achievable with specified observation network charcteristics. *Monthly Weather Review*, 105:1211–1222, 1977.

V. Sherlock, A. Collard, S. Hannon, and R. Saunders. The Gastropod fast radiative transfer model for advanced infrared sounders and characterization of its errors for radiance assimilation. *J. Appl. Meteorol.*, 42:1731–1747, 2003.

D.C. Sorensen. Implicit application of polynomial filters in a k-step arnoldi method. *SIAM J. Matrix Analysis and Applications*, 13:357–385, 1992.

J. L. Steward, I.M.Navon, M. Zupanski, and N. Karmitsa. Impact of non-smooth observation operators on variational and sequential data assimilation for a limited-area shallow water equations model. *Quart. Jour. Roy. Met Soc.*, 138:323–339, 2012.

L.M. Stewart. *Correlated observation errors in data assimilation*. PhD thesis, University of Reading, 2010. Available from http://www.reading.ac.uk/maths-and-stats/research/maths-phdtheses.aspx.

L.M. Stewart, S.L. Dance, and N.K. Nichols. Correlated observation errors in data assimilation. *Int.J.Numer.Meth.Fluids*, 56:1521–1527, 2008.

L.M. Stewart, S.L. Dance, N.K. Nichols, S. English, J. Eyre, and J. Cameron. Observation error correlations in IASI radiance data. In *Mathematics report series 01/2009: http://www.reading.ac.uk/maths/research/maths-report-series.asp*, 2009.

L.M. Stewart, S.L. Dance, N.K. Nichols, J. Eyre, and J. Cameron. Estimating interchannel observation error correlations for IASI radiance data in the Met Office system. *submitted*, 2012.

P. Weston. Progress towards the implementation of correlated observation errors in 4D-Var. Forecasting Research Technical Report 560, UK Met Office, 2011.

D.S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego, 1995.

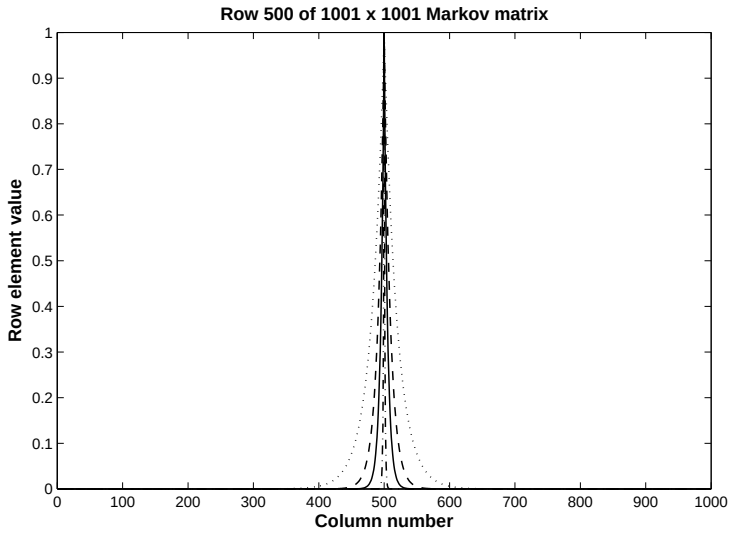


Figure 1: Middle row of a 1001×1001 Markov matrix, (9). Dash-dot line $L_R = 0.01$ m, full line $L_R = 0.05$ m, dashed line $L_R = 0.1$ m, dotted line $L_R = 0.2$ m.

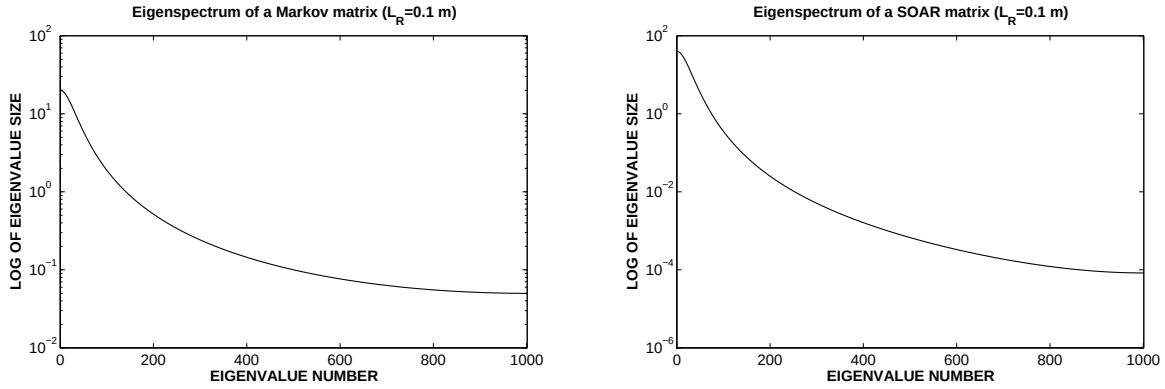


Figure 2: (a)Eigenspectrum of a 1001×1001 Markov error correlation matrix; (b)Eigenspectrum of a 1001×1001 SOAR error correlation matrix

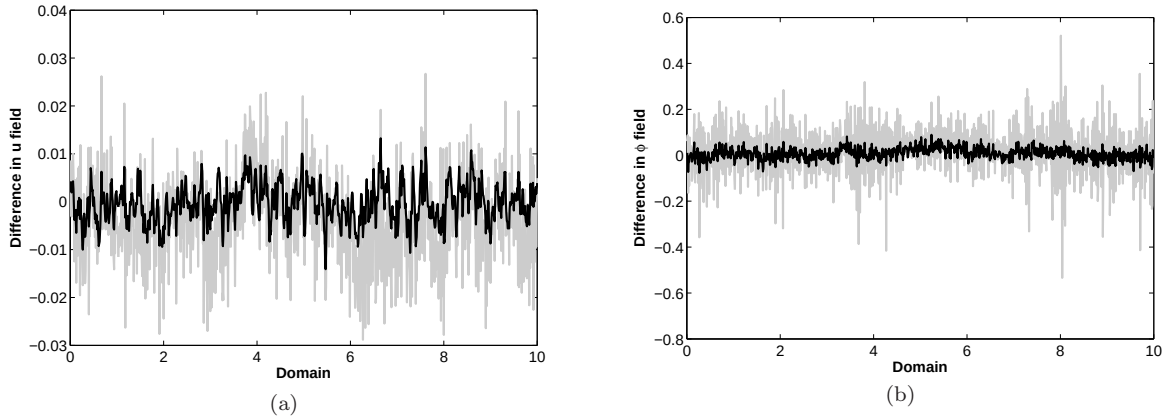
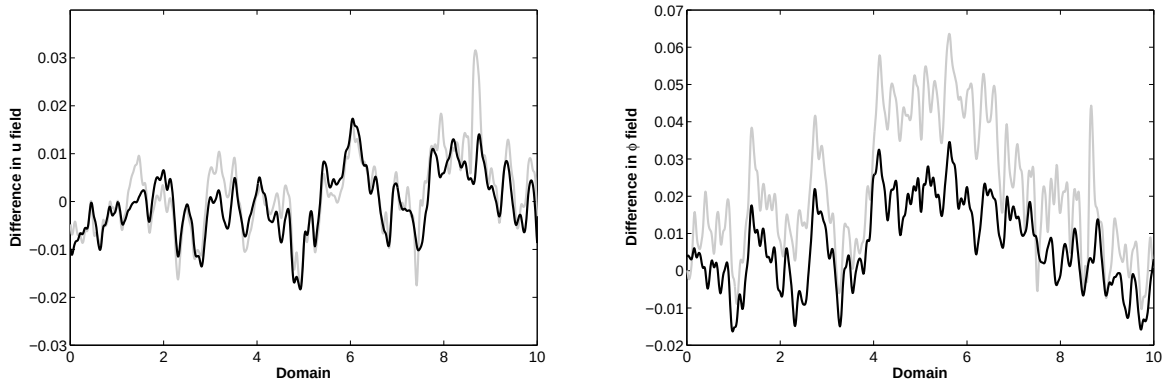


Figure 3: Analysis errors in (a) u field and (b) ϕ field at the start of the time window. The grey line is for a diagonal approximation and the black line is for a Markov approximation with $L_R = 0.2$m.



(a) (b)

Figure 4: Analysis errors in (a) u field and (b) ϕ field at the centre of the time window, $t = 50$. The grey line is for a diagonal approximation and the black line is for a Markov approximation with $L_R = 0.2$m. Note that the scale in panel (b) is different from Fig. 3b.

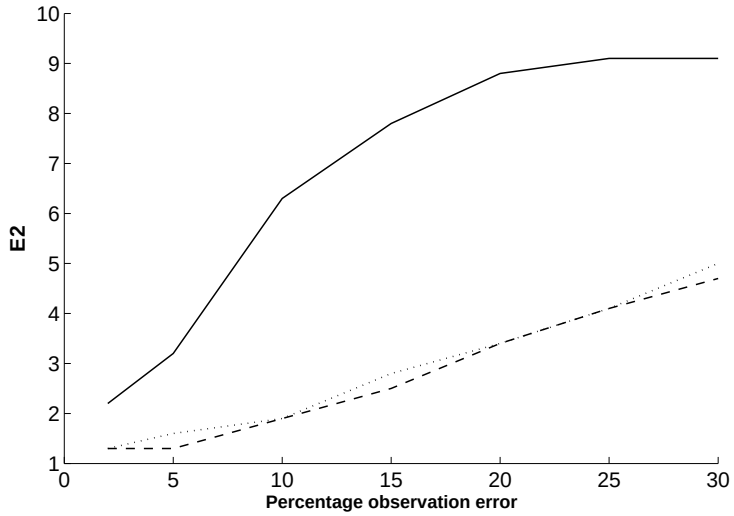


Figure 5: Plot of E2 against level of observation noise for u field. The solid line is for the diagonal approximation, the dashed line for the ED approximation with $k = 50$ and the dotted line for the Markov approximation with $L_R = 0.05$m.

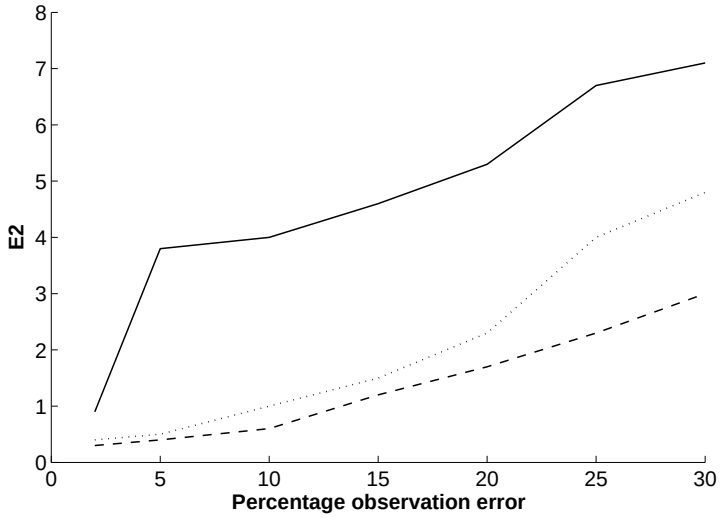


Figure 6: As in Figure 5 but for ϕ field.

Approximation	E1: $\left\|\overline{x}_{R_f} - x^*\right\|_2$	$\left\|\overline{x}_{R_f} - \overline{x}_{R_t}\right\|_2$	E2 (%)
Truth	0.20	0	0
Diagonal	0.30	0.23	7.2
2 × Diagonal	0.31	0.23	7.2
4 × Diagonal	0.31	0.24	7.5
Markov ($L_R = 0.2$)	0.21	0.06	1.9
Markov ($L_R = 0.1$)	0.20	0	0
Markov ($L_R = 0.05$)	0.21	0.05	1.6
Markov ($L_R = 0.01$)	0.27	0.18	5.6
ED ($k = 10$)	0.28	0.19	5.9
ED ($k = 20$)	0.28	0.19	5.9
ED ($k = 50$)	0.25	0.15	4.7
ED ($k = 100$)	0.23	0.10	3.1

Table 1: Analysis errors in u field at $t = 0$ for different approximations to a Markov error covariance matrix ($\left\|\overline{x}_R\right\|_2 = 3.20$)

Approximation	E1: $\left\|\overline{x}_{R_f} - x^*\right\|_2$	$\left\|\overline{x}_{R_f} - \overline{x}_{R_t}\right\|_2$	E2 (%)
Truth	2.35	0	0
Diagonal	3.61	3.04	4.9
2 × Diagonal	3.85	3.32	5.3
4 × Diagonal	4.11	3.61	5.8
Markov ($L_R = 0.2$)	2.41	0.54	0.9
Markov($L_R = 0.1$)	2.35	0	0
Markov ($L_R = 0.05$)	2.42	0.67	1.1
Markov ($L_R = 0.01$)	3.06	2.27	3.6
ED ($k = 10$)	3.97	3.25	5.2
ED ($k = 20$)	3.80	3.03	4.8
ED ($k = 50$)	3.33	2.39	3.8
ED ($k = 100$)	2.77	1.56	2.5

Table 2: Analysis errors in ϕ field at $t = 0$ for different diagonal approximations to a Markov error covariance matrix ($\left\|\overline{x}_R\right\|_2 = 62.64$)

Approximation	E1: $\left\|\overline{x}_{R_f} - x^*\right\|_2$	$\left\|\overline{x}_{R_f} - \overline{x}_{R_t}\right\|_2$	E2 (%)
Truth	0.11	0	0
Diagonal	0.31	0.28	8.8
2 × Diagonal	0.32	0.29	9.1
4 × Diagonal	0.32	0.30	9.4
Markov ($L_R = 0.2$)	0.13	0.07	2.2
Markov ($L_R = 0.1$)	0.15	0.11	3.4
Markov ($L_R = 0.05$)	0.18	0.15	4.7
Markov ($L_R = 0.01$)	0.27	0.25	7.8
ED ($k = 10$)	0.26	0.24	7.5
ED ($k = 20$)	0.23	0.20	6.3
ED ($k = 50$)	0.15	0.11	3.4
ED ($k = 100$)	0.13	0.07	2.2

Table 3: Analysis errors in u field at $t = 0$ for different approximations to a SOAR error covariance matrix ($\left\|\overline{x}_R\right\|_2 = 3.19$)

Approximation	E1: $\left\|\overline{x}_{R_f} - x^*\right\|_2$	$\left\|\overline{x}_{R_f} - \overline{x}_{R_t}\right\|_2$	E2 (%)
Truth	0.57	0	0
Diagonal	3.36	3.32	5.3
2 × Diagonal	3.59	3.55	5.7
4 × Diagonal	3.99	3.95	6.3
Markov ($L_R = 0.2$)	0.81	0.63	1.0
Markov($L_R = 0.1$)	1.18	1.06	1.7
Markov ($L_R = 0.05$)	1.69	1.60	2.6
Markov ($L_R = 0.01$)	2.89	2.84	4.5
ED ($k = 10$)	3.90	3.87	6.2
ED ($k = 20$)	3.71	3.67	5.9
ED ($k = 50$)	1.56	1.45	2.3
ED ($k = 100$)	1.06	0.85	1.4

Table 4: Analysis errors in ϕ field at $t = 0$ for different diagonal approximations to a SOAR error covariance matrix ($\|\overline{x}_R\|_2 = 62.54$)