THE UNIVERSITY OF READING

DEPARTMENTS OF MATHEMATICS AND METEOROLOGY

Correlated observation errors in data assimilation

Laura M. Stewart

Thesis submitted for the degree of

Doctor of Philosophy

December 2009

Abstract

Data assimilation techniques combine observations and prior model forecasts to create initial conditions for numerical weather prediction (NWP). The relative weighting assigned to each observation in the analysis is determined by the error associated with its measurement. Remote sensing data often have correlated errors, but the correlations are typically ignored in NWP. As operational centres move towards high-resolution forecasting, the assumption of uncorrelated errors becomes impractical. This thesis provides new evidence that including observation error correlations in data assimilation schemes is both feasible and beneficial. We study the dual problem of quantifying and modelling observation error correlation structure. Firstly, in original work using statistics from the Met Office 4D-Var assimilation system, we diagnose strong cross-channel error covariances for the IASI satellite instrument. We then see how in a 3D-Var framework, information content is degraded under the assumption of uncorrelated errors, while retention of an approximate correlation gives clear benefits. These novel results motivate further study. We conclude by modelling observation error correlation structure in the framework of a one-dimensional shallow water model. Using an incremental 4D-Var assimilation system we observe that analysis errors are smallest when correlated error covariance matrix approximations are used over diagonal approximations. The new results reinforce earlier conclusions on the benefits of including some error correlation structure.

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Laura Stewart

Acknowledgements

Firstly, I would like to thank my supervisors Dr Sarah Dance and Prof. Nancy Nichols for their personal and academic support, without which I could not have completed this research. I am also grateful to Dr Amos Lawless for his assistance with my work. I would also like to thank my Met Office supervisor Dr John Eyre whose knowledge of my research field was invaluable. Additionally I would like to thank Dr Stephen English, Dr James Cameron and Fiona Hilton from the Met Office, who helped enormously with the operational aspects of my work. I also acknowledge the financial support of the National Environmental Research Council (NERC) and the Met Office.

To all my friends in the Mathematics department, thank you for sharing the good and the bad (!) with me over the last three years. To all my friends outside the world of maths, thank you for tolerating the world of maths over the last three years! To John, thank you for keeping me grounded by reminding me that I was still just a 'skiving student' regardless of my work load(!); I will never forget the love and support you gave me over the years. I would also like to thank my parents for their continued support of my education; their hard work and success has been an inspiration to me. Finally, thank you to Joe, Coral and Georgia who are a constant reminder that there is more to life than burying one's head in a book!

Contents

List of Figures

Acronyms

NWP	Numerical weather prediction
3D-Var	Three dimensional variational data assimilation
4D-Var	Four dimensional variational data assimilation
IASI	Infrared Atmospheric Sounding Interferometer
SWEs	Shallow water equations

Data assimilation notation

x	state vector
x^t	'true' state vector
x^b	background state vector
x^a	optimal analysis state vector
y	observation vector
h	observation operator
H	linearisation of observation operator
B	background error covariance matrix
R	observation error covariance matrix
C	observation error correlation matrix
D	observation error variance matrix
S_a	analysis error covariance matrix
m	nonlinear forward model
M	linear forward model
M^T	adjoint model
K	Kalman gain matrix
J	cost function
J_b	background term of cost function
J_o	observation term of cost function
d_b^o	background innovation vector
d_a^o	analysis innovation vector

Chapter 1

Introduction

Data assimilation techniques are used to exploit information contained in observational data, previous forecasts and atmospheric dynamics for the purpose of weather forecasting. By statistically weighting this contributing information, data assimilation produces the best estimate of the current state of the atmosphere; this is used as the initial conditions for a model forecast. The weighted importance of each component in the assimilation is determined by its associated error. The chaotic nature of the atmosphere requires that the initial conditions be accurately specified to avoid rapid error growth [64], and thus the correct specification of the weighting errors is vital.

1.1 Motivation

In numerical weather prediction (NWP) the governing equations used to describe the behaviour of the atmosphere contain approximately 10^7 variables, and are sampled by order 10^6 observations in a 6 hour synoptic period. The observations are provided by the Global Observing System (GOS) [4] and include in-situ and remotely sensed measurements, each with an associated error structure. We treat observation errors as independent with type, i.e, radar observation errors are independent of aircraft observation errors, but dependency often exists between observations measured by the same instrument. Satellite observations typically have horizontally and vertically correlated errors. Origins of these errors include observation spatial proximity, contrasting model and observation resolutions, and observation pre-processing. Surface-based observations are also affected by correlated errors but their typically lower density means the effects of the correlated error are less significant. The size of the problem in NWP restricts the storage of the extra information provided by the error correlations. In operational weather prediction centres around the world, the data assimilation is most often performed under the assumption of uncorrelated satellite observation errors.

The assumption of zero correlations is often used in conjunction with data thinning methods such as superobbing [5]. This reduces the density of data by averaging the properties of observations in a region, and assigning this average as a single observation value. Under such assumptions, increasing the observation density beyond some threshold value has been shown to yield little or no improvement in analysis accuracy [60], [21]. Although discarding available information may be appropriate when the spatial resolution of the observations is denser than the model grid, recent technological advances have challenged the practicality of such methods. The Unified Model at the Met Office is run operationally at a 4km horizontal resolution, but the increasing demand for nowcasting and convective scale modelling has motivated the move towards a UK area model with resolution of order 1.5km [55]. Under such conditions there is a requirement to retain all the available data to provide details on the appropriate scales, and thus an alternative approach to dealing with observation error correlations is needed.

Approximating observation error correlation is a relatively new direction of research but progress has been made. In [43] circulant matrices were used to approximate a Toeplitz observation error covariance matrix. Results showed that incorrectly assuming uncorrelated observation errors gave misleading estimates of information content. In [34] Fisher proposed assigning a block-diagonal structure to the observation error covariance matrix, with (uncorrelated) blocks corresponding to different instruments or channels. Using this technique, individual block matrices were approximated by a truncated eigendecomposition. On a simple domain, spurious long-range correlations were observed.

1.2 Thesis aims

In this thesis we expand on the existing body of work on modelling observation correlation structure. We first quantify observation error correlation structure for an operational satellite instrument. The statistical results we obtain are new and motivate the need to include satellite observation error correlations in data assimilation algorithms. By performing variational data assimilation experiments in a three-dimensional and four-dimensional framework, we then examine the impact of new and existing approximations to error correlation structure. In undertaking this work we wish to address the following questions:

- **What is the true structure of the observation error correlations?**

  Satellite observations typically carry correlated observation errors; the magnitude and structure of the error covariances can be difficult to quantify [83]. In order to generate a good approximation, we must first have an accurate estimate of the true error correlation structure.

- **What approximations are available to model error correlation structure? What is their impact on data assimilation diagnostics?**

  For an approximate error correlation structure to be implemented operationally it must be computationally feasible. We therefore present approximations which will not be overly demanding on computational time and storage. These approximations are then ranked on their impact on different assimilation retrieval measures.

- **How well do these approximations perform in a data assimilation experiment? Is it better to model observation error correlation structure incorrectly than not at all?**

  It remains unclear whether a better representation of the true error correlation structure will be evident in the analysis of a data assimilation problem. The final experiments quantify analysis accuracy under different diagonal and correlated approximations to a simulated error correlation structure.

1.3 Thesis outline

The thesis is structured as follows. Chapter 2 introduces the concepts of data assimilation and remote sensing. Although data assimilation is a relatively new science, rapid progress has been made. We briefly discuss the evolution of data assimilation algorithms, and focus on the two methods used in the thesis: three-dimensional variational data assimilation (3D-Var) and four-dimensional variational data assimilation (4D-Var). Observing System Experiments (OSEs) at the European Centre for Medium Range Weather Forecasting (ECMWF) and elsewhere have shown that the inclusion of satellite data in a 4D-Var algorithm results in the greatest positive forecast impact over all observation types [4], [89]. Here we review the physics and operational treatment of satellite data, and highlight its importance in current NWP. Details on the nature and origin of observation error covariances are then given. The chapter is concluded with a description of the techniques used to quantify these error covariances.

In Chapter 3 we address the second question posed in Section 1.2 and present possible matrix representations of the observation error covariance structure. Three different approximating structures are described: diagonal [14], circulant [78], [43], and eigendecomposition [34] approximations. We pay careful attention to the feasiblity of including these matrices operationally. The effectiveness of each approximation can be evaluated using several parameters. In the second part of the chapter we describe the following retrieval diagnostics: analysis error covariance matrix, information content, and matrix and vector norms. These properties can be used to determine how well each approximation performs in a data assimilation system.

Chapter 4 addresses the first question posed in Section 1.2 and contains new results on quantifying error correlation structure. We first introduce the Infrared Atmospheric Sounding Interferometer (IASI) satellite instrument, and describe how its measurements are processed in the Met Office incremental 4D-Var assimilation scheme. Using a post-analysis diagnostic based on variational data assimilation theory [25] and statistics from the Met Office system, we successfully quantify the cross-channel error correlations between IASI measurements. Diagnosed error covariances are given for the pre-processing

1D-Var assimilation and the main 4D-Var assimilation. Comparisons are made with the current operational error variances.

More novel results are presented in Chapter 5, where we consider modelling correlation structure in a 3D-Var framework. Being a simpler system than the 4D-Var framework, the results can be analysed more easily. Using information content measures, we quantify the success of each matrix approximation described in Chapter 3 in modelling an empirically derived observation error correlation structure. The impact of each approximation can then be evaluated relative to the truth. Conclusions based on numerical evidence are drawn for different background error structures and constructions of the analysis error covariance matrix. The original results in this chapter address the second thesis question posed in Section 1.2.

Motivated by the results in Chapter 5, Chapter 6 describes the mathematical framework needed to extend this investigation to a 4D-Var setting. We introduce a set of one-dimensional shallow water equations (SWEs) [54], used to represent simplified atmospheric dynamics, and describe the continuous analytical and discretised numerical models. We then develop a new incremental 4D-Var data assimilation system for the 1D SWEs which models observation error correlation structure using diagonal, Markov and eigendecompostion matrix approximations. Finally we describe the coding tests used to test the validity of the model assumptions.

Chapter 7 contains further new results which address the final thesis question posed in Section 1.2. Using the model and data assimilation system described in Chapter 6, this chapter extends the findings in Chapter 5 and examines the impact of correlated error covariance matrix approximations in a 4D-Var framework. We first describe the experiment methodology and the error diagnostics used. We then determine the different realisations of the approximate observation error covariance matrices to be used in the experiments. Assimilation accuracy is then evaluated for each approximation under different simulations of the true error distribution. The novel results motivate further study in this field.

Finally in Chapter 8 we summarise the work done and draw conclusions from these experimental results regarding the effectiveness of modelling observation error correlations in data assimilation algorithms. We also make suggestions for possible further work in this area.

Chapter 2

Data assimilation and remote sensing

2.1 Introduction

In NWP an accurate high-resolution representation of the current state of the atmosphere is needed as an initial condition for the propagation of a weather forecast. Despite the availability of millions of observations, these alone are insufficient to fully represent the state of the atmosphere. Additional knowledge about atmospheric dynamics and physics is needed to compensate for the inadequacies of the observations; these include under-determinancy, measurement error, and observations that are non-linearly related to atmospheric variables. 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.

In an operational setting, the order of the problem in NWP is approximately 10^7 unknown variables with approximately 10^6 observations available over a 6 hour time window [7]. At the UK Met Office the operational forecast suite is run 24 hours a day and 365 days a year. There is a limited amount of computer time in which to run a forecast model: for example, a typical North Atlantic European (NAE) model forecast slot is only 65 minutes long [1]. Therefore data assimilation algorithms must be able to compute the analysis quickly and efficiently for large scale problems.

The complexity of data assimilation in NWP has increased significantly since the first objective analysis algorithms were introduced by Gilchrist and Cressman in 1954 [38]. The Cressman analysis scheme set the analysis equal to a background state plus a weighted contribution of the observations dependent on their spatial proximity to model grid points. The background state was given by the best available approximation to the present state, such as a climatology or a previous forecast. A later extension to this method was a more general algorithm, the successive correction method [6], in which the weights were defined so that the background state had an impact even at the observation points. The procedure was also iterative to enhance the smoothness of the analysis. However, neither of these methods were entirely robust or concerned with the optimality of the analysis. For example, if we had a climatology that we knew was of good quality, then it was possible we would still modify it using values provided from poor quality observations. More statistical techniques were needed to ensure the observations and the background state were weighted in a manner appropriate to their uncertainty and the underlying physical features of the system.

Several statistical assimilation techniques have been developed for meteorology; numerical cost, robustness, and the optimality of the solution generated are all important issues in their operational use. Generally these techniques can be classified as sequential or variational, intermittent or continuous [8]. Sequential assimilation algorithms solve the system of equations needed for an optimal solution explicitly; variational algorithms solve the equations implicitly through the minimisation of a cost function. Intermittent methods assimilate all the observations as if they were taken at the same time, while continuous methods assimilate the observations at the time of measurement.

Clearly intermittent methods are simpler to implement since any algorithm is free of the additional constraint of time, but because real observations are available at different times, continuous methods are more prevalent in operational data assimilation algorithms. The work in this thesis is concerned with variational assimilation algorithms using both intermittent and continuous methods. In the next two sections we describe three-dimensional variational assimilation (3D-Var), an intermittent method, and four-dimensional variational assimilation (4D-Var), a continuous method.

2.2 3D-Var and Bayes' Theorem

NWP is concerned with generating the 'best' analysis given some prior atmospheric information and a set of observations. The best analysis can be defined as that which gives the best subsequent forecast, without compensating for errors in the forecast model [62].

Consider a discretised representation of the true state of the atmosphere $x^t \in \mathbb{R}^n$, 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 $x^b \in \mathbb{R}^n$ and the actual observations $y \in \mathbb{R}^m$, where m is the total number of measurements. The background state and observations will be approximations to the true state of the atmosphere,

$$x^b = x^t + \epsilon^b, \tag{2.1}$$

$$y = h(x^t) + \epsilon^o, \tag{2.2}$$

where ϵ^b and ϵ^o are the background and observation errors, respectively, and h is the possibly nonlinear observation operator mapping from state space to measurement space; for example, a fast radiative transfer model which simulates radiances from an input atmospheric profile. The errors are assumed unbiased and mutually independent,

$$\mathbb{E}[\epsilon^b] = \mathbb{E}[\epsilon^o] = \mathbb{E}[\epsilon^b(\epsilon^o)^T] = \mathbb{E}[\epsilon^o(\epsilon^b)^T] = 0, \tag{2.3}$$

and also to have covariances $B = \mathbb{E}[\epsilon^b(\epsilon^b)^T]$ and $R = \mathbb{E}[\epsilon^o(\epsilon^o)^T]$.

The analysis state is the solution to the inverse problem of determining the 'best' estimate of x^t which satisfies (2.1) and (2.2). This analysis is sometimes known as the maximum *a posteriori* estimate and can be derived in terms of probability distribution functions (pdfs) using Bayesian methods.

Bayes' theorem [56] states that the posterior probability of event A, given that event B occurs, is proportional to the prior probability of A multiplied by the probability of event B given that event A occurs;

$$\mathbb{P}(A|B) \propto \mathbb{P}(B|A)\mathbb{P}(A). \tag{2.4}$$

Applying this idea to data assimilation theory [62]: event A is the state of the system and event B is the sample of observations. Therefore maximising the posterior probability $\mathbb{P}(A|B)$ is equivalent to finding the state of the system which best represents the observations.

Assuming that all pdfs are Gaussian, let $P_b(x)$ be the prior pdf of the state and $P_o(y|x)$ be the conditional probability of the observations given the state;

$$\begin{aligned} P_b(x) &= k_1 \exp\{-\frac{1}{2}(x - x^b)^T B^{-1}(x - x^b)\}, && (2.5) \\ P_o(y|x) &= k_2 \exp\{-\frac{1}{2}(y - h(x))^T R^{-1}(y - h(x))\}, && (2.6) \end{aligned}$$

where x is the analysis state, h is the possibly nonlinear observation operator, and k_1 and k_2 are normalisation factors independent of x. Using Bayes' Theorem, we obtain the posterior (or analysis) pdf of the state,

$$P_a(x|y) = k_3 \exp\{-\frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) - \frac{1}{2}(y - h(x))^T R^{-1}(y - h(x))\}, \tag{2.7}$$

where k_3 is also independent of x. By taking natural logs of both sides of (2.7), we can see that maximising P_a is equivalent to minimising $-\ln(P_a)$, i.e, minimising the quadratic cost function

$$J(x) = \frac{1}{2}(x - x^b)^T B^{-1}(x - x^b) + \frac{1}{2}(y - h(x))^T R^{-1}(y - h(x)). \tag{2.8}$$

The cost function (2.8) measures the distance from the analysis state to the observations and the background state, weighted by the inverse of their respective error covariances. The cost function minimisation can be solved approximately to obtain the best linear unbiased estimate (BLUE) [49], x_a:

$$\begin{aligned} x^a &= x^b + K(y - h(x^b)), && (2.9) \\ K &= BH^T(HBH^T + R)^{-1}, && (2.10) \end{aligned}$$

where H is the linearised observation operator given by $H = \frac{\partial h}{\partial x} |_{x_b}$ and K is the Kalman gain matrix specifying the optimal weighting of the observations in the analysis. When

h is linear the cost function minimisation is solved exactly, and the associated analysis error covariance matrix is given by

$$S_a = (H^T R^{-1} H + B^{-1})^{-1}. \tag{2.11}$$

The BLUE can be seen as a further extension of the initial Cressman analysis algorithm, since the analysis is still obtained through an explicit combination of the observations and the background [49].

Sequential assimilation algorithms approximate (2.9) and (2.10) directly; optimal interpolation is an example of a suitable algorithm previously used operationally [61]. In current operational NWP, the error covariance matrices are too large to be used explicitly in global assimilation problems, and therefore an implicit variational alternative is needed.

The statistical method used operationally by the UK Met Office between 1999 and 2004 was three-dimensional variational data assimilation (3D-Var) [63]. This algorithm seeks the minimum of (2.8) by performing several evaluations of the cost function and its gradient in order to approach the minimum using a suitable descent algorithm [17]. The size of the error covariance matrices is still an issue with the main cost in the cost function evaluation lying in inverting B and R. Although conceptually useful, 3D-Var treats observations as if they were valid at the same point in time, which is clearly an unrealistic assumption. An extension to the standard 3D-Var is the First Guess at the Appropriate Time (3D-FGAT) method [91]. This technique calculates the observation increments $(y - h(x))$ at their appropriate measurement times, but then applies the increments at a single analysis time. A further sophisticated extension of this method is four-dimensional variational data assimilation.

2.3 4D-Var

Four dimensional variational data assimilation (4D-Var) is an extension to 3D-Var which allows the distribution of observations within a time interval (t_0, t_n). The continuous nature of the 4D-Var approach is illustrated in Figure 2.1. The current 4D-Var problem is a modification of the original variational method proposed in [79] where information from a time-sequence of observations was combined with a numerical model. The objective of 4D-Var is to minimise the cost function,

$$\begin{aligned} J(x_0) &= \frac{1}{2}(x_0 - x^b)^T B^{-1}(x_0 - x^b) \\ &\qquad + \frac{1}{2}\sum_{i=0}^{n}(h_i(x_i) - y_i)^T R_i^{-1}(h_i(x_i) - y_i) \qquad (2.12) \\ &\equiv J_b + \sum_{i=0}^{n} J_{o,i} \\ &\equiv J_b + J_o \qquad (2.13) \end{aligned}$$

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

$$x_{i+1} = m(t_i, t_{i+1}, x_i), \qquad (2.14)$$

where x_i is the model state at time t_i, $m(t_0, t_i, x_0)$ is the nonlinear model evolving x_0 from time t_0 to time t_i, x_b is the background field given by a previous forecast, y_i is the observation vector at time t_i, and h_i is the nonlinear observation operator at time t_i. The strong constraint given by equation (2.14) implies the model is assumed to be perfect.

The cost function measures the distance between the model state x_0 and the background at the start of the time interval t_0 (the J_b term), and the sum of the observation innovations $(h_i(x_i) - y_i)$ computed with respect to the time of the observation (the $J_{o,i}$

terms). 4D-Var therefore provides an initial condition such that the forecast best fits the observations within the whole assimilation interval.

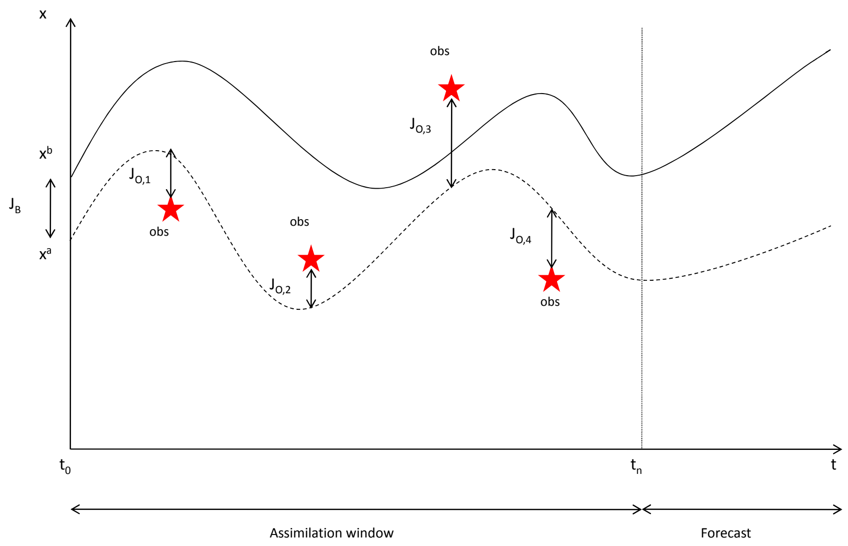


Figure 2.1: Diagrammatic representation of 4D-Var method: minimise the squared distance between the analysis x^a and the background x^b at the beginning of the assimilation window (J_B term) plus the squared distance between the observations (red stars) and the forecast state throughout the assimilation window ($J_{O,i}$ terms).

Although demonstrably superior to 3D-Var [76], the size of the problem in NWP inhibits the direct solution of the 4D-Var cost function (2.12). When 4D-Var methods were originally being investigated it was determined that in order to use a 4D-Var algorithm, a significantly faster computer or a substantial algorithmic improvement was needed [18]. The formulation of an incremental 4D-Var algorithm provided this improvement.

2.3.1 Incremental 4D-Var

Incremental 4D-Var, as proposed by Courtier et al [18], reduces the cost of the 4D-Var algorithm by approximating the full nonlinear cost function (2.12) 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 (2.14). Each cost function minimisation is performed iteratively and the resultant solution is used to update the nonlinear model trajectory. The iterative minimisation procedure is known as the inner loop; the update step is known as the outer loop. Full details of the procedure are described in the following iterative algorithm [52] where k is the iteration number:

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

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. Define an increment $\delta x_0^{(k)} = x_0^{(k+1)} - x_0^{(k)}$.

5. Start the inner loop minimisation. Find the value of $\delta x_0^{(k)}$ that minimises the incremental cost function

$$\begin{aligned} J^{(k)}(\delta x_0^{(k)}) &= \frac{1}{2}(\delta x_0^{(k)} - (x^b - x_0^{(k)}))^T B^{-1}(\delta x_0^{(k)} - (x^b - x_0^{(k)})) \\ &\quad + \frac{1}{2}\sum_{i=0}^{n}(H_i \delta x_i^{(k)} - d_i^{(k)})^T R_i^{-1}(H_i \delta x_i^{(k)} - d_i^{(k)}) \end{aligned} \tag{2.15}$$

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 around the state $x_i^{(k)}$.

6. Update the guess field using

$$x_0^{(k+1)} = x_0^{(k)} + \delta x_0^{(k)}.$$

7. Repeat outer loop (steps 2 - 6) until the desired convergence is reached.

An advantage of this method is that the inner loop cost functions can be simplified; for example, by performing the inner loop minimisation at a lower spatial resolution. Incremental 4D-Var is the data assimilation algorithm currently used at several NWP centres such as the Met Office [76] and ECMWF [75]. However, correctly specifying the observation error structure is still an important issue under this new formulation.

2.3.2 Practical implementation

We now consider the practical details of implementing a 4D-Var system. The issues considered below will be specific to individual assimilation problems. In Chapter 6 we will describe the practical implementation issues relative to a one-dimensional shallow water model which we will use in experiments in Chapter 7.

Minimisation algorithm

There are several algorithms suitable for the inner loop minimisation required in 4D-Var. In this work we use the conjugate gradient method (CGM) [36]. This gradient descent method minimises (2.12) or (2.15) by optimally choosing conjugate search directions such that the algorithm does not minimise in the same direction twice.

The CGM requires the calculation of the cost function (2.15) and its gradient with respect to the model state at time t_0. The gradient of the cost function (2.15) is given by

$$\nabla J = \nabla J_b + \nabla J_o \tag{2.16}$$

where

$$\begin{aligned} \nabla J_o &= -\sum_{i=0}^{n} M_i^T H_i^T R_i^{-1}[H_i \delta x_i^{(k)} - d_i^{(k)}] && (2.17) \\ \nabla J_b &= -B^{-1}[\delta x_0^{(k)} - (x^b - x_0^{(k)})] && (2.18) \end{aligned}$$

where M_i^T is the adjoint of the linear model $M(t_i, t_{i+1}, x^{(k)})$. From (2.15), (2.17) and (2.18) we observe that both a forward linear model M and a backwards adjoint model M^T are required for the calculation of the cost function and its gradient, respectively.

Tangent linear model

In Section 2.3.1 we described how an incremental 4D-Var assimilation requires a linear approximation to a nonlinear model. The linear approximation used in this work is the tangent linear model (TLM) [50]. To generate the TLM we assume that a nonlinear model and its linearised version exhibit similar behaviour for a period known as the validity time. We then consider the nonlinear model m_i applied to a perturbation δx from a state x, and perform a Taylor expansion about x,

$$m_i(x + \delta x) = m_i(x) + M_i(x)\delta x + \frac{1}{2}\hat{M}_i(x)\delta x^2 + \text{higher order terms} \qquad (2.19)$$

where $M_i = \frac{\delta m_i}{\delta x}$ is the Jacobian of m_i found by differentiating the discrete nonlinear model equations with respect to the state x, and $\hat{M}_i = \frac{\delta^2 m_i}{\delta x^2}$ is the matrix of second derivatives. By taking first order terms only, we can approximate the nonlinear model m_i with a linear discrete model M_i,

$$m_i(x + \delta x) = m_i(x) + M_i(x)\delta x. \qquad (2.20)$$

Adjoint model

The adjoint model M^T provides us with a system of model equations, solvable backwards in time to obtain the gradient of the cost function [90]. In practice the discrete adjoint equations are constructed directly from the tangent linear model code using an 'automatic adjoint' method [37]. The principal application of adjoint models is in sensitivity analysis, and further details on the derivation, properties, and applications of adjoint models are given in [27], [28]. In Chapter 6 we comment on how the TLM and adjoint code are constructed for a one-dimensional shallow water model, and describe the coding tests needed to determine their suitability for inclusion in a data assimilation algorithm.

We have seen in Section 2.2 and 2.3 how the complexity of data assimilation techniques has increased in recent years. In order to ensure that current and future data assimilation methods are fully utilised, there is a need for accurate and plentiful observations.

2.4 Remotely sensed data

Millions of observations are available for the running of operational data assimilation algorithms; at the ECMWF, 4-8 million observations are assimilated every 12 hours [4]. As data assimilation techniques improve, as we have seen happen recently with the transition from 3D-Var to 4D-Var, the demands on observation density will only become larger. The current availability of satellite data will likely be a contributing factor in the success of future techniques.

Since the first atmospheric sounders were launched on the Nimbus 3 satellite in 1969

[78], satellite observations have been used to complement the ‘conventional’ observation network. Conventional observations are typically in-situ measurements of temperature, wind, pressure and humidity, observed directly by an instrument on a radiosonde or an aircraft, for example. The static or human dependent nature of these observations results in significant data voids on the globe, e.g, very few surface observations are available over sub-Saharan Africa, and no aircraft observations are available for non-mainstream flight paths. Satellite remote sensing allows us to obtain data from places where it is inconvenient or even virtually impossible to obtain an in-situ measurement.

The main providers of satellite data to global NWP centres are the American (NASA and NOAA), European (ESA and EUMETSAT), and Japanese (JAXA and JMA) space agencies. Global coverage from satellite observations is ensured by the complementarity of the geostationary (GEO) and low earth orbiting (LEO) platforms operational at each centre. Contrary to conventional observations, satellite measurements do not directly relate to desired atmospheric quantities; this indirect nature is the feature of remote sensing rather than a physical ‘remoteness’. The subsequent treatment and utilisation of satellite observations in data assimilation algorithms is therefore a complex physical problem.

2.4.1 Satellite observation physics

Satellite instruments measure the electromagnetic radiation (or radiance) L that reaches the top of the atmosphere at a given frequency ν. Electromagnetic radiation travels in wave form at different frequencies, and is responsible for energy transfer within the atmosphere. The measured radiances are related to geophysical variables through the radiative transfer equation [30]

$$L(\nu) = (I_0)_\nu \tau_\nu(z_0) + \int_{z_0}^{\infty} B_\nu\{T(z)\} \frac{d\tau_\nu(z)}{dz} dz, \tag{2.21}$$

where

$(I_0)_\nu$ is the emission from the earth's surface at height z_0,

$\tau_\nu(z)$ is the vertical transmittance from height z to space,

$T(z)$ is the vertical temperature profile,

and $B_\nu\{T(z)\}$ is the corresponding Planck function profile.

Equation (2.21) is constructed under the assumption of an entirely one-dimensional transmittance along the instrument viewing path with no molecular scattering in and out of the beam. We assume no cloud or rain contributions, but these can be handled in the infra-red and microwave spectrum provided they are either entirely emission or absorption, and there is no significant scattering. The problem of cloudy radiance assimilation is discussed in detail in [30], [57], [70]; we will return to the problem in Section 2.4.4.

The radiative transfer equation is further explained by considering a solitary air parcel at some level in the atmosphere. The radiation emitted to space from this air parcel is determined by its temperature and the atmospheric density of the emitting gas within the parcel. A body at different temperatures emits different amounts of radiation. Atmospheric density decreases exponentially with height, and so the intensity of radiation reaching the top of the atmosphere is less for a parcel at the same temperature but at a higher atmospheric level. Also, the radiance emitted from a parcel of air close to the earth's surface may be entirely absorbed before it reaches the top of the atmosphere. Radiance measurements at different frequencies (or channels) will have different absorption characteristics, and therefore by sensing at different frequencies we obtain information on the vertical profile of the thermodynamic state and composition of the atmosphere.

A detailed overview of the satellite instrument technologies used to observe the atmosphere is given in [29]; we will briefly summarise the main aspects. In general, we categorise the frequencies (or channels) used in NWP into three different types: atmospheric sounding channels (passive instruments), surface sensing channels (passive instruments), and surface sensing channels (active instruments). Passive instruments sense natural radiation emitted by the earth's surface or the atmosphere, while active instruments emit radiation and sense the amount reflected or scattered back by the earth's surface or atmosphere. Details on the features of these channels are given in Table 2.1.

Channel type	Instrument type	Channel location	Use in NWP
Atmospheric Sounding	Passive	Infrared and microwave spectrum where main contribution to measured radiance is from the atmosphere	Atmospheric temperature and humidity
Surface Sensing	Passive	Window regions of infrared and microwave spectrum where the main contribution is from surface emission	Surface temperature emissivity Ocean surface wind Soil moisture
Surface Sensing	Active	Window regions of spectrum that actively illuminate the surface	Ocean winds Cloud monitoring (CloudSat,CALIPSO)

Table 2.1: Typical NWP channel properties

Now consider a channel (i.e, a certain frequency) where we know the primary absorber of radiation is a well-mixed gas with known concentration (i.e, oxygen or carbon dioxide). In equation (2.21) Planck's function $B_\nu\{T(z)\}$ relates the measured radiance intensity at a given frequency with the temperature of the absorbing substance; this is then weighted by the derivative of the transmittance profile $\frac{d\tau_\nu(z)}{dz}$. Therefore a radiance measurement at frequency ν can be interpreted as a weighted average of the atmospheric Planck function profile, where the weighting function is

$$\kappa(\nu) = \frac{d\tau_\nu(z)}{dz}. \tag{2.22}$$

As the derivative of the transmittance profile, an empirically derived weighting function is subject to knowledge of the absorption and density profile of the absorbing gas at a given frequency, as well as the vertical temperature profile.

The weighting function $\kappa(\nu)$ specifies the layer of the atmosphere from which the measured radiation originates. This layer will not correspond to a single modelled level, but rather incorporate several, with a varying radiance contribution from each (as represented by broad, peaked weighting functions). The altitude at which a weighting function peaks will depend on the strength of absorption in the given channel, i.e, frequencies at which the absorption is strong will have high peaking weighting functions. By selecting channels with different absorption strengths, we can build a series of weighting functions, which provide information on the radiance contribution, and hence mean temperature, of many layers.

There are two important characteristics of weighting functions that influence the use of satellite observations in NWP. The first is their broad width. A width of up to several kilometres hinders the ability of satellite sounders such as the High-resolution Infrared Radiation Sounder (HIRS) to identify small scale vertical atmospheric structures. However with the advent of instruments with a high spectral resolution, such as interferometers like the Atmospheric Infrared Sounder (AIRS) and the Infrared Atmospheric Sounding Interferometer (IASI), sharper weighting functions can be built. The second characteristic is the overlapping nature of the weighting functions for one instrument. A consequence of this is a lack of independent data for different atmospheric levels. Despite the several thousand channels observed by the AIRS and IASI instruments, the characteristics of their weighting functions result in the radiance measurements undersampling the smallest scales that are vertically resolved by the NWP model. Therefore the process of obtaining an atmospheric temperature or humidity profile from the set of radiance measurements becomes an ill-posed inverse problem, similar in nature to the general data assimilation problem.

2.4.2 Inverse retrieval problem

The problem of obtaining temperature products from radiances is ill-posed because we have only a finite number of radiance measurements for an unknown continuous function $T(z)$; any one combination of measurements could be from thousands of different profiles. Previous data assimilation schemes such as Optimal Interpolation required the explicit conversion of radiance observations to temperature profiles before the analysis. Typically a 1D-Var retrieval process using background information from a short range forecast was used. The result was an ‘optimal’ temperature profile solution that fits the background information and the measured radiances, respecting the uncertainty in both. Details on this process can be found in [32].

Using retrieval algorithms will however result in a correlation between the errors in the retrieval and the forecast background, both of which are subsequently used in the main assimilation. In a system of the size used operationally, representing these complicated error characteristics is very tricky. The new generation of variational analysis methods such as 3D-Var and 4D-Var avoid this issue and allow the direct assimilation of radiance information. The forecast background still provides the prior information needed to supplement the radiances, but it is not used twice and hence more complicated error characteristics are avoided. This approach also avoids the random and systematic errors introduced by unnecessary pre-processing such as angle adjustment and surface corrections, and allows faster access to data from new platforms (Advanced Microwave Sounding Unit (AMSU) data from NOAA-16 was assimilated operationally 6 weeks after launch [67]). However, although slightly lessened, the significant problems of background and observation error specification are still present.

2.4.3 Current satellite data usage

The Global Observing System (GOS) consists of several different observation types, but satellite data is the dominant contributor. At the ECMWF, satellite data accounts for 95% of the data used in assimilation, 90% of which is radiance data; over 20 million satellite observations are used every day, and this is expected to increase to 28 million by 2010 [4]. The current satellite data sources include radiances (i.e, IASI on Metop, AIRS on AQUA, SEVIRI on Meteosat-9), ozone (SBUV on NOAA-17), bending angles (GRAS on Metop), Atmospheric Motion Vectors (Meteosat-7/9), and sea surface parameters such as wind speed and wave height (Seawinds on QuikSCAT).

The meteorological impact of satellite data in operational NWP systems is illustrated by observing system experiments (OSEs) and information content studies. OSEs are performed regularly to assess the performance of individual components of the GOS; the impact of a GOS component is determined by adding it to the baseline assimilation system. In 2007, eliminating all satellite observations from the ECMWF assimilation system caused a skill reduction of 0.75 days in the northern hemisphere and 3 days in the southern hemisphere [4]. OSEs performed at the Met Office in 2003 and 2007 showed that when IASI and AIRS data was assimilated in 2007, the impact of eliminating microwave sounding data was significantly reduced relative to the 2003 assimilations [4].

Information content studies use a popular information measure called the degrees of freedom for signal (dof_S) to represent the amount of information available from a set of observations [77]. By eliminating different observation types, the contributing information from the missing component can be observed. The dof_S is defined to be

$$dof_S = \text{trace}(I - S_a B^{-1}) \tag{2.23}$$

where I is the identity matrix, and S_a (2.11) and B are the analysis and background error covariance matrices, respectively. In Chapter 3 we discuss information content measures in more detail. These measures are subsequently used in the experiments described in Chapter 5.

Experiments at the ECMWF in 2003 show a 25% reduction in the dof_S when no Advanced TIROS Operational Vertical Sounder (ATOVS) data was assimilated relative to the baseline of assimilating all observations [89]. The size of this drop was not mirrored in the elimination of any other data type. Such experiments demonstrate that the time and money spent on the procurement and utilisation of satellite data is worthwhile.

The evolution of satellite technology is ongoing and the demand for accurate, high-resolution data is increasing. It is therefore important that the procurement of satellite data is in line with these requirements. For example, the Meteosat third generation satellite proposed for launch in 2015 will include an infrared sounder whose spectral, and hence vertical, resolution will be comparable with IASI ($\approx$ 1km), but with improved horizontal ($<$ 10km) and temporal ($<$ 1hour) resolution. The result will be more frequent information on temperature and water vapour profiles suitable for the nowcasting demands of European weather agencies [31]. However, as the quantity and complexity of satellite data increases, important issues in its treatment must be addressed.

2.4.4 Current issues in satellite remote sensing

The horizontal and vertical resolution of numerical forecast models are sampled by millions of observations. Many of these observations are provided by high spectral resolution sounders, such as IASI on Metop and CrIS on NPP and NPOESS. Such observations are also used by climate, chemical and environmental research, and so it is unsurprising that the exploitation of high resolution data is an area of major scientific interest.

Instruments such as AIRS and IASI measure radiation in thousands of different channels and hence provide atmospheric temperature and composition information at a much higher accuracy and vertical resolution than previous sounders. However, better utilisation of this data is needed. Large quantities of high resolution observations are currently omitted from data assimilations because their underlying features are not well understood or cannot be accurately represented. Below we will briefly discuss the reasons for this omission, and the current work in the area.

Channel selection and data compression

The assimilation of all channels from a high spectral resolution sounder is neither feasible nor computationally efficient, and so channel selection methods are used. The aim of the selection process is to choose the set of channels providing the optimal subset of data to be inserted into the assimilation. A desirable set of channels will be large enough to accurately represent atmospheric variability but small enough to be assimilated efficiently within NWP systems.

Specific methods of channel selection based on objective criteria are described in [74]. The iterative method proposed by Rodgers [77] was found to generate the best results in respect to the lowest standard deviations of errors over the vertical profile. This method takes entropy reduction (ER) and degrees of freedom for signal (dof_S) as the objective criteria reflecting an improvement in the analysis error covariance matrix S_a. From the subset of pre-screened channels, the ER or dof_S is calculated for each non-selected channel and the channel with the largest value is chosen for inclusion. Before the next channel is chosen, S_a is updated so that information obtained from previously chosen channels is acknowledged before the next is selected, therefore accounting for redundancy between channels.

The Rodgers iterative channel selection method was performed for the AIRS [35] and IASI instruments [74], [15] and its robustness demonstrated under different specifications of the background error covariance matrix and different atmospheric profiles. However, not all available channels can be incorporated into such channel selection methods. The properties of certain absorbing gases (trace species) or external influencing factors at certain wavelengths mean channels incorporating them are blacklisted. For example, shortwave channels (wavelength $< 5\mu m$) can be affected by sunlight and surface emissivity in ways that cannot be represented easily in the forward model, and so are not chosen in preference to longwave channels that can provide the same information. A channel selection that is too static may lead to wasting crucial information.

NWP centres are now investigating alternative treatments of huge data volumes. One promising method is principal component analysis (PCA) [92]. The nature of PCA techniques is to approximate data vectors with many elements (i.e, IASI observations of 8461 channels) by a new correlated set of data vectors containing fewer elements. The procedure retains most of the variability and information of the initial data. Goldberg et al [40] demonstrated that PCA produces an efficient retrieval of atmospheric temperature, moisture and ozone, and an accurate reconstruction of over 2000 AIRS channels from 60 principal component scores. Also, a PCA-based noise filter for high spectral resolution infrared data was shown by Antonelli et al [2] to significantly reduce the random component of the instrument noise of the observations.

The reconstruction in PCA results in data vectors which are linear combinations of the initial data set. Therefore the errors in the reconstructed data set will be linear combinations of the initial errors, i.e, the error characteristics become much more complicated. It is expected that storing observations in principal component form, i.e, using reconstructed radiances, will only become possible operationally once these error characteristics can be better represented.

Assimilation of 'cloudy' radiances

Approximately 70% of the globe is covered by cloud [93], and therefore much of the satellite data available to NWP centres contains contributions from cloud. Because of the highly nonlinear relationship between satellite retrievals and cloud properties, direct use of cloudy radiances is often avoided in global NWP systems, and approximately 80% of satellite data is rejected because it is cloud contaminated. It is therefore unsurprising that the assimilation of cloudy radiances is of high priority at all NWP centres.

Attempts were previously made to assimilate ‘cloud-cleared’ radiances for AIRS data [57] but the assumptions of homogeneous cloud used in the technique were violated under most atmospheric conditions. Recent work in [70] addressed the feasibility of assimilating cloudy radiances directly. The proposed technique used simple retrieved cloud parameters from a 1D analysis to constrain the radiative transfer calculation in the assimilation process. The results using synthetic AIRS measurements demonstrated improvements in root-mean-square temperature and humidity errors for shallow layer cloud. However, results were less promising when the cases of thick or multi-layer cloud were considered.

A common conclusion from ‘cloudy’ radiance studies is that the physical parametrisation of clouds in radiative transfer modelling is vital to the successful assimilation of ‘cloudy’ radiances. Currently both the Met Office and the ECMWF assimilate some cloudy radiances using schemes similar to those described in [70] with limited cloud parameterisation [68]. It is hoped that a more aggressive use of high resolution infrared radiances to provide information on temperature structure near the cloud top will result in more accurate characterisation of the clouds. This will however lead to additional dependencies and complexities in the charcterisation of the observation errors.

Observation error characterisation

A common issue arises when addressing the utilisation of high resolution satellite observations: the specification of the observation error structure. Clearly the more processing steps involved in assimilating the observations, the more complicated the error characteristics become; for example in PCA. The correct specification of these errors is vital to the success of the assimilation. Studies have shown that increasing observation density beyond some threshold can result in little or no improvement in analysis accuracy [60], or even a degradation [21], when the correlated observation errors are treated as independent. With the new generation of multi-channel advanced sounders, treating observation errors incorrectly will only result in further sub-optimality.

Observation errors are present in all data and their nature is often dependent on the instrument used or the feature they observe. Contrasting model-observation resolutions and inaccurate physics in the radiative transfer equation mean satellite observation errors are highly likely to be correlated. However, in NWP centres observation error correlations are often set to zero regardless of the data type or the processing involved. This is deemed necessary for the computational demands of the data assimilation problem where a non-diagonal observation error covariance matrix can be very expensive to invert; but with the increasing use of high resolution data, the validity of this assumption has been severely challenged. At the Met Office, correlated observation errors are already included in radio occultation assimilation, and there are plans to implement correlated AMV errors at the ECMWF in 2010.

NWP centres have concluded that a better utilisation of satellite data will only be possible when observation error correlations are included in the data assimilation process. The aim of this thesis is to show that this inclusion is both feasible and beneficial. In Section 2.5 we will motivate the problem further by discussing the origin and role of observation error correlations.

2.5 Error covariances

We have seen that the specification of the error covariances for both the background and observations will determine their weighted importance in the final analysis. We now study more closely the origin and structure of the observation error covariances, and discuss their role in producing an accurate forecast.

The uncertainty associated with taking an observation sample is represented through an error vector $\epsilon^o \in \mathbb{R}^m$. The error vector is assumed to have Gaussian distribution with mean zero and error covariance matrix $R = \mathbb{E}[\epsilon^o(\epsilon^o)^T] \in \mathbb{R}^{m\times m}$. The Gaussian assumption does not hold in practice but the resultant pdfs make equation manipulation involving the errors algebraically simpler. The error covariance matrix R is comprised of the individual observation error variances on the diagonal and the error cross-covariances on the off-diagonals. A variance is defined as the mean-square deviation about the mean of the error data; a covariance is a measure of the association between two error variables [10]. In operational systems the error covariance matrix must be symmetric positive-definite.

Consider a simple 2×2 case. Suppose we have a direct measurement $y = (y_1\ y_2)^T$ of a variable $x = (x_1\ x_2)^T$, then the associated error ϵ and error covariance matrix R are given by $\epsilon = (\epsilon_1\ \epsilon_2)^T$ and

$$R = \begin{pmatrix} \mathbb{E}[\epsilon_1\epsilon_1] & \mathbb{E}[\epsilon_1\epsilon_2] \\ \mathbb{E}[\epsilon_2\epsilon_1] & \mathbb{E}[\epsilon_2\epsilon_2] \end{pmatrix} = \begin{pmatrix} \mathrm{var}(\epsilon_1) & \mathrm{cov}(\epsilon_1,\epsilon_2) \\ \mathrm{cov}(\epsilon_2,\epsilon_1) & \mathrm{var}(\epsilon_2) \end{pmatrix} \tag{2.24}$$

$$\equiv \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \tag{2.25}$$

where σ_1^2 and σ_2^2 are the error variances associated with measurement components y_1

and y_2, respectively, and σ_{12} is the error covariance of the two measurement components.

The observation errors can be classified as systematic or random, depending on whether they are constant between consecutive measurements, or vary randomly. Some systematic errors are removed from the observations prior to assimilation; for example, biases are typically removed in order to avoid a biased analysis [23]. The remaining random errors will determine the extent to which the background will be corrected to match the observations.

2.5.1 Origin of observation errors

The size and structure of observation errors will vary according to their origin. For example, a typical wind error in a radiosonde measurement is 2 m/s, while a satellite wind observation may have errors in the range $1 - 10$ m/s. Observation errors not only include errors explicitly generated in taking the observations, but also errors in the treatment of these observations. These two sources are sometimes considered in separate terms [20], but in this work we will contain them in a single error covariance matrix for algebraic convenience. Observation errors are distributed in the horizontal and vertical, and can generally be attributed to four main sources:

- **Instrument noise** - The error associated with an instrument reading under a set of test conditions will be provided by the instrument manufacturers. For example, satellite manufacturers provide the error associated with an instrument reading from a black body at $280K$; this is known as the neδt value. These errors can also be directly measured in space.

- **Forward model error** - This includes errors associated with the discretisation of the radiative transfer equation and errors in the mis-representation of gaseous contributors.

- **Representativity error** - This is present when the observations can resolve spatial scales or features that the model cannot. For example, a sharp temperature inversion in the vertical can be well-observed using radiosondes but cannot be represented precisely with the current vertical resolution of atmospheric models.
- **Pre-processing** - Any pre-processing the observations are subject to will generate errors. For example, if we eliminate all satellite observations affected by clouds and some residual cloud passed through the quality control, then one of the assimilation assumptions is violated and the cloudy observations will contaminate all satellite channels which are influenced by the cloud.

2.5.2 Observation error correlations

In order to represent accurately the observations in a data assimilation system we must be able to correctly determine both the diagonal error variances and the off-diagonal cross-covariances. In order to study the off-diagonal elements of R directly, it can help to transform the error covariances into error correlations using the formula:

$$\rho_{ij} \equiv \mathrm{corr}(\epsilon_i, \epsilon_j) = \frac{\mathrm{cov}(\epsilon_i, \epsilon_j)}{\sqrt{\mathrm{var}(\epsilon_i)\mathrm{var}(\epsilon_j)}}, \tag{2.26}$$

where ϵ_i and ϵ_j are the errors associated with an observation at point i and j in space, respectively. One can then decompose the observation error covariance matrix R into a diagonal variance matrix $D \in \mathbb{R}^{m\times m}$ and a correlation matrix $C \in \mathbb{R}^{m\times m}$: $R =$

$D^{1/2}CD^{1/2}$,

$$C = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1m} \\ \rho_{12} & 1 & \dots & \rho_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ \rho_{1m} & \rho_{2m} & \dots & 1 \end{pmatrix},$$

$$D = \begin{pmatrix} {\sigma_1}^2 & 0 & \dots & 0 \\ 0 & {\sigma_2}^2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & {\sigma_m}^2 \end{pmatrix}, \tag{2.27}$$

where σ_i^2 is the variance of the i^{th} error component, and ρ_{ij} is the correlation level between error components i and j [49].

2.5.3 Current issues in the treatment of observation error correlations

In the current operational assimilation systems at the Met Office and the ECMWF, almost all observation error correlations are assumed to be zero, i.e, the error correlation matrix C is the identity. This is a reasonable assumption for pairs of observations measured by distinct instruments, or for instrument noise from a regularly calibrated instrument. However under certain conditions this assumption is entirely inaccurate.

Observation error correlations can be vertically or horizontally distributed. If observations are used at a higher spatial frequency than the horizontal model resolution, then they will be affected by horizontal correlated errors of representativity because the model will be unable to represent the finer scale spatial structure given by the observations. Vertical errors of representativity will be present if the vertical model resolution is too low to represent a small scale physical feature as represented in the observation. For example, the forecast model may be unable to represent accurately a complex humidity structure at its current vertical resolution, leading to correlations in the errors between satellite channels sensitive to water vapour. Also observation preprocessing can generate artificial error correlations; for example, vertical correlations between satellite channels when PCA is used to compress data.

When observations are assimilated using their true error correlations, as opposed to the assumption of independent errors, the influence of the observations on the analysis is reduced. However, the inclusion of the observation error correlations has been shown to increase the accuracy of gradients of observed fields represented in the analysis [81]. They also act in conjunction with the background error covariance to specify how observation information should be smoothed. In Chapters 5 and 7 we will show how including correlated errors can increase the information content available from a data set, and reduce analysis error in an assimilation.

Unfortunately the magnitude and structure of these error correlations are relatively unknown, and since the number of observations is of order 10^6 [7], the storage and subsequent computational demands of using observation error correlations are deemed infeasible. Hence operationally, observations 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. In [14] Collard calculated the analysis retrieval error under different diagonal approximations to the estimated error covariance structure of AIRS data. Results showed an increase in the temperature and humidity error when diagonal approximations were used. The best approximation using a diagonal error covariance matrix was found to be when the errors were inflated between 2-4 times larger than the standard deviations of the true error covariance matrix. This is equivalent to multiplying the variance matrix D (2.27) by a constant. The variance enlargement was however constrained by the need for a physically accurate error estimate.

The assumption of zero correlations is often used in conjunction with data thinning methods such as superobbing [5]. This 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 [5], [60], [21]. Such studies, combined with earlier examples on how ignoring correlation structure hinders the use of satellite data (i.e, constraining channel selection algorithms [15]), suggest that error correlations for certain observation types have an important role to play in improving numerical weather forecasting.

Approximating observation error correlations in NWP is a relatively new direction of research but progress has been made. In [43] Healy and White 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 [34] 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. This method is potentially expensive if a large number of eigenpairs are needed. The work in this thesis extends on the existing body of work on modelling observation error correlation structure.

2.6 Diagnosing observation error correlations

In order to successfully model observation error correlations, we must have some understanding of the true error structure. This is not a straightforward problem because error covariances cannot be observed directly, only estimated in a statistical sense. Both the background, $y - h(x_b)$, and the analysis, $y - h(x_a)$, innovations are useful sources of information on the statistical properties of the errors, and can be used in several ways to provide a sound statistical basis for a refinement of the analysis system.

2.6.1 Hollingsworth-Lönnberg approach

The most commonly used estimation technique is the observational method, otherwise known as the Hollingsworth-Lönnberg method after the authors who popularised its use in meteorology [47]. This method uses background innovations statistics from a dense observing network, under the assumption that the background errors carry spatial correlations while the observation errors do not.

The premise is to calculate a histogram of background innovation covariances stratified against vertical or horizontal separation. The background innovation is given by

$$c = \mathbb{E}\left[(y - h(x_b))(y - h(x_b))^T\right] \tag{2.28}$$

where y is the observation vector, x_b is the background vector, and h is the observation operator. Under the assumption of mutually independent errors, equation (2.28) becomes

$$c = R + HBH^T \tag{2.29}$$

where H is the linearised observation operator. The i, j-th element of c represents the departure covariance between two points i and j in space.

At zero separation, i.e, when $i = j$, we have $c(i, i) = \sigma_o^2(i) + \sigma_b^2(i)$ where $\sigma_o^2(i)$ is the observation error variance at point i and $\sigma_b^2(i)$ is the background error variance in observation space at point i. At non-zero separation, $R(i, j) = 0$ and the departure covariance is given by the background covariance between points i and j. By calculating $c(i, j)$ for several pairs of spatial points we can create a histogram of the departure covariances scaled by the distance between the points (Figure 2.2). At zero separation we have information on the averaged background and observation errors; at non-zero separation we have information on the averaged background error covariances. By fitting an isotropic correlation model to the histogram and extrapolating this model to zero separation, we obtain a statistical estimate of the observation and background errors separately. This method was shown in the original paper to derive the covariance structure of wind background and observation errors, under the assumption of local homogeneity of the errors [47].

2.6.2 Desroziers' method of statistical approximation

With the increasing need for good error covariance specification, new methods have been proposed for diagnosing error correlations. Dee and da Silva [24] used a maximum likelihood method to estimate information on error statistics. Their work resulted in the derivation of statistical parameters that varied in time. Desroziers and Ivanov [26] used statistics of the analysis innovations to tune background and observation error

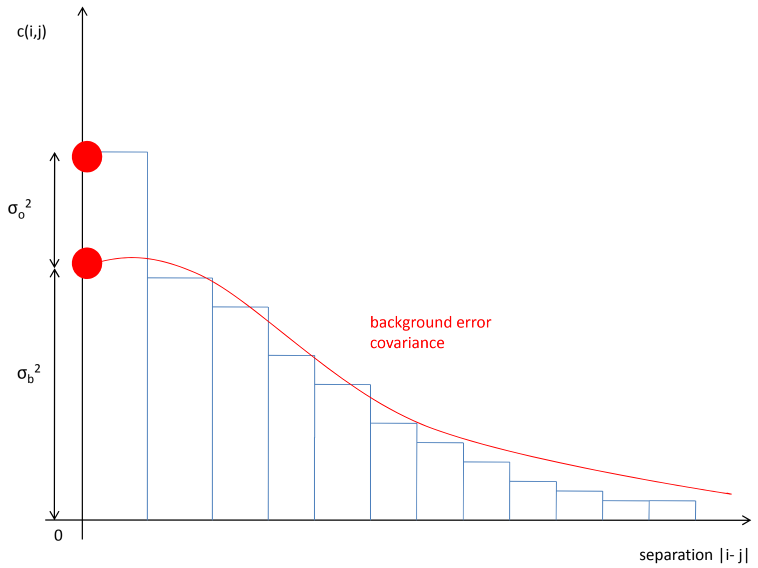


Figure 2.2: Diagrammatic representation of the Hollingsworth-Lönnberg method where $c(i, j)$ is the covariance between spatial points i and j. $c(i, j)$ provides information on the background and observation error variances $\sigma_o^2(i) + \sigma_b^2(i)$ at zero separation, and the background error covariances at non-zero separation. The red line is an isotropic correlation model fit to the histogram and extrapolated to zero separation. At zero separation the model provides information on $\sigma_o^2(i)$ and $\sigma_b^2(i)$ individually.

parameters, resulting in a successful description of the observation error parameters in a 3D-Var framework.

Although several error diagnosis methods have been demonstrated as successful and computationally feasible, the fundamental assumption that the observation errors are uncorrelated is incorrect. If we were to model observation errors as correlated, many of the diagnostics cannot be used. For example, in [24] both background and representativeness errors are likely to be spatially correlated, and therefore the statistical separation of error parameters becomes significantly tougher. A method that address the separation of correlated observation and background errors was proposed by Desroziers et al in 2005 [25]. The principle of the Desroziers' method is to use post-analysis diagnostics derived from linear estimation theory to statistically approximate the covariances of the observations, background and analysis errors in observation space. We describe the method below.

Equations (2.1) and (2.2) show how the background state x^b and the observation vector y are approximations to the true state of the atmosphere x^t. Assuming that the observation and background errors are uncorrelated and mutually independent (2.3), we can derive the BLUE equations (2.9) and (2.10) describing the optimal analysis state x^a. Following [25] we write an alternative expression for the analysis state (2.9) in terms of the background state x^b, the Kalman Gain matrix K (2.10), and the background innovation vector d_b^o,

$$x^a = x^b + Kd_b^o. \tag{2.30}$$

The background innovation vector d_b^o is the difference between the observations y and their background counterparts $h(x^b)$, and can also be described in terms of the observation and background errors,

$$\begin{aligned} d_b^o = y - h(x^b) &= y - h(x^t) + h(x^t) - h(x^b), \\ &\approx \epsilon^o + H(x^t - x^b), \\ &\approx \epsilon^o + H\epsilon^b \end{aligned} \tag{2.31}$$

where H is the linearised verison of h.

Similarly the analysis innovation vector d_a^o is given by the differences between the observations and their analysis counterparts $h(x^a)$,

$$\begin{aligned} d_a^o = y - h(x^a) &= y - h(x^b + Kd_b^o), \\ &\approx y - h(x_b) - HKd_b^o, \\ &\approx (I - HK)d_b^o, \\ &\approx R(HBH^T + R)^{-1}d_b^o. \end{aligned} \tag{2.32}$$

By taking the expectation of the cross product of (2.31) and (2.32), and using the assumption of mutually uncorrelated observation and background errors (2.3), we find a statistical approximation of the observation error covariances,

$$\begin{aligned}
\mathbb{E}\left[d_a^o(d_b^o)^T\right] &= \mathbb{E}\left[R(HBH^T+R)^{-1}d_b^o(d_b^o)^T\right] \\
&\approx R(HBH^T+R)^{-1}\mathbb{E}\left[(\epsilon^o+H\epsilon^b)(\epsilon^o+H\epsilon^b)^T\right] \\
&\approx R(HBH^T+R)^{-1}\left(\mathbb{E}\left[\epsilon^o(\epsilon^o)^T\right]+H\mathbb{E}\left[\epsilon^b(\epsilon^b)^T\right]H^T\right) \\
&\approx R(HBH^T+R)^{-1}(HBH^T+R) \\
&\approx R. \qquad (2.33)
\end{aligned}$$

The relation (2.33) should be satisfied provided the covariance matrices used in $R(HBH^T+R)^{-1}$ are consistent with the true observation and background error covariances $\mathbb{E}\left[\epsilon^o(\epsilon^o)^T\right]$ and $\mathbb{E}\left[\epsilon^b(\epsilon^b)^T\right]$. This diagnostic can be used as a consistency check to ensure the observation error covariances are correctly specified in the analysis. Similar diagnostics can be generated to check the background error covariances in observation space, HBH^T, the analysis errors covariances HS_aH^T, and the sum of the observation and background error covariances, $R+HBH^T$ [25].

In [25] the diagnostics were applied to analyses from the French operational ARPEGE 4D-Var data assimilation system. The results showed that background and observation errors were being overestimated in the analysis. Also by applying the diagnostic (2.33) in a toy problem, Desroziers et al showed that most of the information on observation error covariances can be recovered when they are initially mis-specified. Such results are encouraging because the diagnostic by its construction is nearly cost-free, and it allows the distinction between observation and background correlation structure. However, the relation (2.33) only holds exactly when the errors assumed in the assimilation are equal to those found in reality, i.e, $\mathbb{E}\left[\epsilon^o(\epsilon^o)^T\right] = R$, and the observation operator is linear. Care must therefore be taken when interpreting the results using these diagnostics.

2.7 Summary

In this thesis we are concerned with quantifying and modelling observation error correlation structure in NWP. In order to address the three thesis questions posed in Chapter 1 we must provide the theoretical background for the experiments described in later chapters. In this chapter we introduced the concepts of data assimilation and remote sensing. The role and impact of correlated observation errors in relation to these fields are studied in detail throughout the thesis.

In the first three sections of this chapter we described different data assimilation techniques and their role in the evolution of NWP. We focused on the variational methods of 3D-Var and 4D-Var, and described in detail the resultant cost functions and their practical implementation. These techniques will be used in the experiments performed in later thesis chapters.

In Section 2.4 we considered the procurement and properties of remotely sensed satellite data, namely satellite radiance observations. We described the nature of these observations and how their relationship to atmospheric variables through the radiative transfer equation leads to an ill-posed inverse retrieval problem. Since satellite radiance data accounts for approximately 90% of all data used in the ECMWF assimilation algorithms, optimising its usage is a large area of operational research. We concluded the section by highlighting the current issues in satellite remote sensing including observation error characterisation.

In the penultimate section of the chapter we focused on observation error covariances. These are often ignored in operational data assimilation algorithms, but evidence and intuition suggests that their inclusion will improve the use of satellite data. This will be further investigated in Chapters 5 and 7. Here we described the origin and structure of observation error covariances, and discussed the impact of treating observation errors as independent. We reviewed the current proposed methods of incorporating error correlation structure in data assimilation algorithms; these methods will be further discussed in Chapter 3.

Finally we discussed the different techniques available to quantify error covariance structure. We described the Hollingsworth-Lönnberg method which assumes independent observation errors, and a new method proposed by Desroziers et al [25] in which observation error correlations can be independently derived. The Desroziers' method of statistical approximation will be used later in Chapter 4 to quantify observation error correlation structure for satellite instrument data.

Chapter 3

Matrix representation and retrieval properties

In Chapter 2 we described the structure and properties of the observation error covariance matrix, and commented on its current treatment in operational data assimilation algorithms. In this chapter we will carefully examine the different approximating structures that can be used to represent covariance matrices, and the retrieval properties used to measure the success of the approximation. The theory in this chapter addresses the second thesis aim posed in Chapter 1: what approximations are available to model error correlation structure and what is their impact on data assimilation diagnostics?

3.1 Diagonal approximations

One of the benefits of satellite observations is the vast quantity of data available; however, this in turn is an obstacle to its assimilation. If the observation vector is of size

m then the observation error covariance matrix contains m^2 elements, but by symmetry this is reduced to $(m^2 + m)/2$ 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. However, when the observation errors are correlated, we may have to represent, and subsequently use, the maximum number of elements in the observation error covariance matrix.

From equation (2.8) and (2.12), we know that the inverse of the observation error covariance matrix is the form needed for the calculation of the cost function and its gradient. When the observation error covariance matrix is diagonal, its inverse will also be diagonal. However a non-diagonal matrix, even if sparse, may have a dense inverse. This inverse is required for $2N$ matrix-vector calculations in the cost function and gradient evaluations, where N is the number of assimilation timesteps. A dense inverse may therefore result in excessive additional cost in running a data assimilation algorithm. In operational NWP, this problem is avoided by treating the observation errors as uncorrelated and using diagonal approximations to the true error covariance matrix.

The simplest diagonal approximation of an error covariance matrix is a diagonal of the true variances, or D in equation (2.27). However, by ignoring entirely the correlated component of the observation error, the observations will be overweighted in the analysis because they will appear more or less 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 [46]. This reduces the weighting of the observations in the analysis. The diagonal approximation is now in the form

$$\hat{D} = \begin{pmatrix} d_1\sigma_1^2 & 0 & \dots & 0 \\ 0 & d_2\sigma_2^2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & \dots & 0 & d_m\sigma_m^2 \end{pmatrix}, \tag{3.1}$$

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

The diagonal inflation factors are empirically derived from test data sets; we have no mathematical reasoning to assume that they are truly optimal. However, work in financial mathematics on approximations to a correlation matrix may provide us with the techniques to quantify the optimality of our approximations [44], [87]. Further discussion of this is given in Chapter 8.

In [14], Collard examined the impact of different diagonal observation error covariance approximations on the assimilation of AIRS data. Using three different estimates of the true standard deviation, results showed that diagonal inflation is constrained, between 2-4 times, by the need for a physically accurate error estimate. Collard also concluded that the full potential of the observations, especially with regards to resolving fine scale vertical structure, could not be realised under the assumption of uncorrelated error. Such results suggest an alternative approach to dealing with observation error correlations is needed.

3.2 Circulant approximations

One possible approach to representing the observation error covariance matrix in a more realistic and operationally useable form is described in [43]. In [43], the authors propose that a near symmetric Toeplitz observation error covariance matrix can be well approximated by a circulant matrix. The spectral properties of the circulant matrix allow for ease of use in operational 1D-Var algorithms. Below we describe the form and properties of Toeplitz and circulant matrices, and demonstrate how the approximation is formed.

3.2.1 Toeplitz matrices

Toeplitz matrices are a class of persymmetric matrices, i.e, they are symmetric about their northeast-southwest diagonal, and can be written in the form

$$T_m = \begin{pmatrix} t_0 & t_{-1} & t_{-2} & \dots & t_{-(m-1)} \\ t_1 & t_0 & t_{-1} & \dots & t_{-(m-2)} \\ t_2 & t_1 & t_0 & \dots & \dots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{m-1} & t_{m-2} & \dots & t_1 & t_0 \end{pmatrix}, \tag{3.2}$$

where $T_m = [t_{k,j}; k, j = 0, 1, \dots, m-1]$ and $t_{k,j} = t_{k-j}$ [41]. A standard Toeplitz matrix has $2m - 1$ independent entries; a symmetric Toeplitz matrix has only m independent entries. A simpler class of banded Toeplitz matrices can be defined by the constraint that $t_k = 0, |k| > M$ for some finite M [41].

Toeplitz matrices are dense matrices with a special structure, and arise in many mathematical applications such as signal and image processing. The Toeplitz structure can be exploited, and specific iterative procedures are available to solve the resultant Toeplitz systems. In this work we are not concerned with solving general Toeplitz systems explicitly, but a detailed discussion of the iterative techniques available is given in [69].

3.2.2 Circulant matrices

A circulant matrix is a Toeplitz matrix where each column is a circular shift of its preceding column. A circulant matrix C can be written in the form

$$C = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & c_{m-1} \\ c_{m-1} & c_0 & c_1 & \cdots & c_{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_2 & c_3 & \cdots & c_0 & c_1 \\ c_1 & c_2 & \cdots & c_{m-1} & c_0 \end{pmatrix}, \tag{3.3}$$

where each row is a cyclic shift of the row immediately above it [41]. The inherent properties of circulant matrices make them particularly useful in matrix representation. These can be summarised as:

(i) All circulant matrices have the same eigenvectors, given by

$$y^{(k)} = \frac{1}{\sqrt{m}} \left(1, e^{-\frac{2\pi ik}{m}}, \ldots, e^{-\frac{2\pi ik(m-1)}{m}}\right), \quad k = 0, \ldots, m-1.$$

These are equivalent to the columns of a discrete Fourier transform (DFT) matrix of the form

$$F = \frac{1}{\sqrt{m}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-\frac{2\pi i}{m}} & e^{-\frac{2\pi i}{m}\times 2} & \dots & e^{-\frac{2\pi i}{m}\times(m-1)} \\ 1 & e^{-\frac{2\pi i}{m}\times 2} & e^{-\frac{2\pi i}{m}\times 4} & \dots & e^{-\frac{2\pi i}{m}\times 2(m-1)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 1 & e^{-\frac{2\pi i}{m}\times(m-1)} & e^{-\frac{2\pi i}{m}\times 2(m-1)} & \dots & e^{-\frac{2\pi i}{m}\times(m-1)^2} \end{pmatrix}. \qquad (3.4)$$

(ii) The eigenvalues of a circulant matrix C, given by $\lambda_k = \sum_{j=0}^{m-1} c_k e^{-\frac{2\pi ikj}{m}}$ where j sums over the rows of C, can be obtained by applying a DFT to the first row of C.

(iii) The inverse, product and sums of circulant matrices are also circulant.

If we consider the eigendecomposition of a circulant matrix

$$C = F^{-1}\Lambda F,$$

where Λ is the diagonal matrix of the eigenvalues of C and F is the matrix of eigenvectors, then we can see how its inverse can be similarly decomposed:

$$C^{-1} = F^{-1}\Lambda^{-1}F.$$

From the properties listed above, we see that multiplying a vector by C or C^{-1} is simply equivalent to applying three DFTs and one vector-vector product: one DFT to the first row of C to calculate the eigenvalues, one DFT to represent the matrix of eigenvectors F, and one inverse DFT to represent the inverse of F. The only storage requirements for this process is the first row of C, and the use of Fast Fourier transforms will make the computation of matrix-vector products very fast.

3.2.3 Toeplitz-circulant approximations

A circulant approximation C to a symmetric Toeplitz matrix T can be described by only its first row and contains fewer individual elements than the original Toeplitz form. The first row of C is found by reflecting the first row of T with the reflection axis between the columns $\left[\frac{m}{2}\right]_+$ and $\left[\frac{m}{2}+1\right]_+$ where $\left[\frac{m}{2}\right]_+$ is the smallest integer value greater than $\frac{m}{2}$ [43]. For example if $m = 5$ and we have a Toeplitz matrix of the form,

$$T = \begin{pmatrix} x & y & z & s & t \\ y & x & y & z & s \\ z & y & x & y & z \\ s & z & y & x & y \\ t & s & z & y & x \end{pmatrix},$$

then the first row of C is the reflection of the first row of T between the elements z and s, i.e. $(x\ y\ z\ z\ y)$. The remaining rows are found by performing a cyclic shift to the right of the previous row to give the approximating circulant matrix

$$C = \begin{pmatrix} x & y & z & z & y \\ y & x & y & z & z \\ z & y & x & y & z \\ z & z & y & x & y \\ y & z & z & y & x \end{pmatrix}.$$

From this example case, we can see that the disagreement between the two matrices is confined to the off-diagonal corners. In reality these corners of an error correlation martrix represent the correlations between horizontally or vertically spatially distant errors, which are in general, likely to be smaller than those close together. Therefore the circulant matrix approximation may contain spurious long-range correlations, since small values in the corners of Toeplitz matrix are replaced with moderately large ones.

In [41], the approximation of a Toeplitz matrix by its circulant equivalent is formalised. It is shown that as the size of the matrix $m \to \infty$, the difference between T and C converges in the Frobenius norm, and C^{-1} becomes a good approximation to T^{-1}, i.e, $C^{-1}T \cong I$.

In some meteorological cases, such as for apodised 1D-Var IASI radiance measurements, the observation error correlation matrix may be close to a symmetric Toeplitz form [43]. In image processing problems, approximating a Toeplitz matrix by its circulant equivalent is widely used [16], and the theory in [43] extends this idea to 1D-Var retrievals of high resolution satellite measurements. It is demonstrated that correlation matrices with a symmetric Toeplitz structure can be approximated with circulant matrices, and the manipulation of such matrices is not overly complicated. In Chapter 5 we will perform further assimilation experiments using circulant matrix structures

3.2.4 Markov special case

The Toeplitz-circulant approximation discussed above is potentially useful because it involves less matrix storage and potentially fewer matrix operations than using a Toeplitz matrix explicitly. However, there also exists a Toeplitz matrix whose properties allow for a simple matrix inversion needed in the calculation of the data assimilation cost function. Known as the Markov matrix, this matrix is the resultant covariance matrix from a first-order autoregression process [92].

A series of data in space, such as temperature measurements T at different levels in the atmosphere, can be written in the form,

$$T_{k+1} - \mu = \rho(T_k - \mu) + \epsilon_{k+1} \tag{3.5}$$

where k is the atmospheric level, μ is the mean of the spatial series, ρ is the autoregressive parameter, and ϵ_{k+1} is the residual error associated with the regression [92]. Using ideas from time series analysis applied to spatial data, we can describe equation (3.5) as a first-order autoregressive process or an AR(1) model. This is the continuous analog of a first-order Markov chain, i.e, the data can take on infinitely many values on a real line. The Markov property of the process states that the probability of a future state is only dependent on the probability of the present state and is independent of the probability of any previous states. This does not mean series values separated by more than one step are independent, rather that the information on the future state is contained entirely in the present state.

By treating the values of ϵ as mutually independent, uncorrelated with the value of T, and Gaussian distributed with mean zero and variance σ_ϵ^2, the covariance matrix of the AR(1) process (3.5) can be derived to be

$$R(i, j) = \sigma_t^2 \rho^{|i-j|}, \tag{3.6}$$

where σ_t^2 is the variance of the time series [78].

In [78] an AR(1) process is used to model a vertical column of temperature departures from the mean. Here, the AR(1) covariance matrix, or Markov matrix, is written as

$$R(i, j) = \sigma_t^2 \exp\left\{\frac{-|i-j|\delta z}{h}\right\} \tag{3.7}$$

where δz is the distance between vertical levels and $h = -\frac{\delta z}{\ln \rho}$ is the length scale chosen so that the interlevel correlation is $\frac{1}{e}$. To show that equation (3.7) is of the form (3.6), we write the correlation matrix associated with (3.7) as

$$C = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{m-1} \\ \rho & 1 & \rho & \dots & \rho^{m-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{m-2} & \dots & \rho & 1 & \rho \\ \rho^{m-1} & \dots & \rho^2 & \rho & 1 \end{pmatrix} \tag{3.8}$$

where $\rho = \exp\left\{-\frac{\delta z}{h}\right\}$. This matrix has a tri-diagonal inverse,

$$C^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \tag{3.9}$$

In current data assimilation algorithms, the inverse of the observation error correlation matrix is required for the calculation of the cost function and its gradient. In order for this to be operationally feasible, the storage requirements and number of matrix product operations of the inverse matrix must be sufficiently small. The storage needed for reconstructing matrix (3.9) 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 matrix approximation (3.8) is a realistic possibility. In Chapters 6 and 7 we will model error correlation structure using Markov matrix approximations and evaluate their success.

3.3 Eigendecomposition approximation

Following [34] we assume that the observation error covariance matrix has a block-diagonal structure with blocks corresponding to different instruments, or groups of channels. It is unlikely there will be significant correlation between blocks, and certain blocks may even be diagonal because the observation errors are uncorrelated. For those instruments or channels whose observation errors are likely to be correlated, we can use a correlated approximation such as those described in Sections 3.1 and 3.2. However, these approximations do not attempt to incorporate any prior knowledge of the error correlations. A correlated matrix approximation which attempts to utilise a potentially known error correlation structure was proposed in [34].

Recall the matrix decomposition $R = D^{1/2}CD^{1/2}$ from Section 2.5.2. In [34] the observation error covariance matrix is approximated using a truncated eigendecomposition $\hat{C}$ of the error correlation matrix C,

$$R = D^{1/2}(\alpha I + \sum_{k=1}^{K} \left(\lambda_k - \alpha)v_k v_k^T\right) D^{1/2} \equiv D^{1/2}\hat{C}D^{1/2}, \tag{3.10}$$

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 (3.10) is easily obtainable and is given by

$$R^{-1} = D^{-1/2}(\alpha^{-1} I + \sum_{k=1}^{K} \left(\lambda_k^{-1} - \alpha^{-1})v_k v_k^T\right) D^{-1/2} = D^{-1/2}\hat{C}^{-1}D^{-1/2}. \tag{3.11}$$

The representation (3.10) allows the retention of some of the true correlation structure, with the user choosing how accurately to represent the inverse error covariance matrix (3.11) through the choice of K.

In [34] the leading eigenpairs of C are found using the Lanczos algorithm. However, if the correlation matrix is available explicitly, then the eigenspectrum can be calculated directly using a suitable algorithm. The method was demonstrated successfully in [34] for observation errors with Gaussian correlation structure and unit variance. However, spurious long-range correlations were present when too few eigenpairs were used in the approximation. In Chapters 5 and 7 we will apply this method to different realisations of observation error correlation structure.

3.4 Summary of matrix representations

The approximations described in Sections 3.1 to 3.3 have all been proposed for modelling observation error correlation structure in data assimilation algorithms. We have reviewed both diagonal and correlated approximations. We described the properties of three different correlated matrix approximations and discussed their potential benefit to reducing the expense of the cost function calculations needed in 3D-Var and 4D-Var. In Chapters 5 and 7 we will use these matrix representations to model different realisations of error correlation structure.

The success of a data assimilation algorithm can be described by several measures. By using the same observations and model framework, and varying the modelled observation error correlation structure, any effect on the value of the measure can be attributed to the observation error correlation approximation used. In the second half of this chapter we will describe several popular metrics used in data assimilation studies.

3.5 Analysis error covariance matrices

An obvious measure of how useful an observation set is to a data assimilation algorithm is the error reduction in the state variable, i.e, the analysis error covariance matrix. The smaller the trace of the matrix S_a, the better the reduction in error variance. Recall from Section 2.2, under the assumptions of mutually independent background and observation errors, and the linearity of H, the analysis error covariance matrix is derived as in [49] to be

$$S_a = (H^T R^{-1} H + B^{-1})^{-1}; \tag{3.12}$$

substituting in the Kalman gain matrix K we obtain

$$S_a = (I - KH)B. \tag{3.13}$$

In 3D-Var data assimilation, S_a can be calculated explicitly by (3.13). In 4D-Var the analysis error covariance matrix is often inferred from the matrix of second derivatives of the cost function known as the Hessian [8]. It can also be calculated sequentially [77] but this relies on the original observation error covariance matrix being diagonal.

Equation (3.13) applies when all the statistical assumptions on the observation and background errors are accurate. However, when we use the incorrect observation error covariance matrix it is not appropriate to use (3.13). If we knowingly using an incorrect error covariance matrix in the analysis error calculations, then we need to take account of the true observation errors in order to calculate the analysis error covariance matrix correctly. This can be achieved through the addition of an extra term to the analysis error covariance matrix, as in [78], giving

$$
\begin{aligned}
S_a^* &= S_a + KR'K^T && (3.14)\\
K &= BH^T(HBH^T + R_f)^{-1} && (3.15)\\
R' &= R_t - R_f && (3.16)
\end{aligned}
$$

where S_a^* is the correct analysis error covariance matrix, R_t and R_f are the true and false observation error covariance matrices, respectively, and S_a and K are both evaluated at R_f.

However, we can argue that we always knowingly use an incorrect R matrix because it is impossible to know the observation errors exactly. By treating R_f as an upper bound for the true error covariance matrix, we can use the same error analysis by using R_f in equation (3.12). In conclusion, when comparing the analysis error covariance matrix for different structures of R, three approaches are possible:

A1 Assume we are using the correct R matrix, R_t, and evaluate S_a at this value

$$S_a = (H^T R_t^{-1} H + B^{-1})^{-1} \tag{3.17}$$

A2 We knowingly use an incorrect R matrix and include an additional term in the analysis error covariance matrix to accurately model this

$$S_a^* = (H^T R_f^{-1} H + B^{-1})^{-1} + KR'K^T \tag{3.18}$$

A3 Accept that we are using an incorrect R matrix but don't know the truth, and evaluate S_a at this value

$$S_a = (H^T R_f^{-1} H + B^{-1})^{-1} \tag{3.19}$$

Approaches **A1** and **A2** are used in [14] to determine the impact of various diagonal approximations to an empirically determined observation error correlation structure. In [43] **A1** and **A2** are also used to calculate the trace of the analysis error covariance matrix for Toeplitz, circulant and diagonal error covariance matrix structures.

3.6 Information content

When we ignore observation error correlations and use processes such as data thinning and superobbing we are neglecting a portion of the data, and information that could be utilised is lost. Quantifying the information provided by an observation or an observing system is used in the development of satellite instruments [74], [89] as well as in the assessment of operational data assimilation systems [33]. The two most frequently used information measures are entropy reduction (or Shannon Information Content) and degrees of freedom for signal. These measures are generally applicable and can be defined without reference to a specific retrieval method [77].

3.6.1 Entropy reduction

Entropy is a real valued functional that characterises pdfs [82]. Recall from Section 2.2, that pdfs can be used as a measure of knowledge of the state and observation vectors. If $P_b(x)$ represents the knowledge of the state vector before the observations and $P_o(x|y)$ represents the knowledge after the observations are taken, then their respective entropies are defined to be

$$\begin{aligned} S[P_b(x)] &= -\int P_b(x)\mathrm{log}_2[P_b(x)]dx, & (3.20)\\ S[P_o(x|y)] &= -\int P_o(x|y)\mathrm{log}_2[P_o(x|y)]dx. & (3.21)\end{aligned}$$

The Shannon Information Content (SIC), or entropy reduction, due to the use of the observations is then given by

$$SIC = S[P_b(x)] - S[P_o(x|y)]. \tag{3.22}$$

Under the assumption of Gaussian pdfs, it is algebraically convenient to use natural logs as opposed to $\log_2$ [78]. This results in a rescaling of the entropy definition by $\ln 2 = 0.69$, but makes equation manipulation considerably easier. Using this approach, we can rewrite equations (3.20) and (3.21) as

$$\begin{aligned} S[P_b(x)] &= n\ln(2\pi e)^{1/2} + \frac{1}{2}\ln|B| & (3.23) \\ S[P_o(x|y)] &= n\ln(2\pi e)^{1/2} + \frac{1}{2}\ln|S_a| & (3.24) \end{aligned}$$

where $|B|$ and $|S_a|$ are the determinants of the matrices B and S_a, respectively [78]. Combining equations (3.22), (3.23) and (3.24) we can write the SIC due to the observations as

$$SIC = \frac{1}{2}\ln\frac{|B|}{|S_a|}. \tag{3.25}$$

3.6.2 Degrees of freedom for signal

The degrees of freedom in a set of observations is a measure of the amount of information from the data that has been utilised. The number of degrees of freedom for signal (dof_S) indicates the number of independent quantities deemed measured by the observations; the remaining degrees of freedom, known as the degrees of freedom for noise, provide information about the uncertainty in the data [78]. The closer the dof_S is to the total number of degrees of freedom, the more information the observations have provided.

We have an initial covariance matrix B and performing an analysis to minimise the variance in observed directions gives us a posterior covariance matrix S_a. The size of the eigenvalues in each matrix represents the size of the uncertainty in the direction of the associated eigenvector; by comparing the eigenvalues of the two, we can determine the reduction in uncertainty.

To this end, we take a non-singular square matrix L, as in [33], such that $LBL^T = I$ and $LS_aL^T = \hat{S}_a$, where B and S_a are both symmetric positive definite. This transformation is not unique as we can replace L by X^TL where X is an orthogonal matrix. Now if we take X to be the matrix of eigenvectors of $\hat{S}_a$, then we simultaneously reduce B to the identity matrix and $\hat{S}_a$ to a diagonal matrix of its eigenvalues, Λ;

$$\begin{aligned} X^TLBL^TX &= X^TX = I, \\ X^TLS_aL^TX &= X^T\hat{S}_aX = \Lambda. \end{aligned}$$

After this transformation, the diagonal elements (eigenvalues) of the transformed matrix LBL^T are unity and each corresponds to an individual degree of freedom. The eigenvalues of $\hat{S}_a$ may therefore be interpreted as the relative reduction of variance in each of the independent directions. Hence if n is the total number of degrees of freedom (also the number of components of the state vector), then the dof_S is given by

$$dof_S = n - \text{trace}(\Lambda). \tag{3.26}$$

By the properties of a matrix trace we can write (3.26) as

$$\begin{aligned} dof_S &= n - \text{trace}(\hat{S}_a), \\ &= n - \text{trace}(B^{-1}S_a). \end{aligned} \tag{3.27}$$

Note that equations (3.25) and (3.27) describe the SIC and dof_S in terms of the scaled analysis variances; therefore, information inferred from one measure can be directly related to the other. Also, we must be careful to use a consistent definition for the analysis error covariance matrix as described by equations (3.17) - (3.19). This issue is addressed in the information content experiments performed in Chapter 5.

Information content studies have been performed for some of the structures described in Sections 3.1 and 3.2. In [14] the number of dof_S was calculated for different diagonal approximations to a non-diagonal error covariance matrix. In [78] the SIC and number of dof_S were calculated in a simulated study using a Markov matrix as the true observation error covariance matrix. In this work both information measures were found to be significantly larger when the full error covariance matrix was used in preference to a diagonal approximation of the same variances. In [85] approaches **A1**, **A2** and **A3**, described in Section 3.5, were used to evaluate information content under different diagonal and eigendecompostion approximations to a SOAR distributed error correlation matrix [3]. An eigendecomposition approximation with a sufficient number of eigenpairs was shown to retain the most information relative to the truth. Further information content studies are performed in Chapter 5.

3.7 Norms

The final quality retrieval measures we consider in this chapter are particular vector and matrix norms. These can be used to evaluate assimilation accuracy and compare covariance matrix approximations, respectively.

3.7.1 Vector norms

A vector norm is a measure of distance in vector space [36]. The norm $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the following properties for vectors $x, y \in \mathbb{R}^n$ and real number α:

- $f(x) \geq 0$ with equality if and only if $x = 0$;
- $f(x + y) \leq f(x) + f(y)$;
- $f(\alpha x) = |\alpha| f(x)$.

A commonly used norm for measuring vectors is the 2-norm:

$$\|x\|_2 = (|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2)^{1/2} = (x^T x)^{1/2}. \tag{3.28}$$

However, the 2-norm is not used explicitly as a retrieval measure; it is directly related to the root mean square error which is commonly used as a diagnostic [5], [70], [54]. Assuming the data is unbiased, the root mean square error (rms) is given by

$$\begin{aligned} rms &= \left(\frac{1}{n}\left(|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2\right)\right)^{1/2}, \qquad (3.29) \\ &= \left(\frac{1}{n} x^T x\right)^{1/2}, \\ &= \frac{1}{\sqrt{n}} \|x\|_2 . \end{aligned}$$

3.7.2 Matrix norms

Although not used explicitly in assessing data assimilation algorithms, matrix norms are a useful measure of how accurately an error covariance matrix is approximated. Matrix norms act as a distance measure on a space of matrices [36]. A matrix norm $f : \mathbb{R}^{m \times n} \to \mathbb{R}$ holds the following properties for matrices $A, B \in \mathbb{R}^{m \times n}$ and real number α:

- $f(A) \geq 0$ with equality if and only if $A = 0$;
- $f(A + B) \leq f(A) + f(B)$;

- $f(\alpha A) = |\alpha| f(A)$.

The matrix norm we will use is the Frobenius norm (sometimes called the Euclidean matrix norm), which is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}, \tag{3.30}$$

where a_{ij} are the elements of the matrix A. If A is a symmetric positive-definite matrix, such as an error covariance matrix, then the Frobenius norm can be described in terms of the eigenvalues of A,

$$\begin{aligned}
\|A\|_F &= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}, && (3.31)\\
&= \left(\mathrm{tr}(A^T A)\right)^{1/2}, && (3.32)\\
&= \left(\sum_{k=1}^{n} \lambda_k^2\right)^{1/2}, && (3.33)
\end{aligned}$$

where λ_k is an eigenvalue of A.

For the purpose of this work, we are interested in the difference between an observation error covariance matrix R_t and its approximation R_f. The Frobenius norm of the difference is given by

$$\begin{aligned}
\|R_t - R_f\|_F &= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |r_{ij} - \hat{r}_{ij}|^2} \\
&= \left(\mathrm{tr}\left((R_t - R_f)^T (R_t - R_f)\right)\right)^{1/2} && (3.34)\\
&= \left(\sum_{k=1}^{n} \mu_k^2\right)^{1/2} && (3.35)
\end{aligned}$$

where r_{ij} and $\hat{r}_{ij}$ are elements of matrices R_t and R_f, respectively, and μ_k is an eigenvalue of $R_t - R_f$. It is also possible to calculate the Frobenius norm of the difference between the respective analysis error covariance matrices using R_t and R_f from the formulae (3.14)-(3.16):

$$\|S_a^* - S_a\|_F = \left\|K(R_t - R_f)K^T\right\|_F. \tag{3.36}$$

where S_a^* and S_a are the analysis error covariance matrices of R_f and R_t, respectively.

3.8 Summary

In this chapter we addressed one of the main thesis questions posed in Chapter 1: what approximations are available to model error correlation structure and what is their impact on data assimilation diagnostics? In Sections 3.1 to 3.3 three different types of approximating structures were described: diagonal, circulant and eigendecomposition approximations. We focused on the unique properties of the approximating matrices that make them potentially useful in variational data assimilation algorithms. Clearly diagonal approximations can be used very cheaply in matrix-vector products, but the tri-diagonal inverse structure of a Markov matrix and the DFT representation of a circulant matrix demonstrate cheap ways of including some correlation structure in the error covariance matrix. The representation of the inverse matrix of all approximations avoids the need for a potentially expensive explicit inversion of R.

To determine whether an approximation is suitable, it is necessary to quantify its impact when used in a data assimilation scheme. In Sections 3.5 to 3.7 we described three popular retrieval measures used in the assessment of data assimilation algorithms: analysis error covariance matrices, information content and norms. These are not the only available measures, but in contrast to diagnostics like the effect on forecast accuracy, are simple to quantify in the subsequent framework. When calculating these measures, attention must be paid to the assumptions under which the observation error correlations are specified, i.e, do we assume a correct or an incorrect R matrix has been used? Incorrect assumptions can lead to misleading estimates of analysis accuracy and information content. If we have an accurate specification of the true error correlation structure, then this problem is mitigated because we are more certain of the true specification of C (and hence R). In the next chapter we will demonstrate how an accurate specification of C can be determined.

Chapter 4

Quantifying observation error correlations

In Chapter 2 we described the variational formulation of operational data assimilation algorithms, where the information provided by the observations and a first-guess model background is weighted by the inverse of their respective error covariance matrices. We discussed how the error characteristics of remotely sensed observation types are typically not accurately represented in data assimilation algorithms. This can result in a negative impact on forecast accuracy and an inefficient utilisation of observations [14], [43], [60]. Quantifying observation error correlations is not a straightforward problem because they can only be estimated in a statistical sense, not observed directly. However, attempts have been made to quantify error correlation structure for different observation types such as Atmospheric Motion Vectors [7] and satellite radiances [83].

Two methods of diagnosing the correct error covariance matrices using post-analysis diagnostics are the Hollingsworth-Lönnberg method [47] and the Desroziers' method [25] described in Section 2.6. The Hollingsworth-Lönnberg method is typically used when the background errors carry spatial correlations while the observation errors do not, while the Desroziers' method can be applied when observation error correlations are present. The results in [25] demonstrated the success of the Desroziers' method in diagnosing cross-correlations from data which is assumed uncorrelated in the assimilation (i.e, the observation error covariance matrix is set as diagonal) but in reality carries correlations. In this chapter we extend these studies to operational data.

The work in this chapter aims to answer the first thesis question posed in Chapter 1: what is the true structure of observation error correlations? Using Infrared Atmospheric Sounding Interferometer (IASI) data as the observation type, we will apply the diagnostic proposed by Desroziers (2.33) to quantify the cross-channel correlations between measurements. First we will introduce the IASI instrument and describe the likely origin of IASI observation error correlations. We then apply the diagnostic to IASI measurements processed using the Met Office incremental 4D-Var data assimilation scheme. Technical and practical details of the process will be given, including the pre-processing of IASI observations through 1D-Var retrievals. The diagnosed error covariances will be produced for both the 1D-Var retrieval procedure and the 4D-Var assimilation. The results described have previously been published as a technical report [86].

4.1 Infrared Atmospheric Sounding Interferometer (IASI) data

The IASI instrument is an infrared Fourier transform spectrometer which measures the infrared radiation emitted by the earth's surface and atmosphere [9]. The first IASI instrument was launched on the MetOp-A satellite in 2006 as part of the EUMETSAT European Polar System (EPS). Its spectral interval of 645-2760cm^{-1} is divided into three bands and sampled by 8461 channels at a resolution of 0.5cm^{-1}. Band one, from 645-1210cm^{-1}, is used primarily for temperature and ozone sounding, band two (1210-2000cm^{-1}) for water vapour sounding and the retrieval of N_2O and CH_4 column amounts, and band three (2000-2760cm^{-1}) for temperature sounding and the retrieval of N_2O and CO column amounts.

IASI measurements of radiances, r, are expressed as black-body equivalent brightness temperatures, T, through Planck's function [59]

$$r = \frac{2h\nu^3 c}{\exp\{\frac{hc\nu}{kT}\} - 1},$$

where k is Boltzmann's constant, h is Planck's constant, c is the speed of light and ν is the wavenumber. Planck's function is used in the radiative transfer equation (2.21) described in Secton 2.4.1. The radiative transfer model used in the assimilation of IASI radiances is the radiative transfer model for TOVS (RTTOV) [66].

IASI data is an important component of the global observing system. The assimilation of IASI radiances is operational at the Met Office [45], Meteo-France [72] and ECMWF [13], and is in the testing stage at several other national weather centres. At the Met Office, forecast accuracy has improved through the assimilation of IASI radiances [45] from the general channel subset determined in [15]. However, the assimilation of channels sensitive to water vapour has only shown a weak impact on forecast accuracy. The suspected cause of this underperformance is attributed partially to the mis-representation of the cross-channel observation error correlation structure.

4.1.1 Observation error correlations

The IASI observation errors are treated as horizontally and vertically uncorrelated. The assumption of horizontally independent observation errors is supported by intelligent thinning of the data ensuring that no observations are assimilated at a higher density than model resolution. This is clearly a very inefficient use of the data, but it reduces the complexity of the subsequent assimilation of the radiances.

Ensuring vertically independent observation errors is more difficult. Because of the nature of the IASI instrument, radiance measurements are sensitive to the temperature profile over several atmospheric levels. This distribution is represented by the broad channel weighting functions of the instrument (Figure 4.1). Therefore the errors in adjacent channels (i.e, those close to each other in wavelength) can potentially be correlated; for example, if the sensitivity of the signal to a trace gas present in several adjcent channels is mis-represented. The current IASI channel selection procedure deals with this issue by avoiding the assimilation of adjacent channels. However, this cannot be rigorously enforced because adjacent channels in certain wavelength bands are needed to provide fine scale information on atmospheric profiles; for example, channels in the longwave CO_2 band provide information on temperature and humidity. Therefore some level of error correlation structure will exist between selected channels.

Additionally, correlated errors of representativity are present between channels that observe spatial scales or features that the model cannot. Although the IASI observation spacing of 25km is similar to the Met Office NWP model grid spacing of 40km, IASI is sensitive to small-scale variations within its 12km field-of-view which the NWP model does not attempt to represent. For example, the NWP model may be unable to represent

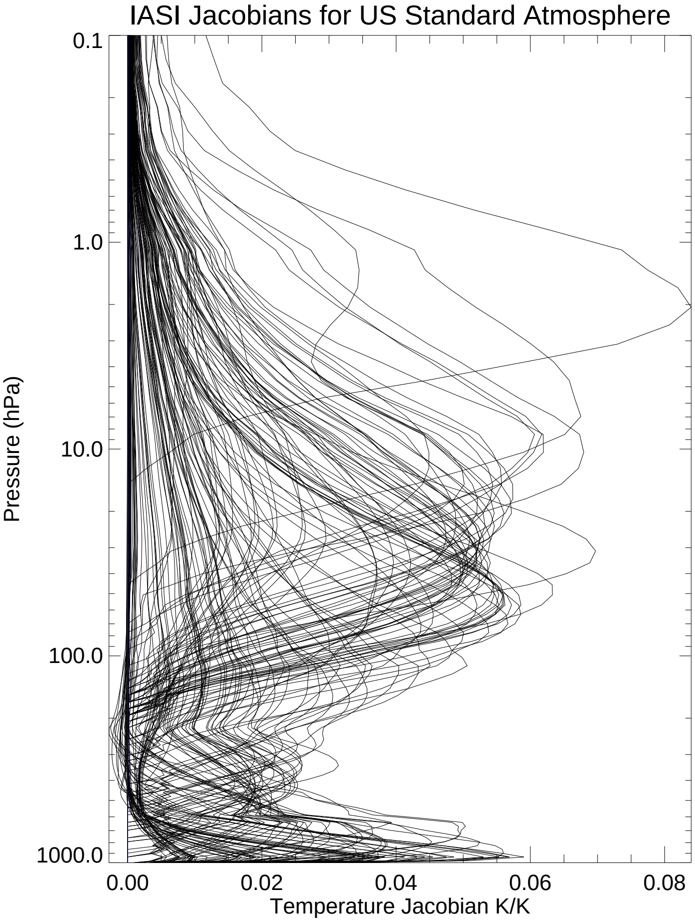


Figure 4.1: IASI weighting functions (provided by Fiona Hilton)

accurately a complex humidity structure at its current resolution, leading to correlations between channels sensitive to water vapour.

Finally, errors in the forward model may be correlated between channels. These include errors in the spectroscopy, an inaccurate discretisation of the radiative transfer equation, and mis-representation of the gaseous contributors in certain channels.

4.1.2 Processing

Any preprocessing performed on the original IASI radiances prior to their assimilation is likely to create errors. At the Met Office before IASI observations are assimilated directly into the NWP model, they are subject to pre-screening and quality control procedures. This is performed in the Observation Processing System (OPS). A schematic of the IASI observations processing path is shown in Figure 4.2.

IASI has the potential to provide observations in 8461 channels, but at present only observations from a subset of 314 are used. IASI measured brightness temperatures from this subset are fed into the OPS and processed using a code specifically written for satellite measurements. This code, known as the SatRad code, implements a 1D-Var assimilation on the bias-corrected brightness temperature measurements, y, and an accurate first-guess model-profile from a short range forecast, x_b. The solution is the state vector x that minimises the cost function,

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - h(x))^T R^{-1}(y - h(x)), \tag{4.1}$$

where h is the observation operator mapping from state space to measurement space, B is

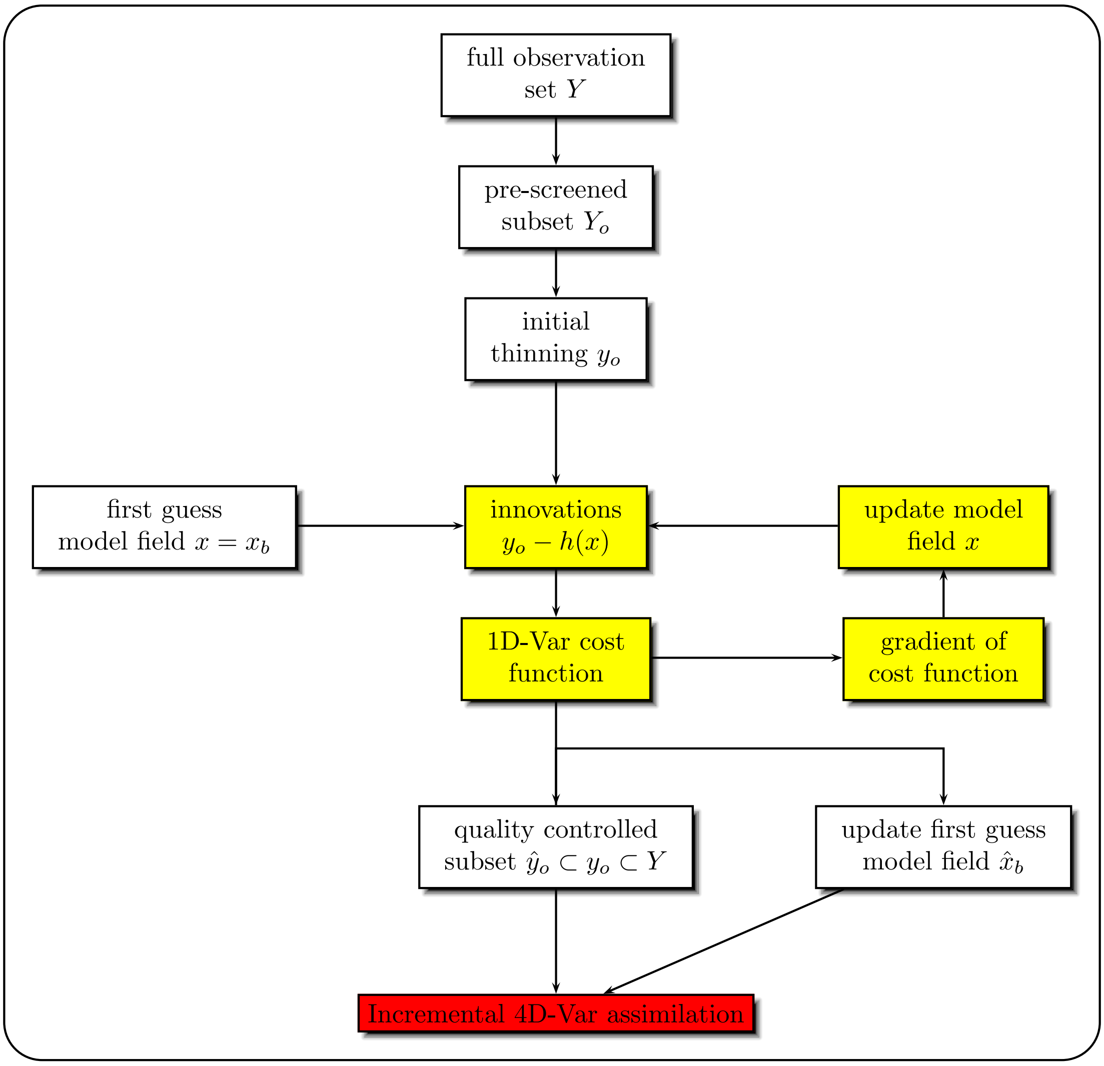

full observation set Y
pre-screened subset Y_o
initial thinning y_o
first guess model field $x = x_b$
innovations $y_o - h(x)$
update model field x
1D-Var cost function
gradient of cost function
quality controlled subset $\hat{y}_o \subset y_o \subset Y$
update first guess model field $\hat{x}_b$
Incremental 4D-Var assimilation

Figure 4.2: Schematic of IASI radiance processing and 1D-Var retrieval (yellow boxes).

the background error covariance matrix and R is the observation error covariance matrix. The observation operator h is comprised of a Radiative Transfer for TIROS Operational Vertical Sounder (RTTOV) radiative transfer model [80], [66]; it accurately predicts brightness temperatures given first-guess model fields of temperature and humidity on 43 fixed pressure levels between 0.1 and 1013hPa, as well as surface air temperature, skin temperature and surface humidity.

The OPS has two main functions: the first is quality control on the brightness temperature measurements, and the second is providing an accurate estimate of the model variables not assimilated in 4D-Var. The assimilation performs a local analysis of the model state at the location of every satellite observation; an observation is suitable for 4D-Var assimilation if its 1D-Var analysis generates a good convergence and a suitable a posteriori cost [88]. Each observation has an associated cost which is scaled to be ideally one, and if the distribution of the costs about 1 were plotted, the quality control procedure would be equivalent to eliminating those observations for which the costs lie in the tails of the distribution.

Unsuitably high costs and slow convergence are caused by inconsistencies between the background and the observations; for example, if the background assumes a clear sky but the observation is affected by cloud. If we consider the prior and likelihood distributions of the background and the observations, then the 1D-Var assimilation finds the solution with maximum probability that satisfies both the background and the observation distribution. If the distributions are highly overlapping, then the solution state will exist with a high probability; if the distributions have a small overlap then the solution state is improbable, convergence to it will be slow, and its cost will be high. Identifying and eliminating these observations in 1D-Var enables a stable and fast convergence in 4D-Var.

The 1D-Var assimilation also provides estimates of the atmospheric variables not represented in 4D-Var. The control vector in 4D-Var is comprised of a subset of the full state vector variables, and those variables, such as skin temperature, which are not included are unmodifiable. It is therefore crucial to the success of the assimilation that these variables are accurately specified prior to the 4D-Var assimilation. For example, if a poorly specified skin temperature is fixed then information from the observations will not be able to improve it. Therefore the control variables, which can be modified, will be fitted incorrectly to the observations. The full state vector is used in the 1D-Var assimilation, and the analysis values of those variables not present in the control vector are passed to 4D-Var.

When the 1D-Var assimilation is performed in the OPS, the forward model is separately fitted to each individual column of observations, so the position of the observations, and hence any resolution conflicts, is already determined. Therefore, it can be argued that the representativity errors will appear in the background matrix B, and so correlations in representativity error within the observation error covariance matrix R will be low. Hence, from the OPS diagnostics (2.33) we expect any error correlations to be mainly attributed to forward model error and pre-processing error.

The OPS produces a quality controlled subset of brightness temperature measurements suitable for assimilation in the Met Office incremental 4D-Var assimilation system [76]. As with the 1D-Var procedure, 4D-Var assimilation aims to minimise a cost function penalising distance from the solution state to the observations and the first-guess background profile (2.15). In 1D-Var, each set of radiance measurements is assimilated at its own horizontal location, while in 4D-Var all observation types are assimilated together at model grid points. The algorithm generates an optimal analysis increment which is used to update the solution state at the start of the assimilation time window. From this starting state, the nonlinear model is run over the time window to generate a forecast. The forecast model fields are output at the model grid points at pre-determined times, and can be interpolated in time and space to the observation locations. In the 4D-Var assimilation, all observation information is fitted to the resolution provided by the model, and so correlated representativity errors are expected to be contained wholly in R.

4.2 Application of the Desroziers' diagnostic

We now describe the methodology for generating the Desroziers' diagnostic (2.33):

$$\mathbb{E}\left[d_a^o(d_b^o)^T\right] \approx R \tag{4.2}$$

where $d_b^o = y - h(x^b)$ is the background innovation vector and $d_a^o = y - h(x^a)$ is the analysis innovation vector. The suitability of the diagnostic (derived from 3D-Var assimilation theory) for 4D-Var assimilation is shown in Appendix B. The diagnostic is calculated for two situations: firstly using the analysis output from the OPS and secondly using the analysis output from the incremental 4D-Var assimilation. The background and analysis increment statistics are generated from the assimilation of only clear sky, sea surface IASI observations. Observations will be from both day and night time, with the exception of daytime observations from shortwave channels which will be eliminated. Using only IASI observations in the assimilation avoids the difficulties of attributing the diagnosed error correlation structure to different observation types. We will now discuss the technical and practical details of the procedure.

4.2.1 Technical details

First we calculate the diagnostic for the OPS retrievals. As previously mentioned, before satellite radiances are assimilated into 4D-Var, they are passed through the OPS for quality control. Within the OPS, a 1D-Var assimilation is performed on the equivalent brightness temperatures and a first-guess background, producing an analysis retrieval. The first set of statistics will be generated using the background, d_b^o, and analysis, d_a^o, innovations from the 1D-Var analysis.

Calculating the diagnostic using the 4D-Var retrievals is more complicated. The initial

OPS run analyses those atmospheric quantities not present in the 4D-Var state vector, and passes them to 4D-Var with a quality controlled set of brightness temperatures; these are used to produce an optimal analysis increment. Along with the forecast value at the start of the time window, the increment is run through the Unified Model (UM) [19] over a 6 hour time window to generate an analysis trajectory. Using the same observation set, the analysis fields can be passed back through OPS (the second OPS run), only this time as the background input. We can therefore use the background innovations generated by OPS as the d_a^o innovation statistics for the 4D-Var assimilation. This process is shown in Figure 4.3.

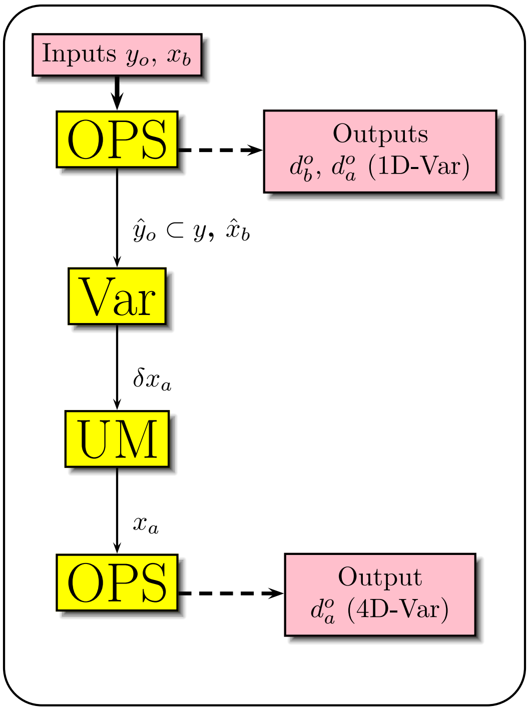


Figure 4.3: Met Office assimilation process: y_o is the initial observation set, $\hat{y}_o$ is the quality control observation subset, x_b is the background, $\hat{x}_b$ is the quality control background, δx_a is the analysis increment, and x_a is the analysis. The yellow boxes represent assimilation steps and the pink boxes represent assimilation inputs and outputs.

Clearly we only want to generate our statistics from those observations that are deemed suitable to process in the Var system, i.e, those that pass the OPS quality control. These are easily identifiable since OPS assigns all observations a quality control flag value: zero if the observation is passed to Var, one if the observation was accepted by Var but spatially thinned out, and greater than one if the observation was rejected. However, the observations passed to Var in the second OPS run will not be the same as those passed in the initial run, because the backgrounds are different. To ensure that the same observations are used to generate the d_b^o innovations in the initial OPS run and the d_a^o innovations in the second OPS run, we match observations using their latitude and longitude values.

4.2.2 Practical calculations

The aim of this chapter is to use the background and analysis increment statistics generated from the assimilation of IASI data, to provide a consistency check on the observation error covariances used in the assimilation. We are interested in the correlations between channels used in (i) the 1D-Var assimilation in OPS (183 channels), (ii) the 4D-Var assimilation (139 channels). Using the diagnostic (4.2), for each channel i, we compute the observation error covariance with channel j by averaging the product of the background and analysis innovations over the total number of observations used in the assimilation N,

$$\begin{aligned} R(i,j) &= \frac{1}{N}\sum_{k=1}^{N}\{(d_a^o)_i\,(d_b^o)_j\}_k - \left(\frac{1}{N}\sum_{k=1}^{N}\{(d_a^o)_i\}_k\right)\left(\frac{1}{N}\sum_{k=1}^{N}\{(d_b^o)_j\}_k\right) \\ &= \frac{1}{N}\sum_{k=1}^{N}\{y_i^o - y_i^a\}_k\{y_j^o - y_j^b\}_k \\ &\qquad - \left(\frac{1}{N}\sum_{k=1}^{N}\{y_i^o - y_i^a\}_k\right)\left(\frac{1}{N}\sum_{k=1}^{N}\{y_j^o - y_j^b\}_k\right), \end{aligned} \tag{4.3}$$

where y_i^o is the brightness temperature value in channel i, and y_i^a and y_i^b are the analysis and background counterparts, respectively. We subtract the mean innovation values to ensure our diagnostic is unbiased.

The diagnostic (4.3) is only an approximation of the observation error covariance matrix. It is only strictly valid when $\mathbb{E}\left[\epsilon^o(\epsilon^o)^T\right] = R$ and $\mathbb{E}\left[\epsilon^b(\epsilon^b)^T\right] = B$, i.e, when the errors assumed in the assimilation are equal to those found in reality. Under such circumstances the resulting matrix would be exact and symmetric. However, we are knowingly using an incorrect specification of the observation error covariances, and so by construction the matrix may not be symmetric. Since an error covariance matrix is required to be symmetric positive definite, we could approximate R with the symmetric component of our diagnosed matrix

$$R_{\text{sym}} = \frac{1}{2}(R + R^T). \tag{4.4}$$

4.3 Results for 1D-Var retrievals

We now perform a set of experiments using the techniques described in the previous section. First we consider the diagnostic (4.2) applied to the OPS analyses. A set of analyses are produced by the 1D-Var assimilation of IASI data from the 17th July 2008 at 00z, 06z, 12z and 18z, within the Observation Processing System. The total number of observations used to produce the statistics is 27,854; 9,131 of which are suitable for use in the 4D-Var assimilation and 18,723 of which are thinned out. Figures 4.4 and 4.5 show the global location of all the observations used in the OPS assimilation, and the size of their background innovations for MetDB channel 1 (sensitive to stratospheric temperature) and MetDB channel 279 (sensitive to water vapour), respectively. Using the formula (4.3), we calculate the observation error covariances for this data.

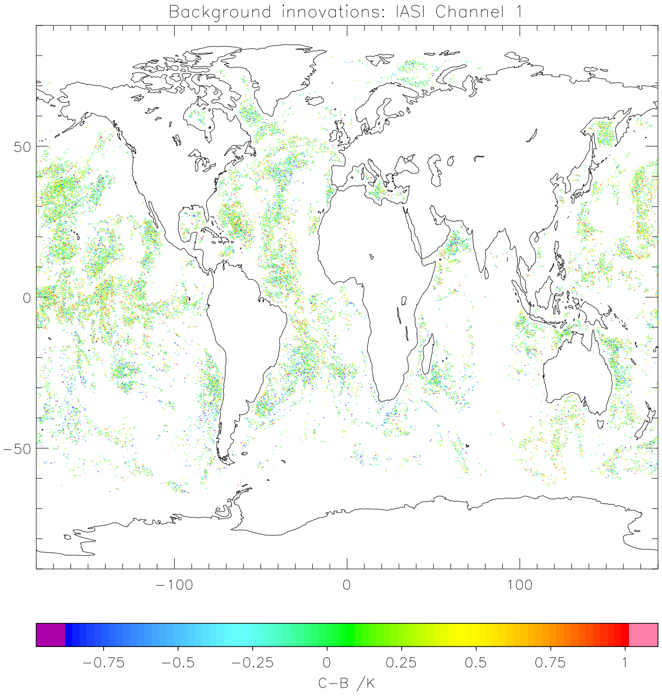


Figure 4.4: Global location and background innovation value $C - B$ (degrees Kelvin) for observations in MetDB channel 1

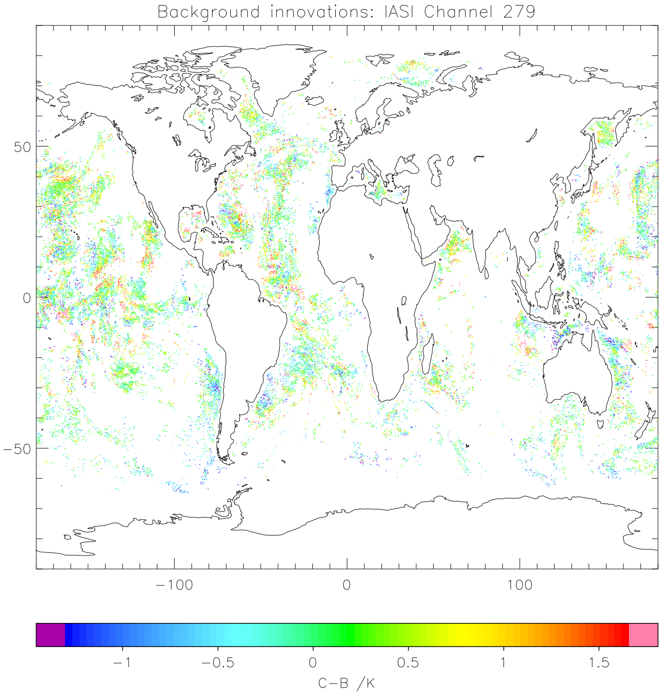


Figure 4.5: Global location and background innovation value $C - B$ (degrees Kelvin) for observations in MetDB channel 279

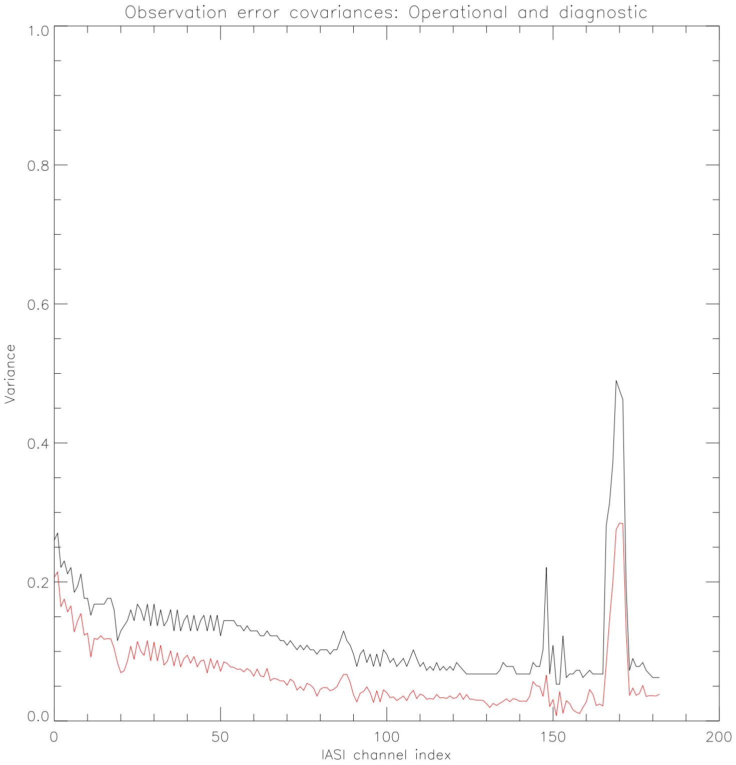


Figure 4.6: Operational error variances (black line) and diagnosed error variances (red line) (K^2)

Figure 4.6 shows the operational observation error variances used in the 1D-Var assimilation (black line) and the error variances diagnosed by (4.3) (red line) for all the 183 channels used in the OPS. The error standard deviations used in the OPS (square root of the variances) are comprised of the instrument noise plus a forward model error of 0.2K. The channels numbers correspond to the index of the MetDB channel used in the OPS, i.e, MetDB channel number 1 has OPS channel index 0 (the first channel used in the OPS) and MetDB channel 280 has OPS channel index 182 (the last channel to be used in the OPS). Figure 4.7 shows a typical IASI spectrum for all 314 channels; the channels used in OPS are highlighed by the red asterisks (the full list of corresponding channel numbers and indices can be found in Appendix A).

The structure of the operational and diagnostic error variances in Figure 4.6 is very similar. The diagnosed error variance is significantly lower than the current operational

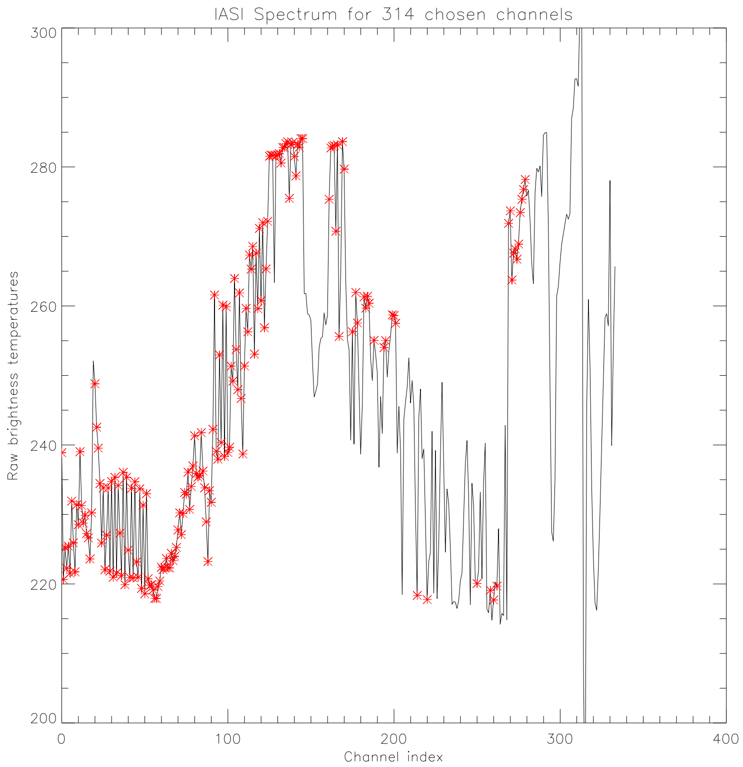


Figure 4.7: Channels used in OPS (red asterisks) on a typical IASI spectrum (black line)

variance for all channels. The largest difference is in the OPS indexed channels 145 - 180 which are highly sensitive to water vapour. In conclusion, the results suggest that the error variances are being overestimated especially in channels sensitive to water vapour.

Figure 4.8 shows the observation error covariances for the 183 channels used in the OPS. The error covariance plot is heavily diagonally dominant; the diagonal values in Figure 4.8 correspond to the values plotted on the red line in Figure 4.6. The darker colours towards the top of the diagonal in Figure 4.8 correspond to the larger variance values in channels 165-172 in Figure 4.6. The correlations are relatively weak between channels with OPS index under 120, with the exception of channel 20 (MetDB number 21). However, channel 20 is a high-peaking channel in the temperature sounding band, which is not used in the 4D-Var assimilation because of the stratospheric ringing of its innovations.

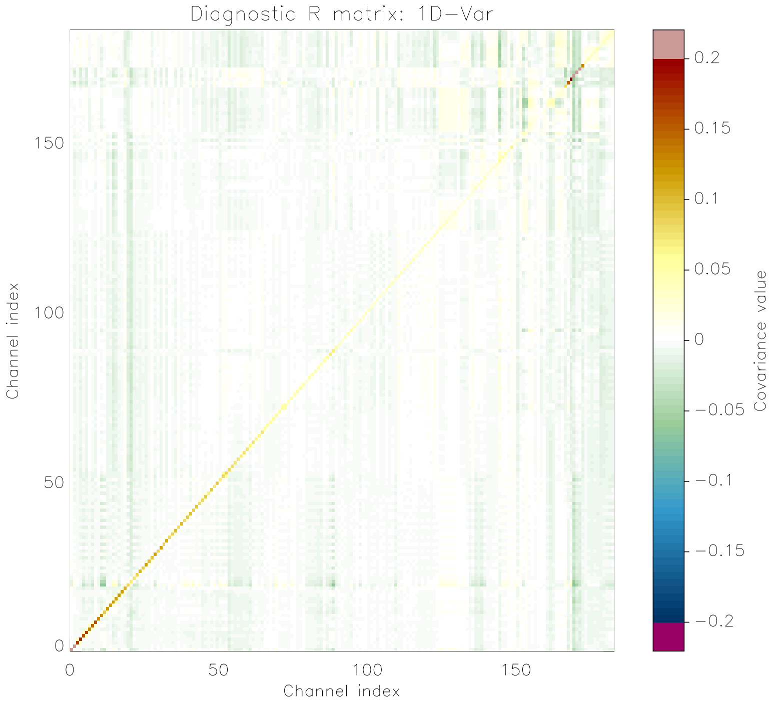


Figure 4.8: Diagnosed observation error covariance matrix for the OPS (K^2)

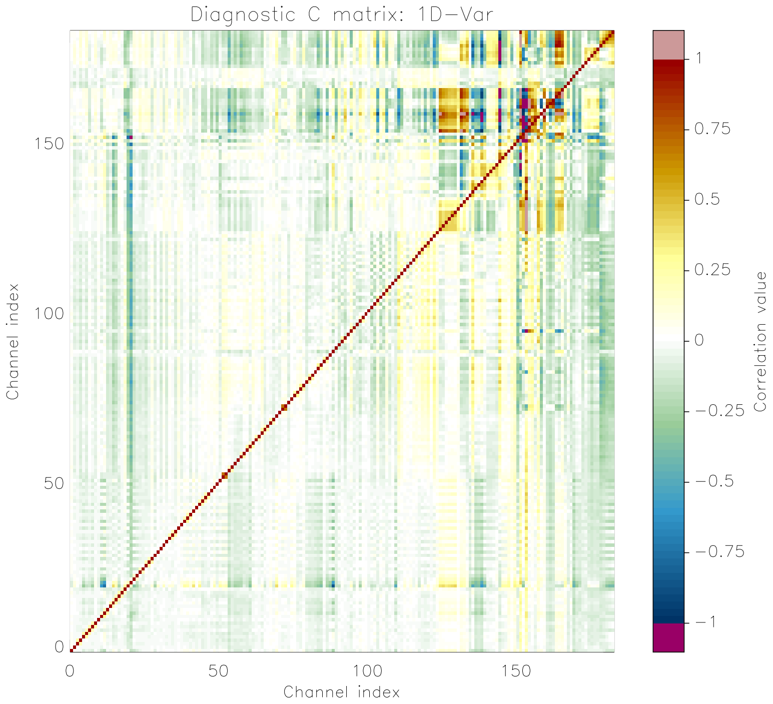


Figure 4.9: Diagnosed observation error correlation matrix for the OPS

The error correlation matrix can be determined easily from the error covariance matrix using the identity $R = D^{1/2}CD^{1/2}$ from Chapter 2.5.2; the diagnosed error correlation matrix is shown in Figure 4.9. The correlation structure shown in Figure 4.9 is not uniformly symmetric, suggesting that the iterative procedure for updating the error variances (as proposed by Desroziers [25]) could be beneficial.

From the results in Figure 4.9 we can conclude that when the IASI observations are analysed in the OPS, observation error cross-correlations are very small for most channels. This can be explained by recalling that only forward model and adjacent channel error correlation is expected to appear in the observation error covariance matrix. When we use the analyses from the 4D-Var assimilation, we expect the cross-channel correlations to be larger because correlated error of representativity will also be contained in R.

4.4 Results for 4D-Var retrievals

We now calculate the observation error covariances using the analysis innovations derived from the 4D-Var assimilation of IASI data. The set of analyses are produced by the assimilation of data from 17th July 2008 at 18:00. A total of 2,073 observations are used to produce the statistics; this is a subset of the 6,539 observations that are used in the OPS at this time. We expect a stronger correlation structure than that diagnosed using the 1D-Var retrievals.

Figures 4.10 and 4.11 show the diagnosed observation error covariances and correlations, respectively, for the 139 channels used in 4D-Var. Comparing Figure 4.10 to Figure 4.8, we observe that the variances are notably larger in the 4D-Var matrix (up to 0.8 in Var channel 263). Also there exist larger off-diagonal covariances in Figure 4.10. There are

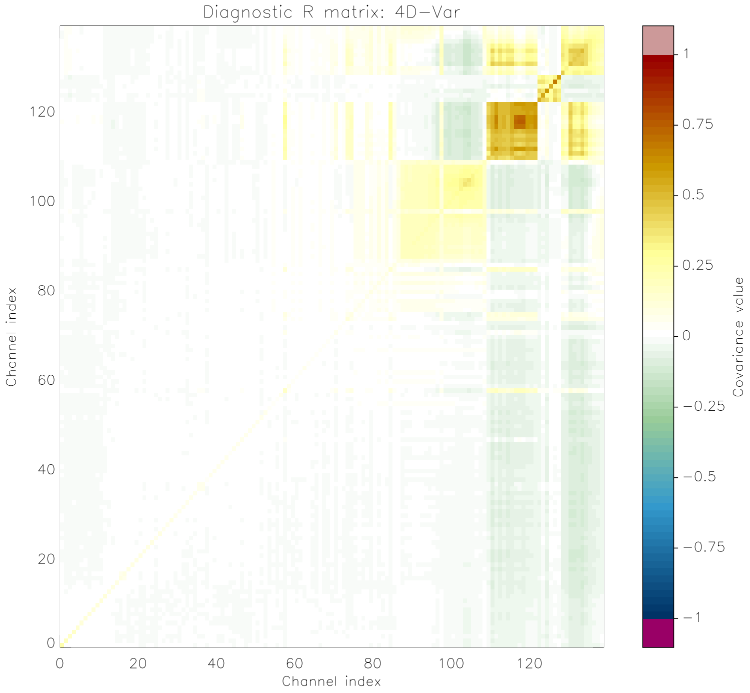


Figure 4.10: Diagnosed observation error covariance matrix for 4D-Var assimilation (K^2).

four significant block structures of covariance centred around the diagonal: the first for Var channels between 124-171 (index 86-108) which are window channels sensitive to surface temperature and emissivity, the latter three for Var channels 176-202 (index 109-121), 215-263 (index 122-127), and 270-280 (index 128-138) which are sensitive to water vapour. The block structure implies the channels in each block have highly correlated errors. This can be seen more clearly in the Figure 4.11 plot of the observation error correlation matrix.

Figure 4.12 shows a typical IASI spectrum; the channels present in each of the block structures in Figure 4.11 are marked by coloured asterisks, and channels outside these blocks are marked by black asterisks. Combined with the table of spectral information in Appendix A, we see that the channels that carry significant correlations between them have similar spectral properties. For example, the channels between 215-263 used in Var

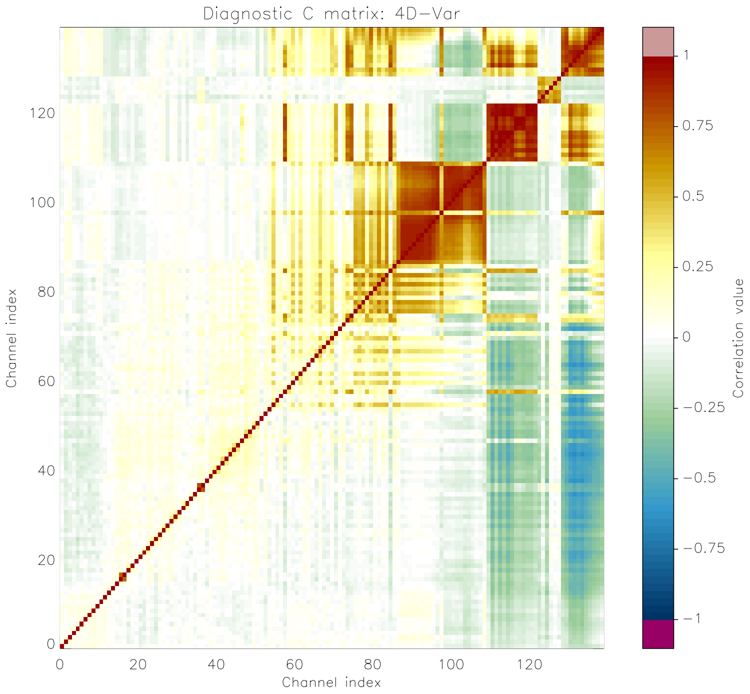


Figure 4.11: Diagnosed observation error correlation matrix for 4D-Var assimilation

(index 122-127) have typical brightness temperature measurements between 217-222K, and a Q jac peak (see Appendix A) at 208.16hPa. Var channel blocks 176-202 and 270-280 also have similar brightness temperature measurements and a strong correlation structure in Figure 4.11.

In addition to the block diagonal structure, Figure 4.11 also shows bands of correlation surrounding the first, and largest, block structure. Using the summed Q jac value (see Appendix A) as a measure of the sensitivity of a channel to water vapour, we observe that channel index 60 (Var channel 98) has a value of 0.113 and a significant band of correlation, while channel index 61 (Var channel 99) has a value of 0.022 and a near zero correlation value with the surrounding channels. The largest summed Q jac values are found in the low-peaking Var channel blocks 176-202 and 270-280 (up to 1.000 in channel 272). From these results and the diagnosed block correlation structure of

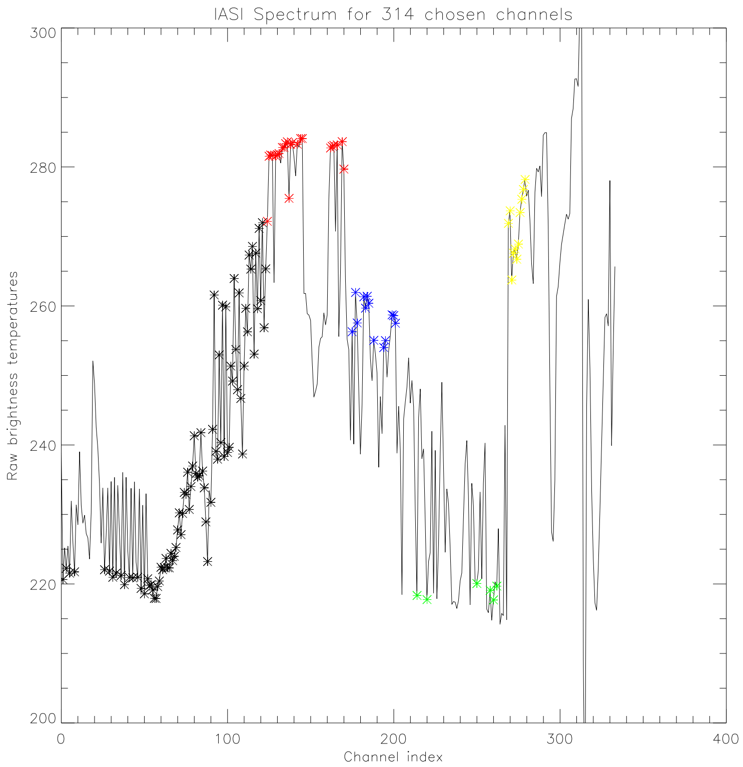


Figure 4.12: Channels used in 4D-Var on a typical IASI spectrum: Var channel indexed 0-85 (black), 86-108 (red), 109-121 (blue), 122-127 (green), 128-138 (yellow)

channels sensitive to water vapour in Figure 4.11, we can infer that strong correlated errors of representativity exist between channels sensitive to water vapour. This implies that some fine scale humidity structure is represented in the IASI observations but not in the NWP model.

Although correlations are largest in those channels highly sensitive to water vapour, a weaker level of correlation is also present in the channels used in temperature sounding. Figure 4.13 shows two fainter blocks of correlation centred on the diagonal for channel indices 0-10 (Var channels 2-42) and 11-50 (Var channels 44-88) ; channels 14 and 24 are highly correlated with their neighbouring channel within these blocks. Many of the channels within these blocks are adjacent to each other, and therefore we would expect some level of error correlation structure. The differences in measurements between these channels can be used to capture fine scale information on humidity and temperature

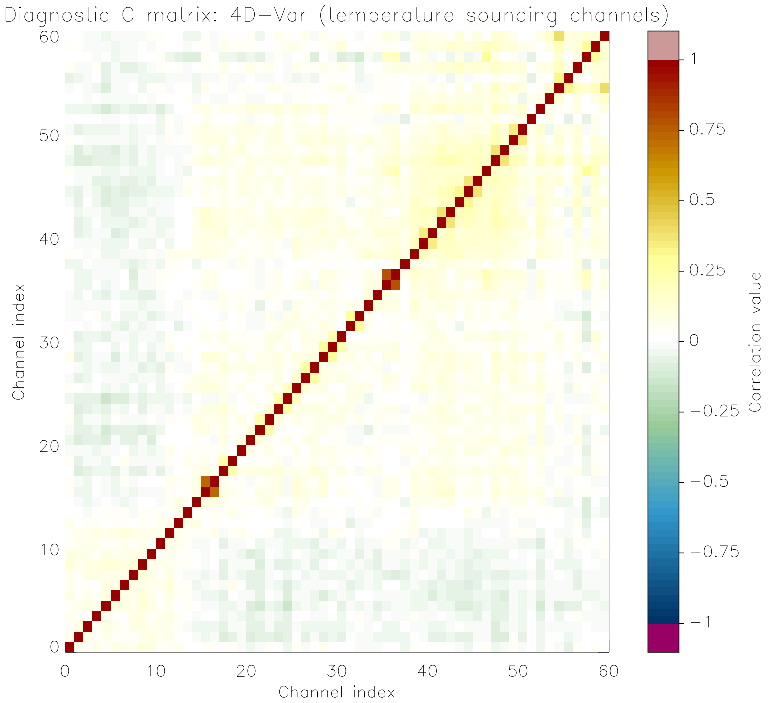


Figure 4.13: Diagnosed observation error correlations in temperature sounding channels indexed 1-60

profiles; it is therefore desirable to include even a weak level of correlation structure in an attempt to lower the operational error variances and hence retain more information.

An important feature of the matrices displayed in Figure 4.9 and 4.11 is their non-symmetry. Although the matrices are predominantly symmetric, water vapour sensitive channels 109-121 and 18-138 in Figure 4.11, for example, display asymmetric correlations between channels. The departures from symmetry can be attributed to departures from the assumption of using the correct observation and background errors in the construction of (2.33). Our results demonstrate that observation errors are correlated between certain channels, but the observation error covariance matrix used in 4D-Var is diagonal. This emphasises the need to include a correlation structure in the error covariance matrix.

We conclude this section by comparing the diagnosed error variances with those currently used operationally. In the previous section, we found that the error variances were being over-estimated in the OPS. Figure 4.14 shows the observation error variances used in 4D-Var (black line), the error variances diagnosed using (4.3) (red line) and the first off-diagonal from the symmetrised diagnosed error covariance matrix (green line). For all channels, the diagnosed variances are considerably less than those being used operationally, implying an overestimation of observation error variances in 4D-Var. However, the size of the first off-diagonal covariance value (green line) indicates why this overestimation might take place. For most of the Var channels indexed 86 upwards, the first off-diagonal covariance value is very close in size to the diagonal variance value, therefore ignoring this value and other off-diagonal covariances will result in over-weighting the observations in the analysis. We conclude that it is therefore necessary to inflate the error variances if we choose to ignore large off-diagonal covariances, as previously suggested in [14].

4.5 Summary and conclusions

In order to model successfully observation error correlations, an accurate knowledge of the true correlation structure is needed. This structure varies with observation type and instrument. In this chapter we have successfully used a post-analysis diagnostic derived from variational data assimilation theory to obtain the cross-channel error correlations for IASI observations. We first introduced IASI data and commented on its current usage at the Met Office. The technical and practical details of applying the Desroziers' diagnostic to the assimilation of IASI observations were then discussed. Background and analysis innovation statistics were acquired through the assimilation of IASI observations

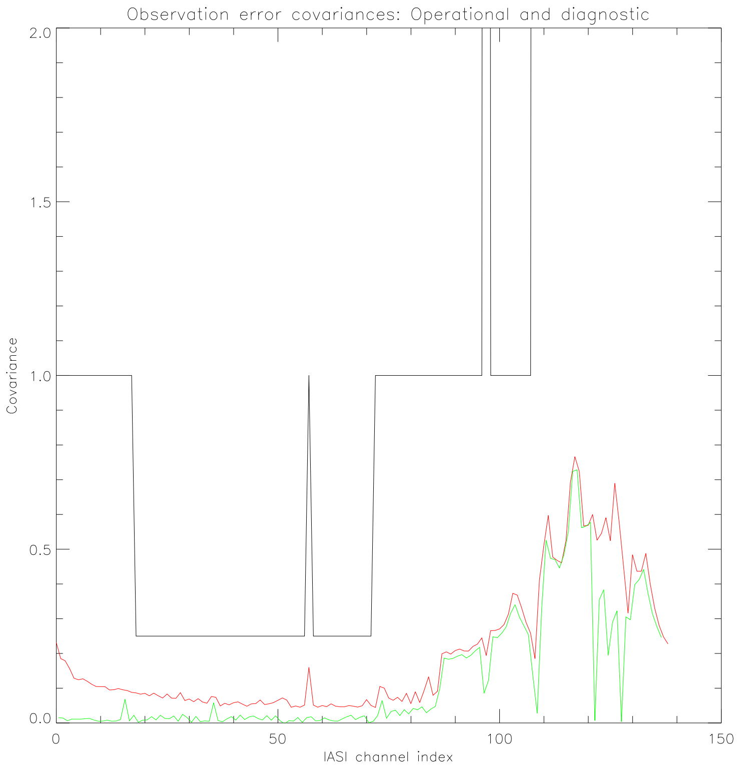


Figure 4.14: Operational error variances (black line), diagnosed error variances (red line), and first off-diagonal error covariance (green line) in K^2. Operational error variances not shown off the top of the plot are all equal to $4K$.

in the OPS and then the incremental 4D-Var assimilation system.

The current treatment of vertical IASI observation errors is to assume independence between channels, i.e, the observation error covariance matrix is diagonal. The new results in this chapter have challenged the validity of this assumption. The statistics from the 4D-Var assimilation showed large off-diagonal error covariances in channels highly sensitive to water vapour, and additional correlation structure in channels in the temperature sounding band. Observation error correlations were shown to be significant between neighbouring channels with similar spectral properties, leading to a block structure in the error covariance and correlation matrix.

However, the statistics from the 1D-Var assimilation identified predominantly uncorrelated errors between channels, with some weak correlation in those channels sensitive to water vapour. These findings suggest that correlated observation errors in IASI data can largely be attributed to errors of representativity.

The application of the post-analysis diagnostic to both the 1D- and 4D-Var assimilation procedures recorded observation error variances considerably smaller than those currently being used operationally. We can attribute this over-inflation to the assumption of uncorrelated errors. In the 4D-Var assimilation, the diagnosed error covariances between certain channels are very large, and ignoring these will lead to a mis-weighted representation of the observations in the analysis. Therefore inflating the variances is necessary if all observation errors are assumed independent. If we are to change this assumption, a suitable representation of the error correlation structure is needed.

The diagnosed values of observation error covariances and correlations generated here provide a realistic starting point for future work on including observation error correlation structure in variational data assimilation. The block diagonal structure in the error correlation matrix highlights the potential use of Markov representations for each of the blocks, for example. Although the diagnosed matrices are not entirely symmetric, the data provides us with an approximation of the 'true' correlation structure, and an approximating symmetric matrix (4.4) can be generated. Against this matrix it is possible to compare analytic error correlation structures by examining features such as information content and analysis accuracy.

In the next chapter we run some initial statistical experiments comparing the matrix approximations described in Chapter 3 against an empirically derived true error correlation structure. The aim is to determine if an estimation of true error correlation structure can retain important features of the data.

Chapter 5

Information content studies in a 3D-Var framework

In Chapter 3 we described several matrix approximations suitable for representing error correlation structure in data assimilation algorithms. In Chapter 4 we demonstrated the quantitative behaviour of observation error correlations between channels of the IASI instrument. The results showed an error correlation matrix with close to block diagonal structure and strong off-diagonal correlations present between channels with similar spectral properties. In this chapter we present new results which quantify the success of each of the matrix approximations described in Chapter 3 in modelling an empirically derived error correlation structure. Specifically we address the second of our thesis aims: what is the impact of error covariance approximations on data assimilation algorithms?

To this end we examine the information content provided by a 2D set of observations when assimilated using their true error correlation matrix and the proposed approximations. The assimilation technique we use is three-dimensional variational assimilation (3D-Var) which was introduced in Section 2.2. Although we are only considering two spatial dimensions, we shall use the terminology 3D-Var for the sake of convention. We use the 3D-Var method because the equations for calculating information content are available explicitly without the added complications of four-dimensional variational assimilation.

We begin the chapter by recalling two measures of information content: Shannon Information Content (or entropy reduction) and the degrees of freedom for signal as described in Section 3.5.2. We describe the formula for each of these measures under the different possible constructions of the analysis error covariance matrix. The empirically derived observation error covariance matrix against which we test our approximations is diagnosed from pairs of Atmospheric Motion Vector and radiosonde collocations. We use these results over the diagnosed structure in Chapter 4 because a known correlation function has previously been fitted to the empirical data [7].

Results for diagonal and correlated approximations to the diagnosed error structure demonstrate (a) the usefulness of including some level of correlation structure if sufficient information is to be retained; (b) the sensitivity of the information content to the background error correlation structure; and (c) the importance of knowing the true observation error correlation structure even when approximations are to be made.

5.1 Information content

The overall aim of the experiment is to calculate the information content for a set of observations when (a) the true error covariance matrix, R_t, is used, and (b) an approximate error covariance matrix, R_f, is used in the assimilation process. We will use the measures of Shannon Information Content (SIC) and degrees of freedom for signal (dof_S) to quantify the information provided by the observation set. Recall from Section 3.5.2, the SIC and dof_S can both be described in terms of the scaled analysis variances:

$$\begin{aligned} SIC &= \frac{1}{2}\ln\frac{|B|}{|S_a|}, && (5.1)\\ dof_S &= n - \text{trace}(B^{-1}S_a), && (5.2) \end{aligned}$$

where B is the background error covariance matrix, S_a is the analysis error covariance matrix, and n is the total number of observations.

The formula for the analysis error covariance matrix, S_a, varies under the specification of the observation error covariance matrix R. In Section 3.5.1, three possible specifications for the analysis error covariance matrix were given. Below we will describe their application to the information content measures.

The first approach is when we use the correct error covariance matrix R_t. The analysis error covariance matrix is then given by

$$S_a^{(1)} = (H^T R_t^{-1} H + B^{-1})^{-1}, \tag{5.3}$$

and hence the information measures can be written as

$$\begin{aligned} SIC^{(1)} &= \frac{1}{2}\ln|B(S_a^{(1)})^{-1}|, && (5.4)\\ &= \frac{1}{2}\ln|B(H^T R_t^{-1} H + B^{-1})|, && (5.5)\\ &= \frac{1}{2}\ln|BH^T R_t^{-1} H + I|, && (5.6)\\ dof_S^{(1)} &= n - \text{trace}(B^{-1}(H^T R_t^{-1} H + B^{-1})^{-1}), && (5.7)\\ &= n - \text{trace}((H^T R_t^{-1} HB + I)^{-1}). && (5.8) \end{aligned}$$

The second approach is when we knowingly use the incorrect error covariance matrix R_f and include an extra term in the analysis error covariance matrix to model this [33]. The analysis error covariance matrix becomes

$$S_a^{(2)} = (H^T R_f^{-1} H + B^{-1})^{-1} + K(R_t - R_f)K^T, \tag{5.9}$$

where K is the Kalman Gain matrix (2.10) evaluated at R_f. The information measures under these conditions are given by

$$\begin{aligned} SIC^{(2)} &= \frac{1}{2} \ln \frac{|B|}{|(H^T R_t^{-1} H + B^{-1})^{-1} + K(R_t - R_f)K^T|}, && (5.10) \\ dof_S^{(2)} &= n - \text{trace}(H^T R_f^{-1} H B + I)^{-1} - \text{trace}(B^{-1} K(R_t - R_f)K^T). && (5.11) \end{aligned}$$

Finally the third approach is when we know that we are using an incorrect error covariance matrix but do not know what the true structure is. We therefore use R_f as the true error covariance matrix. The analysis error covariance matrix is then defined to be

$$S_a^{(3)} = (H^T R_f^{-1} H + B^{-1})^{-1}, \tag{5.12}$$

and hence the information measures can be written as

$$\begin{aligned} SIC^{(3)} &= \frac{1}{2} \ln |B H^T R_f^{-1} H + I|, && (5.13) \\ dof_S^{(3)} &= n - \text{trace}(H^T R_f^{-1} H B + I)^{-1} && (5.14) \\ &= dof_S^{(2)} + \text{trace}(B^{-1} K(R_t - R_f)K^T). && (5.15) \end{aligned}$$

Using the third approach can potentially give a mis-leading estimate of the information content. This can be seen in equation (5.15), where if we treat the incorrect error covariance matrix as the truth then the degrees of freedom for signal has an extra term, $\text{trace}(B^{-1} K(R_t - R_f)K^T)$, compared to the true value, $dof_S^{(2)}$. Although this approach is sometimes unavoidable because the true error covariance matrix is not accurately known, the mis-calculation must be taken into account when interpreting any results.

5.2 Data structure

We now describe the empirically derived correlation matrix against which we will test our approximations. In [7] Bormann et al considered an observation set of Atmospheric Motion Vector (AMV) / radiosonde collocations, and derived a spatial error correlation structure. AMVs (known as satellite winds) are very important in NWP because of their excellent spatial and temporal global coverage. AMVs are derived operationally by tracking clouds in the infrared, water-vapour or visible channels, or clear-sky features in the water-vapour channels [7]. Correlated observation errors in AMVs can stem from tracking similar cloud structures in neighbouring channels and the use of temperature profiles (with correlated errors) in height assignment. However, to avoid added complexity in their assimilation, observation errors are assumed independent between neighbouring observations. To justify this assumption, AMVs are thinned to a lower resolution, and the observation error variances are inflated. However, error correlations will still remain.

In [7] the spatial correlations of the random errors in AMVs were derived from the analysis of pairs of collocations between AMVs and radiosonde observations. Using these statistics and assuming spatially uncorrelated sonde errors, the spatial AMV error correlations were obtained over dense sonde networks using a modified Hollingsworth-Lönnberg method. Any correlations between the AMV-sonde differences of two observation points were attributed to the spatially correlated AMV errors.

The authors quantified the isotropic correlations by deriving a least squares fit of correlation function to the empirical correlation data. The correlation function used was one derived in [20],

$$R(r) = R_0 \left(1 + \frac{r}{L}\right) \exp\{-r/L\} \tag{5.16}$$

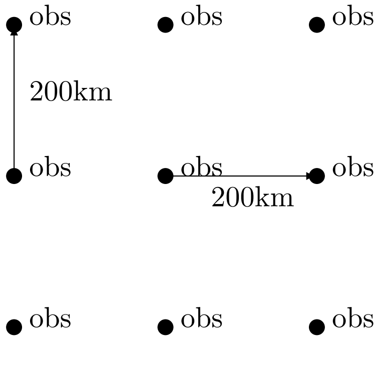
where r is the distance between observation stations, L is the length scale and $R_0 > 0$ is the intercept.

The parameter values R_0 and L, and the correlated part of the AMV error σ, were derived for five satellites and for different pressure levels in the northern hemisphere, tropics and the southern hemisphere. The full details of these results are given in [7]. We will use the results for the GOES-10 satellite in the northern hemisphere at all pressure levels; the parameter values are shown in Table 5.1.

Parameter	Value
R_0	0.42
L	190km
σ	3.5m/s

Table 5.1: Diagnosed values of the intercept, length scale and error standard deviation for the correlation function (5.16) applied to the spatial AMV correlations for the satellite GOES-10 at all levels in the northern hemisphere.

Using the diagnosed parameter values from [7] and the specified correlation function (5.16), we construct an observation error covariance matrix for an idealised data set. Consider N^2 observations on a regular flat $N \times N$ grid, with 200km spacing between adjacent observation points, i.e, for a 3×3 grid, there are 9 observations and the grid domain is 400km × 400km



We assume that every observation is taken directly, i.e, $h = H = I$. We use two different structures for the uniform background error correlations: firstly the identity matrix, and secondly the Gaussian correlation function $B_{ij} = \exp\{-r_{ij}^2/2L_B^2\}$, where r_{ij} is the Euclidean distance between point i and j, and $L_B = 190$km is the background length scale. The background error variance is set as $\sigma_B^2 = 1$m/s at each grid point; this is chosen to be smaller than the observation error variance in line with the experiments performed in [7]. We quantitatively evaluate the information content under different treatments of the error correlation structure relative to the empirically diagnosed truth.

5.3 Matrix approximations

We now propose several approximations to the empirically determined error covariance matrix described above. We wish to compare the information content available from a simulated observation set. The five different approaches to observation error correlation structure are:

(1) **Use the true error covariance matrix R_t**;

This involves using a matrix described by (5.16) with the parameters in Table 5.1. Figure 5.1 shows the correlation function used in the construction of R_t.

(2) **Set the correlations to zero in R_t**;

This is the simplest form of a diagonal approximation as described in Section 3.1. It is equivalent to setting the observation error correlation matrix as the identity matrix.

(3) **Set the correlations to zero in R_t and inflate the error variances**;

This form of diagonal approximation is often used operationally in NWP centres (see Section 3.1). We inflate the error variances by a constant scale factor d of between 2 and 8 to compensate for the elimination of the off-diagonal error covariances. This is in line with previous information content studies perfomed in [14].

(4) **Describe R_f by a circulant approximation**;

By construction the true error covariance matrix has a symmetric Toeplitz structure and can therefore be approximated by its equivalent circulant matrix using the technique described in Section 3.2.3 [43]. This allows us to use a series of discrete Fourier transforms to perform any computations involving the inverse of R_f.

(5) **Describe R_f by a truncated eigendecomposition (ED) approximation**;

The leading eigenvalues and eigenvectors of the true error correlation matrix C_t can be calculated using the inbuilt MATLAB function *eigs()* [65]. We can then approximate R by a truncated eigendecomposition of the error correlation matrix using the formula [34]

$$R_f = D^{1/2}(\alpha I + \sum_{k=1}^{K} \left(\lambda_k - \alpha\right) v_k v_k^T\right) D^{1/2} = D^{1/2} C_f D^{1/2}, \tag{5.17}$$

where (λ_k, v_k) is an eigenvalue, eigenvector pair of C_t, K is the number of leading eigenpairs used in the approximation, C_f is the approximate error correlation matrix, and α is chosen such that trace(R_f)=trace(D). This method was previously described in Section 3.3.

We can write α explicitly in terms of the leading eigenpairs. If Λ is the diagonal matrix of the leading eigenvalues and V is a matrix of the corresponding eigenvectors stored columnwise, then alpha is calculated by

$$\alpha = \frac{N^2 - \sum_{j=1}^{K} \sum_{i=1}^{N^2} \Lambda(j,j) V(i,j)^2}{N^2 - \sum_{j=1}^{K} \sum_{i=1}^{N^2} V(i,j)^2}. \tag{5.18}$$

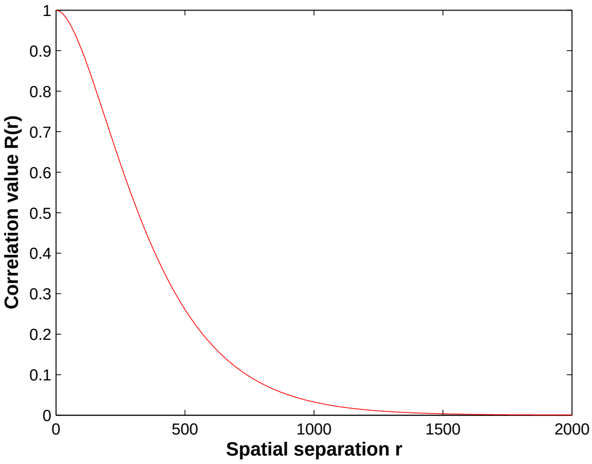


Figure 5.1: Correlation function (5.16) with length scale $L = 190$ against grid point spatial separation.

5.4 Results

We first present the information content values calculated using the second approach to the formulation of the analysis error covariance matrix $S_a^{(2)}$, i.e, adding an extra term to S_a when the incorrect observation error covariance matrix R_f is used. As anticipated, under simplified assumptions of observation error correlations, information content is lost. Both the SIC and dof_S are directly proportional to the number of observation points; so increasing the area and volume of observations provides access to more information.

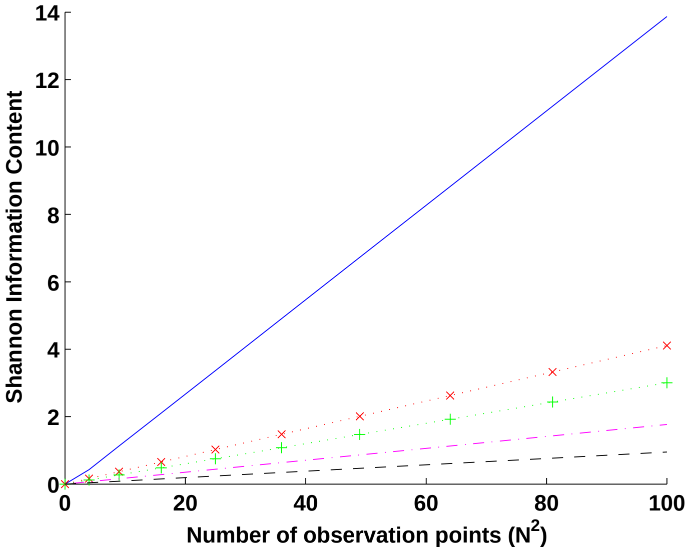


Figure 5.2: SIC for different grid sizes using R_t (blue line), diagonal approximation (2) (red crossed line), and diagonal approximation (3) with $d = 2$ (green plus line), $d = 4$ (pink dot-dashed line) and $d = 8$ (black double-dashed line).

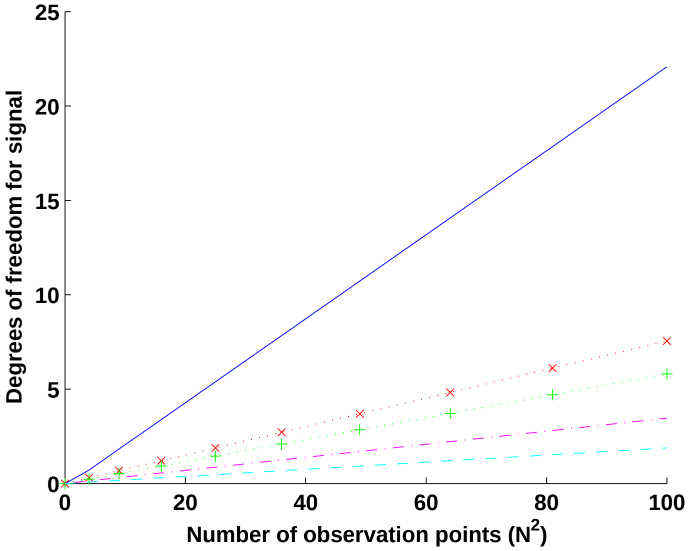


Figure 5.3: dof_S for different grid sizes using R_t (blue line), diagonal approximation (2) (red crossed line), and diagonal approximation (3) with $d = 2$ (green plus line), $d = 4$ (pink dot-dashed line) and $d = 8$ (cyan double-dashed line).

5.4.1 Uncorrelated background errors

Figures 5.2 and 5.3 show the SIC (5.10) and dof_S (5.11), respectively, when R_t (1) and the diagonal approximations (2) and (3) are used; the background error covariance matrix is the identity matrix. We see that using a diagonal approximation is very detrimental to the information content. As the number of observation points increases, the greater the difference in information content between R_t and the diagonal approximations. The depletion in information increases with the scale of variance enlargement used in approximation (3). Variance enlargement is shown to have a detrimental effect on the information; more so than a simple diagonal approximation (2).

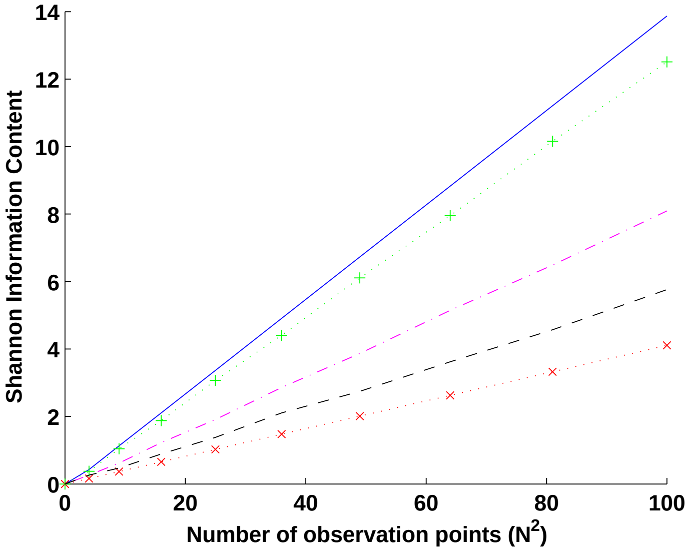


Figure 5.4: SIC for different grid sizes using R_t (blue line), diagonal approximation (2) (red crossed line), and ED approximation (4) with half the eigenpairs (green plus line), a fourth of the eigenpairs (pink dot-dashed line) and an eighth of the eigenpairs (black double-dashed line).

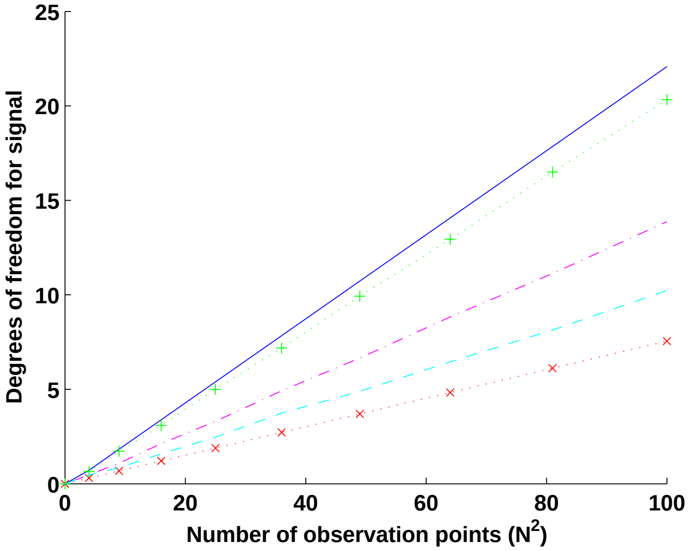


Figure 5.5: dof_S for different grid sizes using R_t (blue line), diagonal approximation (2) (red crossed line), and ED approximation (4) with half the eigenpairs (green plus line), a fourth of the eigenpairs (pink dot-dashed line) and an eighth of the eigenpairs (cyan double-dashed line).

In approach (5) we approximate the error correlations by forming a truncated eigendecomposition of R_t. Figures 5.4 and 5.5 show that the more eigenpairs used in the decomposition, the smaller the difference in information between R_t and the eigenpair approximation. The eigenpair approximation retains a higher percentage of the information available than the diagonal approximations. For all grid sizes, even when R_t is described by an eighth of its eigenpairs, the resultant loss of information is considerably less than for any diagonal approximation. However, in describing R_t by its leading eigenpairs, using too few will lead to spurious error correlations as suggested by Fisher [34]. Figures 5.6 and 5.7 show the true error covariance matrix for a 10×10 grid and one reconstructed using an eighth of the eigenpairs, respectively. Spurious long range correlations are noticable in the ED reconstruction. Under this set up the correlations are not large enough to discount the approach, but care must be taken for larger problems so as to avoid potential spurious correlations in the analysis error.

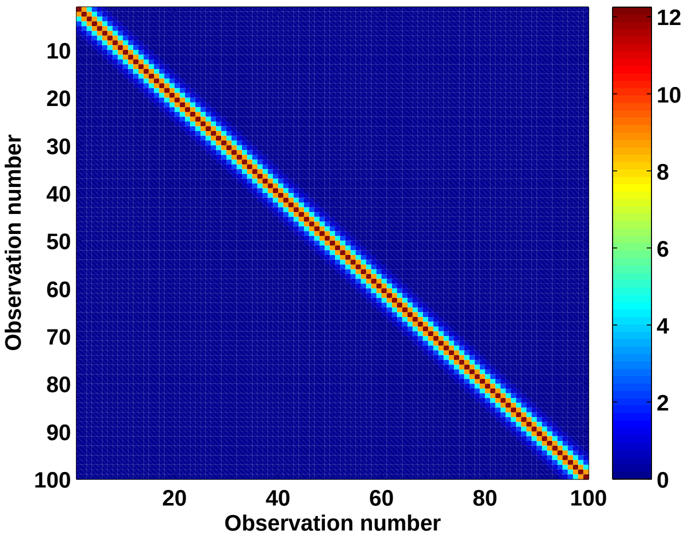


Figure 5.6: Observation error covariance matrix R_t.

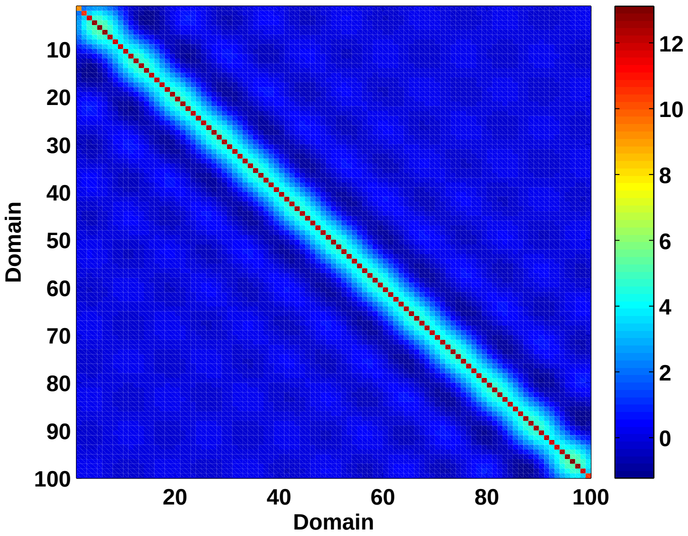


Figure 5.7: ED approximate observation error covariance matrix constructed using an eighth of the eigenpairs.

5.4.2 Correlated background errors

We now compare the results using the same formulation of the analysis error covariance matrix but instead using a correlated background error covariance matrix. The background error correlations follow a Gaussian distribution as described in Section 5.3. When an ED approximation (5) is used, the results are qualitatively the same as when $B = I$, but quantitatively the information content is smaller. This is because when the background errors are treated as correlated there is less prior uncertainty in the problem, so the same reduction in uncertainty through the use of the observations gives a smaller information content, i.e, less difference has been made. However when diagonal approximations (2) and (3) are used, the nature of the results differs from those when the background errors were independent.

Figures 5.8 and 5.9 show that when the background error correlations follow a Gaussian distribution with the same length scale as the observation error correlations, using a diagonal matrix with the variance inflated between 2-4 times retains more information

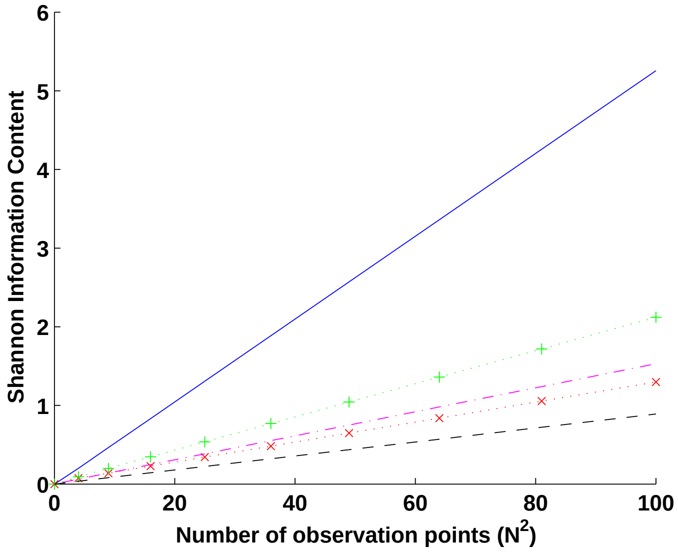


Figure 5.8: As for Figure 5.2 but with correlated background errors.

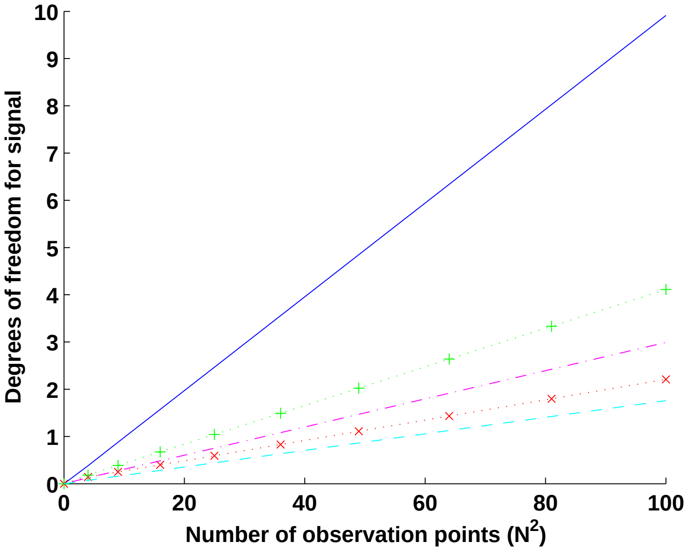


Figure 5.9: As with Figure 5.3 but with correlated background errors.

than a simple diagonal approximation. But once the variance is inflated to 8 times the truth, the matrix approximation (3) performs worse than the diagonal approximation. Information is still significantly depleted relative to the truth. These results support the findings in [14] where Collard showed that under larger levels of observation noise, a 2-4 times error variance inflation retained the most information relative to the truth.

These features of Figures 5.8 and 5.9 can be qualitatively explained by considering the impact of error correlations on the influence of the observations and the background in the analysis. Including background error correlations decreases the uncertainty of the background because the prior entropy is reduced, and hence impacts the relative influence of the observations. Under these circumstances it is more important that the observations are assigned their correct weighting; potentially because they will have a greater influence on the analysis. This is done through inflating the error variances as in (3). From Figure 5.10 we see that for this experiment an inflation factor of approximately 2 produces the optimal value of SIC; while using an inflation factor of over 5 times will deplete the information relative to a simple diagonal approximation (2).

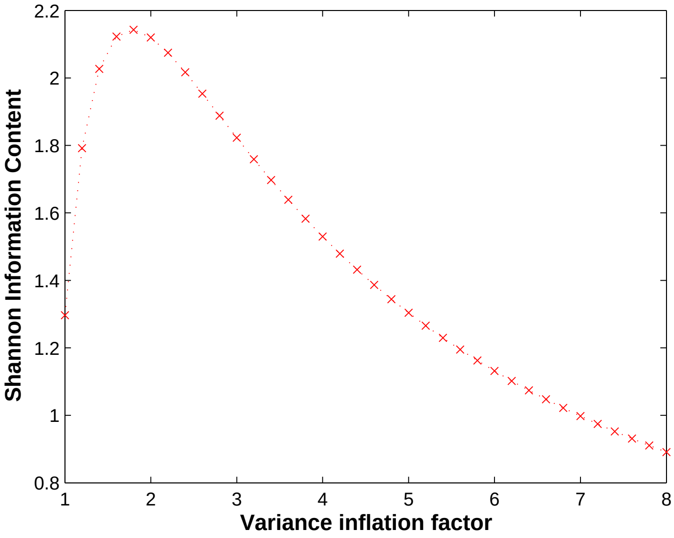


Figure 5.10: SIC for diagonal approximations (3) using different variance inflation factors.

For a correlated background error covariance matrix, we also examine the impact of using a circulant approximation to R_t (4). Figures 5.11 and 5.12 show the SIC and dof_S for a circulant and a diagonal approximation. There are two noticeable features of the plots. The first is the negative information for a 2×2 grid size, i.e, a 200km × 200km grid domain with 4 observations. This depletion in information suggests that using a circulant observation error correlation matrix is a very poor approximation to the truth for very small grid sizes. Using this correlation structure is detrimental to analysis accuracy. It is however unlikely that observation sets will be this small and we will therefore focus on larger grid domains (and therefore larger observation sets in our problem).

The second feature of the plots is the parallel linear increase in information with the number of observations (and hence domain size) relative to the truth. For a 4 × 4 grid upwards, the circulant approximation retains most of the information content relative to the truth (9.4340 dof_S compared to 9.9139 for a 10 × 10 grid). By examining the structure in Figure 5.6 we can explain this behaviour. The error covariance matrix R_t has a thin band of significant correlation centred around the diagonal. By construction the circulant approximation reflects the first row of R_t in roughly the central column. Most of the reflected values are zero or very close to zero; this is consistent with R_t. The exception will be the top right corner of the circulant matrix which will contain spurious non-zero values reflected from the top left corner of R_t. Therefore elementwise the circulant matrix will be a very good approximation to R_t.

The successful elementwise approximation is reflected in the difference between the matrices in the Frobenius norm (3.34). For a 10 × 10 grid the Frobenius norm of the difference between R_t and its circulant approximation is 16.65, compared to 143.44 for a diagonal approximation, and 188.63 for a diagonal approximation with twice the error variance. Importantly the difference in the respective inverse matrices is also small relative to the diagonal approximations. The difference in the Frobenius norm between R_t^{-1} and the inverse circulant matrix is only 0.47 compared to 4.45 for a diagonal approximation. We can conclude that for larger observation sets, a circulant approximation

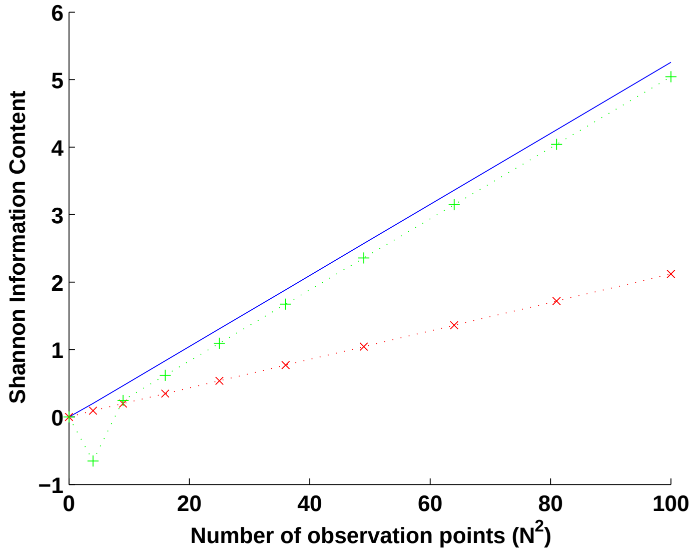


Figure 5.11: SIC for different grid sizes using R_t (blue line), diagonal approximation (2) (red x line) and a circulant approximation (green + line).

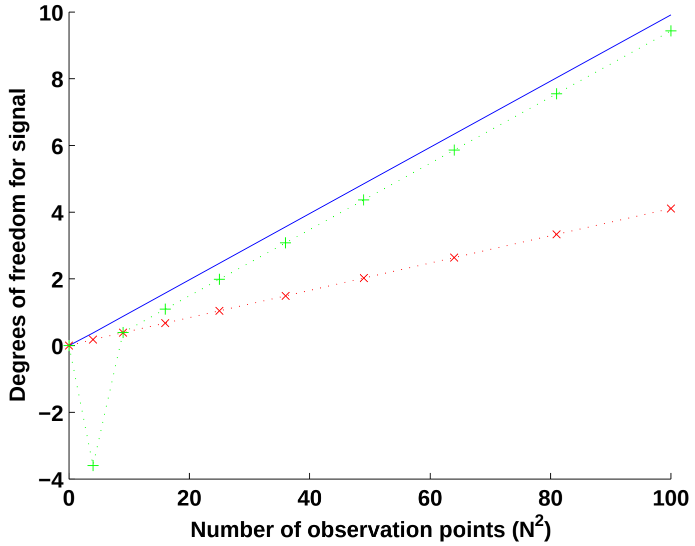


Figure 5.12: dof_S for different grid sizes using R_t (blue line), diagonal approximation (2) (red x line) and a circulant approximation (green + line).

performs very well in terms of information content.

5.4.3 Alternative analysis error covariance matrix

Finally we consider the impact of the formulation of the analysis error covariance matrix on the information content results. Previous results have used the approach where an additional term including the difference $R_t - R_f$ is added to S_a when the incorrect observation error covariance matrix was used. However, if the correct observation error covariance matrix is unknown then this approach cannot be used. The alternative is simply to treat the incorrect error covariance matrix as the truth, resulting in analysis error covariance matrix $S_a^{(3)}$ (5.12).

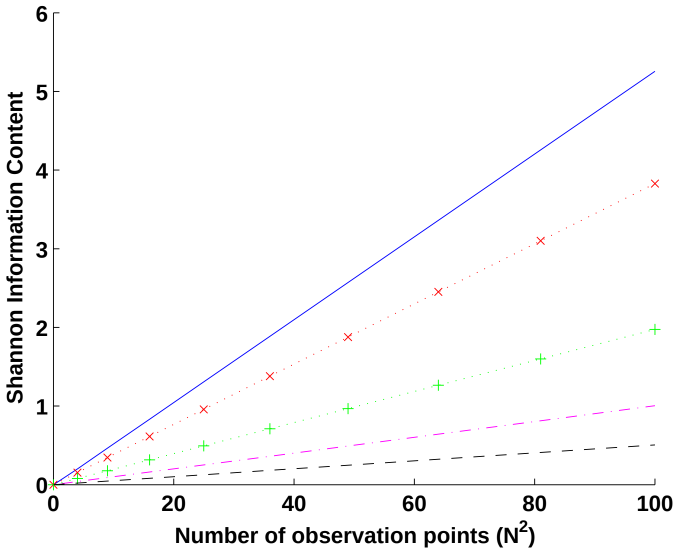


Figure 5.13: As for Figure 5.8 but using $S_a^{(3)}$.

Figures 5.13 and 5.14 show the SIC and dof_S, respectively, when R_t (1) and the diagonal approximations (2) and (3) are used, and the background errors have Gaussian distributed correlations. The results suggest that using a diagonal approximation (2) retains the most information, and is not overly detrimental to information content. Also, the results infer that inflating the error variance in a diagonal approximation (3) has an increasingly detrimental effect on the information content.

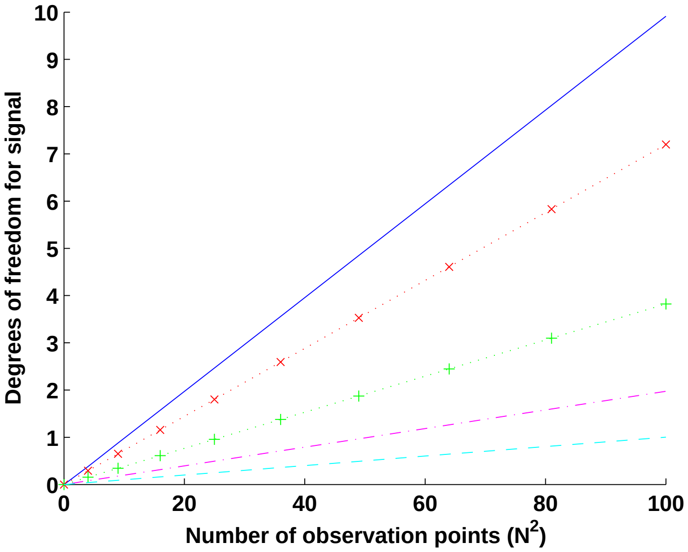


Figure 5.14: As with Figure 5.9 but usig $S_a^{(3)}$.

Approximation	dof_S using $S_a^{(2)}$	dof_S using $S_a^{(3)}$
Truth	9.9139	9.9139
Diagonal	2.2067	7.1970
2× Diagonal	4.1113	3.8226
4× Diagonal	2.9912	1.9736
8× Diagonal	1.7576	1.0033
ED (50 eigenpairs)	8.8670	10.3205
ED (25 eigenpairs)	5.2322	9.4934
ED (12 eigenpairs)	3.3039	8.3366

Table 5.2: dof_S for different matrix approximations using $S_a^{(2)}$ amd $S_a^{(3)}$.

Comparing these results to the previous approach in which the use of the incorrect observation error covariance matrix was incorporated, i.e, using $S_a^{(2)}$, we observe a difference in conclusions. When $S_a^{(3)}$ is used the information content is being mis-estimated using all approximations. Table 5.2 shows the dof_S when $S_a^{(2)}$ and $S_a^{(3)}$ are used for a 10×10 grid. The dof_S under the diagonal approximation is the largest overestimation (increased from 2.2067 to 7.1970), but the overestimation when using the ED approximation is more than the dof_S under the true observation error covariance matrix (10.3205 compared to the truth of 9.9139 when 50 eigenpairs are used). We can infer that using the incorrect observation error covariance matrix with no comparison to the truth can lead to misleading and inflated information content values, and subsequently incorrect conclusions. This highlights the importance of knowing the true error correlation structure so as to enable comparisons and the use of $S_a^{(2)}$.

5.4.4 Summary

Before any modifications are made to operational data assimilation frameworks, the proposed changes must be shown to demonstrate improvements. One measure of how effectively a data assimilation algorithm treats observation data is information content. The amount of information obtained from the observations can identify both wasteful and efficient data assimilation techniques.

We have evaluated the loss of information content under four different treatments of correlated observation errors. Experiments have been performed under independent and Gaussian correlated background errors. Information content was shown to be significantly degraded when approximating R_t with a diagonal matrix and ignoring error correlations. This implied a correlated approximation to modelling the errors was needed. One such approach was the approximation of R_t through its leading eigenpairs (5); this retained much of the information available even with fewer than half the available eigenpairs. But addressing Fisher's concerns [34], we found that spurious long range correlations were present even for larger observation sets.

A second approach to modelling correlation structure was using a circulant matrix approximation. The circulant matrix was shown to provide a good elementwise approximation to R_t, and retained nearly all the available information for a suitably large grid size. Both correlated approaches demonstrated the benefits of including some level of correlation structure.

Both the qualitative and quantitative information content results were shown to be sensitive to the specification of the background error correlations and the construction of the analysis error covariance matrix. The structure of B influenced the impact of a diagonal approximation with variance inflation. When $B = I$, a diagonal approximation retained more information than all inflated diagonal approximations; where as when B had a Gaussian correlation structure, a diagonal matrix with a 2-4 times variance inflation retained more information than a simple diagonal approximation.

There are three approaches to the construction of the analysis error covariance matrix. The first is for when R_t is used, and the latter two are for when an approximation R_f is used. The final two approaches either incorporate the additional error in using R_f or treat R_f as the truth. The final results of the section showed that treating R_f as the truth resulted in misleading and inflated information content values. This may however be the only approach if the true error correlation structure of the data is unknown.

5.5 Conclusions

The experiments presented in this chapter evaluated the loss in information content when ignoring error correlations, using simplified diagonal matrix structures, and using approximate correlation structures. Our new results showed that information content was severely degraded under the assumption of independent observation errors, but the retention of an approximated correlation structure gave clear benefits. In addition we examined the effect of background error correlation structure on the information content measures. We concluded that when background errors are correlated, it is more important that the observations have the correct weighting in the analysis, created by an appropriate correlation structure.

We began the chapter by describing the two measures of information content (SIC and dof_S) for different constructions of the analysis error covariance matrix. We then introduced diagnosed error correlations for a set of AMV observations [7]. By fitting a correlation function to empirical correlation data, the authors in [7] were able to quantify the spatial error correlations between AMV observations. The experiments pre-date the Desroziers' method used in quantifying cross-channel IASI error correlations in Chapter 4, and instead used a modified Hollingsworth-Lönnberg technique.

Different diagonal and correlated approximations to the previously diagnosed error covariance matrix were proposed in Section 5.3. The success of each of these was then evaluated in terms of the information content provided by a set of simulated observations. The results showed the importance of including some approximate correlation structure, with diagonal approximations retaining considerably less information than their correlated counterparts. A circulant matrix approximation was shown to be the approximation that retained the most information content for larger grid sizes.

The results also highlighted the sensitivity of information content to the background error correlations and the construction of the analysis error covariance matrix. If the approximate observation error covariance matrix is treated as the truth, then the information content can be overestimated even relative to the real truth. This reinforces the previous conclusion that it is important to know accurately the correct error correlation structure for an observation type, even if an approximation to this structure is to be made.

We have addressed the second of the thesis aims and evaluated several approximations available to model error correlation structure. We have quantified their impact on the data assimilation diagnostic of information content. However our model is a relatively simple test problem and we have used a simple 2D model framework and 3D-Var data assimilation scheme. Using the results in this chapter as motivation, we extend our research on correlated matrix approximation structures to a 4D-Var data assimilation scheme. In the initial assimilation experiments in the following chapters we take the proposed matrix approximations described in Chapter 3, and used in Chapter 5, and apply them in an incremental 4D-Var data assimilation algorithm of the type used at the Met Office. We can then address the final thesis aim, and determine the behaviour of these approximations in a 1D shallow water model data assimilation experiment.

Chapter 6

Modelling correlation structure in a 1D shallow water model

In NWP a set of governing equations is used to describe complex atmospheric and oceanic motions. However, new research ideas can be difficult and time consuming to implement directly into such a sophisticated framework. The Shallow Water Equations (SWEs) are often used as a test bed for atmospheric research, providing an intermediate step between conception and operational implementation. They have been shown capable of describing important aspects of the dynamic properties we wish to model, such as geostrophic motion in three-dimensions [71].

In the final chapters of the thesis we will study the behaviour of a data assimilation algorithm under different approximations to the observation error covariance matrix. We will use the SWEs as the model in the assimilation. In this chapter we introduce the SWEs and describe the data assimilation system applied to them. We focus on the implementation of two correlated approximations to the observation error covariance matrix.

We start by describing the continuous and discrete form of the SWEs; details on the discretisation technique are provided. Our attention is then focused on the data assimilation system of interest: incremental 4D-Var. The SWEs are one-dimensional so the incremental 4D-Var system becomes two-dimensional; we shall however use the terminology 4D-Var for the sake of convention. We discuss the practical issues surrounding the implementation of the algorithm, specifically generating the approximations to the observation error covariance matrix. This matrix is used in the cost function calculations of the 4D-Var algorithm, where matrix-vector products involving its inverse are required. We generate new equations used for calculating these matrix-vector products when the observation error covariance matrix is approximated with a Markov or an eigendecomposition (ED) matrix. These equations demonstrate a feasible method of incorporating error correlations in data assimilation algorithms. In the penultimate section we discuss various methods of determining convergence and solution accuracy. We conclude the chapter by describing the coding tests necessary to ensure the validity of the assumptions used in constructing the shallow water model.

6.1 Model framework

We begin the chapter by considering a one-dimensional shallow-water system describing the irrotational flow of a single-layer, inviscid fluid over an object. Although multi-dimensional shallow water models are available, the one-dimensional shallow water model (SWM) retains the key properties of the more detailed models whilst being significantly simpler to develop. In the work described here, the added time dimension leads to a two-dimensional problem. The one-dimensional model has previously been used to represent atmospheric phenomena such as air flow over mountains [48], and practical problems such as hydraulic flow in power plants [11]. A thorough description of inviscid multi-dimensional shallow water theory is given in [71].

6.1.1 The continuous analytical model

The continuous equations describing 1D shallow water flow are given in [54] by

$$\begin{aligned} \frac{Du}{Dt} + \frac{\partial \phi}{\partial x_D} &= -g\frac{\partial h_o}{\partial x_D}, & (6.1) \\ \frac{D(\ln \phi)}{Dt} + \frac{\partial u}{\partial x_D} &= 0, & (6.2) \end{aligned}$$

where

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

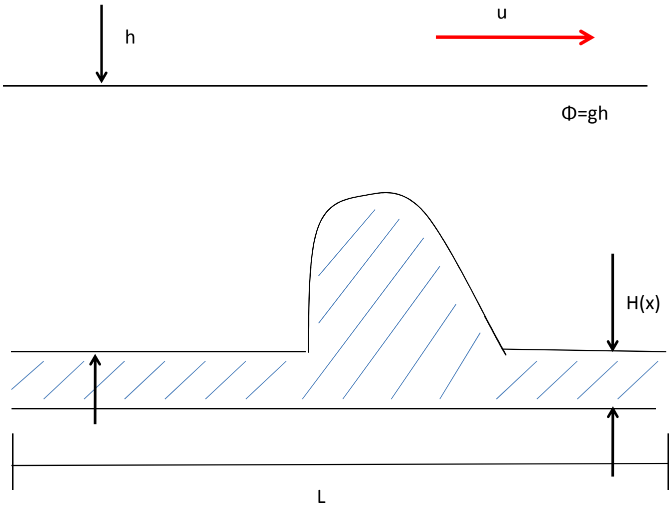
and $h_o = h_o(x_D)$ is the height of the bottom orography, u is the fluid velocity, $\phi = gh$ is the geopotential where g is the gravitational acceleration and $h > 0$ is the depth of fluid above the orography. The problem is defined on the domain $x_D \in [0, L]$. The spatial boundary conditions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

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

The fundamental parametric condition which characterises shallow-water theory is

$$\frac{h}{L} \ll 1;$$

hence the 'shallow' nature of the problem. A schematic of one-dimensional shallow water flow is shown in Figure 6.1.



90

Figure 6.1: Shallow water model schematic where H is the height of the bottom orography, u is the fluid velocity, $\phi = gh$ is the geopotential, g is the gravitational acceleration, h is the depth of the fluid above the orography and L is the length of the domain.

6.1.2 The numerical model

In many modern data assimilation systems a linearised version of the nonlinear forecasting model is required; for example in incremental 4D-Var to model the evolution of a perturbation. Therefore in order to use the SWEs as a model test-bed, their linearised form is needed. Two possible linearisation approaches are compared in [54]. The discrete method involves the linearisation of the discrete form of the nonlinear model to give the tangent linear model (TLM). The semi-continuous method involves the linearisation of the continuous nonlinear model equations and their subsequent discretisation to give the perturbation forecast model (PFM). In [52] Lawless et al showed that the linearisation techniques performed similarly well for noisy observations in a 1D SWM framework.

Therefore in this work we will employ the discrete method.

The nonlinear SWEs are discretised using a two-time-level semi-implicit, semi-Lagrangian scheme (SISL). The SISL scheme is chosen to match closely the numerical integration scheme used operationally at the Met Office [22]. In a Lagrangian scheme the advection in a shallow water system is studied by tracking the position of a set of water parcels. A set of originally regularly spaced parcels at one time step may evolve to be very close to each other at the next time step, and therefore some areas may be poorly resolved [84]. A semi-Lagrangian scheme tracks a different set of parcels at each time step; chosen so that their positions at the next time step (known as the arrival point) are at regularly spaced grid points. The point from which the parcel originates is known as the departure point. Figure 6.2 shows example departure and arrival points at two time levels.

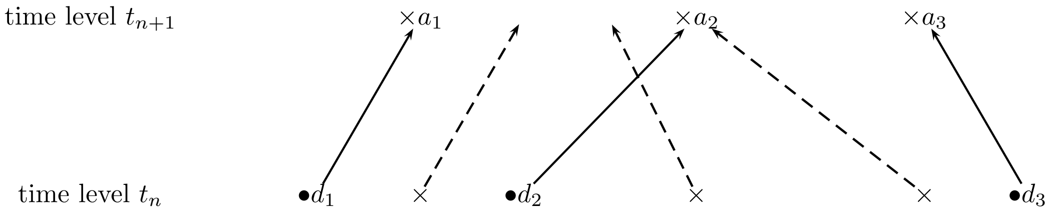


Figure 6.2: A semi-Lagrangian scheme with departure points (d_1, d_2, d_3) and arrival points (a_1, a_2, a_3). The paths taken by water parcels from the determined departure points are shown by the full lines, and the paths taken by water parcels from the regular grid points are shown by the dashed lines.

Applying the semi-Lagrangian method to the 1D SWEs, we denote a_u and d_u as the arrival and departure points for the u variable, respectively, and a_ϕ and d_ϕ similarly for the ϕ variable. The discretised form of the nonlinear model is given by

$$\frac{u_{a_u}^{n+1} - u_{d_u}^{n}}{\Delta t} + (1-\alpha_1)\left(\frac{\partial \phi}{\partial x_D} + g\frac{\partial h_o}{\partial x_D}\right)_{d_u}^{n} + \alpha_1\left(\frac{\partial \phi}{\partial x_D} + g\frac{\partial h_o}{\partial x_D}\right)_{a_u}^{n+1} = 0 \quad (6.3)$$

$$\frac{(\ln\phi)_{a_\phi}^{n+1} - (\ln\phi)_{d_\phi}^{n}}{\Delta t} + (1-\alpha_2)\frac{\partial u}{\partial x_D}\Big|_{d_\phi}^{n} + \alpha_2\frac{\partial u}{\partial x_D}\Big|_{a_\phi}^{n+1} = 0 \quad (6.4)$$

where the superscripts indicate the time level and coefficients α_1 and α_2 are time-weighting parameters chosen to meet the stability requirements of the scheme [50].

Equations (6.3) and (6.4) can be solved iteratively to derive the u and ϕ variables at each time level [54]. The TLM is given by the linearised version of (6.3) and (6.4); the result is a model describing how a perturbation to a solution state evolves over time under a linear approximation. By treating the TLM code as a sequence of linear operations, the adjoint model code can be derived directly. The details of this process are described in [12].

6.2 Data assimilation system

The data assimilation system in which we will run the SWEs is the incremental 4D-Var method described in Section 2.3.1. This method simplifies the minimisation problem posed in the full 4D-Var framework to a series of quadratic minimisations constrained by a linear model, and is used operationally by the Met Office and ECMWF. Because of the one-dimensional nature of the SWEs, the 4D-Var assimilation system is reduced to two-dimensions. Our experiments are performed using a modified Fortran 90 code for an irrotational shallow water model taken from the NERC Data Assimilation Research Centre website [51]. The code was originally designed to investigate incremental 4D-Var using non-tangent linear models, but modifications to the framework allow us to investigate the effect of different observation error correlation structures on the assimilation.

6.2.1 Correlated observation noise

The 1D SWM code is run using an identical twin experiment in which the model is assumed perfect. The identical twin experiment involves running the SWM forward in time from some initial conditions to generate a solution trajectory over a time window. From this solution, known as the truth trajectory, observations are sampled and perturbed by noise, if required. The observation set and a first-guess background field are then assimilated using an incremental 4D-Var algorithm; the resultant analysis propagated over the time window is known as the analysis trajectory. By comparing the analysis and truth trajectories, we can measure the success of the data assimilation method.

The original code in [51] was written such that the observations used in the assimilation were sampled from the truth trajectory and perturbed using random noise with variance σ^2. In order to compare the impact of using different error covariance matrix approximations to model correlated noise, we must first modify the distribution of the observation noise.

In our new code, samples of correlated noise are generated outside the main program. A MATLAB random number generator *mvnrnd()* [65] is run and returns random vectors chosen from a multivariate normal distribution with mean zero and covariance matrix R. The specification of the covariance matrix is dependent on the proposed correlation structure of the observation errors. We assume that the errors in the u and ϕ observations are mutually uncorrelated and so the observation error vectors for the u and ϕ field are generated independently. The observations in the main program are now sampled from the truth trajectory and then perturbed using the externally generated error vectors. This allows the user to determine the level of correlation in the observation errors.

The one-dimensional construction of the shallow water model 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.

6.2.2 Background error covariance matrix

The noise used to perturb the background trajectory is created using the same method used to generate the observation errors described in Section 6.2.1. We treat the background errors as uncorrelated, and so the covariance matrix used to generate the background noise will be a diagonal matrix comprised of the error variances. The background error variances are set as half those of the observation errors. This matrix will be used in the assimilation as the background error covariance matrix.

6.3 Observation error covariance matrices

In order to assess the impact of modelling correlated observation error structure in the SWM, we run the identical twin experiments using different approximations to the true error covariance matrix used in the generation of the correlated observation errors. 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 observation error covariance matrix is used explicitly in the calculation of the incremental 4D-Var cost function (2.15) and its gradient (2.16), needed for the inner loop minimisation algorithm. The contributions to the cost function and gradient values are calculated separately for the u and ϕ fields because of the assumption of mutually independent errors.

The inherited code contained three representation options for the observation error covariance matrix: a zero matrix (i.e, ignoring the term), the identity matrix, and a diagonal matrix of the true error variances. These options were sufficient under the previous specification of uncorrelated observation noise, but when correlated observation errors are present, we need a more sophisticated approximation to the error correlation structure. Below we describe the new implementation of two proposed correlated approximations to an observation error correlation matrix: a circulant matrix and an eigendecomposition (ED) matrix. Both approximations have previously been considered in the 3D-Var investigations in Chapter 5.

6.3.1 Markov matrix

The Toeplitz matrix approximation we use is the Markov error covariance matrix described in Section 3.2.4. A convenient feature of this matrix is its tri-diagonal inverse, which is known explicitly (3.9). Using the specification of the Markov inverse and the assumption that all observation error variances are the same in each field at each point, we can find an explicit form for the matrix-vector products used in the calculation of the cost function (2.15) and its gradient (2.16):

$$(R_M^{-1}\mathbf{x})_i = \begin{cases} \frac{1}{\sigma^2(1-\rho^2)}[x_i - \rho x_{i+1}] & i = 1 \\ \frac{1}{\sigma^2(1-\rho^2)}[-\rho x_{i-1} + (1+\rho^2)x_i - \rho x_{i+1}] & i = 2, \dots, N-1 \\ \frac{1}{\sigma^2(1-\rho^2)}[-\rho x_{i-1} + x_i] & i = N \end{cases} \tag{6.5}$$

and

$$\mathbf{x}^T R_M^{-1}\mathbf{x} = \frac{1}{\sigma^2(1-\rho^2)}\left[(1+\rho^2)\sum_{i=1}^{N} x_i^2 - \rho^2(x_1^2 + x_N^2) - 2\rho\sum_{i=1}^{N-1} x_i x_{i+1}\right]$$

where R_M^{-1} is the Markov matrix inverse, $\mathbf{x}$ is the incremental innovation vector, σ^2 is the observation error variance associated with the field, N is the number of observations, and ρ is the correlation level. The correlation level is given by

$$\rho = \exp\left(\frac{-\Delta x}{L_R}\right),$$

where Δx is the spatial separation and L_R is the correlation length scale. A technical note is that the second of these expressions is not used explicitly in the code, because once $R_M^{-1}\mathbf{x}$ is calculated, we need only calculate the dot product of this vector and $\mathbf{x}$ to find $\mathbf{x}^T R_M^{-1}\mathbf{x}$.

One of the aims of the thesis is to demonstrate how the inclusion of observation error correlation structure is feasible in operational data assimilation algorithms. Equation (6.5) has demonstrated an inexpensive approach to modelling error correlation structure; the number of operations used in calculating $R_M^{-1}\mathbf{x}$ is the same order of magnitude ($O(n)$) as that if the observation error covariance matrix was diagonal. In Chapter 7 we will test how effective this proposed method is at modelling error correlation structure.

6.3.2 Eigendecomposition (ED) matrix

A truncated eigendecomposition of the true observation error correlation matrix can also be used to model error correlation structure. In Chapter 5, this approximation was shown to retain a significant amount of information content relative to a diagonal approximation. In Chapter 7 we will investigate the effect of this approximation and the previously proposed Markov matrix approximation on the accuracy of an assimilation analysis.

The construction of the ED matrix and its inverse were given in Section 3.3. In our new code, the leading eigenpairs needed for the representation are pre-computed using the MATLAB function *eigs()* [65] which decomposes the true error correlation matrix into its leading eigenvalues and eigenvectors. If the option to model the correlation structure using an ED matrix is chosen, then the leading K (specified by the user) eigenpairs are read in and stored for use in the main program.

Using equation (3.11) from Chapter 3 and (5.18) from Chapter 5, we can calculate the value of α and use the leading K eigenpairs $(\lambda_k, \mathbf{v}_k), k = 1, \ldots, K$ to implicitly represent the inverse error covariance matrix. Again assuming that all the observation error variances are the same in each field at each point, we have an explicit form for the matrix vector products $R_E^{-1}\mathbf{x}$ and $\mathbf{x}^T R_E^{-1}\mathbf{x}$ needed for the calculation of the cost function and its gradient:

$$\begin{aligned}
(R_E^{-1}\mathbf{x})_i &= \left(\alpha^{-1}D^{-1}\mathbf{x} + D^{-1/2}\sum_{k=1}^{K}(\lambda_k^{-1} - \alpha^{-1})\mathbf{v}_k\mathbf{v}_k^T D^{-1/2}\mathbf{x}\right)_i \\
&= \left(\frac{1}{\alpha\sigma^2}\mathbf{x} + \frac{1}{\sigma^2}\sum_{k=1}^{K}(\lambda_k^{-1} - \alpha^{-1})\mathbf{v}_k s_k\right)_i \\
&= \frac{1}{\alpha\sigma^2}x_i + \frac{1}{\sigma^2}\sum_{k=1}^{K}(\lambda_k^{-1} - \alpha^{-1})v_{ik}s_k \qquad (6.6)
\end{aligned}$$

$$\mathbf{x}^T R_E^{-1}\mathbf{x} = \frac{1}{\alpha\sigma^2}\mathbf{x}^T\mathbf{x} + \frac{1}{\sigma^2}\sum_{k=1}^{K}(\lambda^{-1} - \alpha^{-1})s_k^2$$

where R_E^{-1} is the ED matrix inverse, $\mathbf{x}$ is the incremental innovation vector, $D = \sigma^2 I$ is the diagonal matrix of the error variances, $s_k = \mathbf{v}_k^T\mathbf{x}$ is the dot product of $\mathbf{v}_k$ and $\mathbf{x}$, and v_{ik} is the i^{th} component of the k^{th} eigenvector. As with the Markov matrix, the expression for $\mathbf{x}^T R_E^{-1}\mathbf{x}$ is unnecessary if $R_E^{-1}\mathbf{x}$ has already been calculated.

We can choose K to be small in order to reduce our storage costs, but the length of the eigenvector, N, is still likely to be large. In order to reduce the number of operations performed we calculate the inverse values of α, σ^2 and λ_k prior to the calculation. We also use the inbuilt Fortran 90 function DOT_PRODUCT() to calculate $\mathbf{x}^T\mathbf{x}$ and s_k, and calculate s_k outside the sum since it only needs calculating once for each eigenpair $(\lambda_k, \mathbf{v}_k)$.

Again the above equations demonstrate a feasible method of incorporating error correlation structure in an operational data assimilation algorithm. We now look at the conditions necessary for the successful treatment of the two matrices described above.

6.4 Convergence and solution accuracy

To produce the analysis trajectory most consistent with the available data over a particular time window, the incremental 4D-Var algorithm must ensure the cost function has been solved to sufficient accuracy. A good approximation to the error covariance matrices will ensure that the cost function is accurately specified, but in order to obtain a desired accuracy of solution, the inner and outer loop minimisations of the algorithm must be run until some specified tolerance level is reached. The tolerance level will influence the cost and exactness of the solution; if the tolerance level is set too small then excessive iterations will be performed, and if the tolerance level is too large then the minimisation may not be solved precisely enough.

The inner loop of the assimilation algorithm is responsible for minimising a series of quadratic cost functions constrained by a linear model; the minimisation algorithm used in our experiments is the conjugate gradient method [36]. Because the minimisation is essentially an approximation to the full nonlinear cost function, we need not solve it too accurately, but it needs to have converged sufficiently [53]. Several stopping criteria options exist for the conjugate gradient minimisation, and a good review of their use in a similar model framework is given in [53].

In these experiments we will use the relative change in gradient stopping criteria favoured in [53]. This requires that the inner loop minimisation is terminated when

$$\frac{\left\|\nabla J_m^{(k)}\right\|_2}{\left\|\nabla J_0^{(k)}\right\|_2} < \epsilon_I, \tag{6.7}$$

where the subscripts indicate the inner loop iteration index, k indicates the outer loop iteration index, and ϵ_I is the user set tolerance. In other words, the solution is assumed to have converged when the ratio of the 2-norm of the inner loop gradient after m iterations and at the start of the outer loop is less than a certain tolerance.

The outer loop of the assimilation algorithm is responsible for updating the linear model trajectory. In practical data assimilation, the outer loops are not usually run to complete convergence and only a few are performed. However, if we are to examine the impact of different approximations to the observation error covariance matrix, we need the same level of convergence to be obtained under each approximation, so as to draw consistent conclusions. Therefore we use enough outer loops so that some convergence criterion is satisfied. The convergence criterion we use is the relative change in function

$$\frac{|J^{(k+1)} - J^{(k)}|}{1 + |J^{(k)}|} < \epsilon_o, \tag{6.8}$$

where the superscripts indicate the outer loop iteration index and ϵ_o is the user set tolerance. This is one of the proposed criterion in [53].

When the tolerance levels for the inner and outer loop convergence criteria are achieved we can be sure that the solution to the minimisation problem has converged to some level of accuracy. However we do not know how close the computed solution is to the 'true' solution of the problem. By testing the gradient $\nabla J^{(k)}$ at the converged solution of the k^{th} outer loop, we can determine the limiting accuracy in this solution [39]. The best numerical accuracy we would expect in the converged solution $\overline{x}$ is the same order as the normalised gradient $\frac{\|\nabla J^{(k)}(\overline{x})\|_2}{|J^{(k)}(\overline{x})|}$. Experiments showed that as we decrease the outer tolerance ϵ_o, the solution accuracy increases. Setting $\epsilon_o = 0.01$, we computed the normalised gradient and found it to be of order 10^{-2}; from [39], this is our expected converged solution accuracy. We confirmed this accuracy by comparing the converged solutions using $\epsilon_o = 0.01$, $\epsilon_o = 0.001$ and $\epsilon_o = 0.0001$. The converged solution was found to be accurate to approximately two decimal places when using $\epsilon_o = 0.01$, as expected.

6.5 Coding tests

To conclude this chapter we will describe the pre-assimilation tests necessary to ensure the validity of the model assumptions. Both the TLM and its adjoint are constructed under the assumption of linearity. It is important to test the robustness of this assumption prior to their use in the minimisation algorithm.

We can verify the TLM is coded correctly by performing two tests: the correctness test and the validity test. The correctness test checks if the evolution of a perturbation in the TLM is comparable with that of the same perturbation in the nonlinear model [73]. The validity test identifies the time window for which the assumption of linearity in the TLM is valid. Details on the successful application of these tests to the TLM in the inherited SWM code are given in [54].

We can then test the adjoint model by ensuring that the cost function gradient computed in the assimilation is realistic; this is done using the gradient test [58]. The gradient test is based on a Taylor series expansion of the cost function,

$$J(x_0 + \alpha\delta x_0) = J(x_0) + \alpha\delta x_0^T \nabla J(x_0) + O(\alpha^2)$$

which rearranged gives

$$\chi(\alpha) \equiv \frac{J(x_0 + \alpha\delta x_0) - J(x_0)}{\alpha\delta x_0^T \nabla J(x_0)} = 1 + O(\alpha) \tag{6.9}$$

where α is a small scalar and $\delta x_0 = \frac{\nabla J}{\|\nabla J\|}$ is a unit vector in the gradient direction. If the adjoint code is working correctly and the cost function and its gradient are well calculated, results will show $\chi(\alpha)$ approaching 1 as α decreases to 0. An exception will be when α is very close to machine accuracy. In Chapter 7 we perform the gradient test under different approximations to the observation error covariance matrix to ensure the cost function gradient is calculated correctly.

6.6 Summary

In this chapter we have described the framework of a one-dimensional shallow water model and the practicalities of its use in an incremental 4D-Var data assimilation algorithm. We started by considering the continuous and discrete form of the SWEs, and explained how the TLM and adjoint code might be derived. We then considered the use of the SWM in an incremental 4D-Var data assimilation algorithm; because of the one-dimensional nature of the SWM, the incremental 4D-Var algorithm becomes two-dimensional. We focused on the specification of the observation error covariance matrix and its approximations. A method for generating correlated observation error noise was described, and its application to background errors was indicated.

In Section 6.3 we discussed the role of the observation error covariance matrix in the proposed data assimilation algorithm, and provided two possible correlated approximations. We found that by their construction, both of these approximations allowed for the inclusion of observation error correlation structure without excessive computation cost. Equations specifying their use in the data assimilation algorithm were given.

We then described how the use of different error covariance matrices could impact on convergence and solution accuracy in the data assimilation algorithm. Convergence criteria for the inner and outer loop of the cost function minimisation were given. Also the expected level of solution accuracy was determined. Finally we described several coding tests used to test the validity of the model assumptions.

In the next chapter we take the theory described in this chapter, and apply it to several assimilation experiments. The focus will be on determining how well the matrix approximations described in Section 6.3 perform relative to the assumption of a diagonal error covariance matrix.

Chapter 7

Shallow water equations statistical tests

In the previous section we described how an incremental 4D-Var assimilation using 1D-SWEs could be extended to include correlated observation errors. A new approach to modelling the observation error correlation structure was required. The two correlated error covariance matrix representations given in Section 6.3 are now tested against diagonal approximations in the modified assimilation system. The impact of each approximation on the analysis error in the assimilation is examined. The aim of the experiments in this penultimate chapter is to address the final thesis question posed in Chapter 1: how well do approximations to error correlation structure perform in a data assimilation experiment? For the purpose of this chapter we decompose this into three separate questions:

- Is it better to model observation error correlation structure incorrectly than not at all?

- Which matrix approximation is the most robust to changes in the true error correlation distribution?
- Can we identify a suitable matrix approximation even when the observation error correlation structure is unknown?

The new results presented in this chapter will address these three questions.

The chapter is structured as follows. We begin by outlining the data assimilation experiments performed, including details on the model set up, the simulated error correlation structures, and the analysis error diagnostics. We then describe three separate experiments performed to address the questions posed above. The results shown are new and original. The first experiment tests the performance of different matrix approximations in modelling a Markov error correlation structure. We provide details on the choice of the matrix approximations and ensure they have been suitably coded using the model tests described in Section 6.5. The results suggest that a Markov error correlation structure is better modelled using a Markov matrix with a mis-estimated length scale rather than a diagonal matrix.

In the second experiment we investigate the robustness of the matrix approximations in modelling correlated observation error. This is done by replicating the first experiment but instead using a SOAR error correlation structure [3] in the true error covariance matrix. Results show that a Markov matrix is the most robust approximation, but that an ED approximation even with a small number of eigenpairs is an improvement on using a diagonal approximation. We then extend these experiments and investigate the impact of the size of the observation error variances on the conclusions made.

Finally in the third experiment we treat the true error correlation structure as unknown, and use the diagnostics proposed in Chapter 4 to diagnose both the true error correlation structure, and an approximate Markov correlation structure. This tests the robustness of our approximations when we are unsure of the error correlation structure being modelled. The results are encouraging, and show that the diagnostic can be used successfully to replicate the true error correlation structure, and diagnose a suitable Markov approximation. We conclude that a Markov matrix is the most robust approximation, but including some form of correlation structure is preferable to none at all.

7.1 Experimental methodology

The experiments performed in this chapter model a flow field described in [48] in which shallow water motion is forced by some orography. Using the SWEs (6.1)-(6.2) described in Chapter 6, 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 [54]. 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 [48]. We consider a 1D 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 on which the assimilation is run 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 $g = 10\text{ms}^{-2}$. The time-weighting parameters for the numerical scheme (6.3) and (6.4) are set as $\alpha_1 = \alpha_2 = 0.6$ to satisfy the stability conditions.

An identical twin experiment is performed by running the nonlinear shallow water model forward in time from the initial conditions, to generate a true model solution at each assimilation timestep (known as the truth trajectory). Observations of fluid velocity u and geopotential ϕ are sampled from the truth trajectory, and random noise with multivariate normal distribution $N(\mathbf{0}, R_t)$ is added as error. It is assumed 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. The initial background is taken as the truth trajectory plus random noise from a multivariate normal distribution $N(\mathbf{0}, B_t)$, where B_t is a diagonal matrix of the background error variances. An incremental 4D-Var data assimilation algorithm is then run using this observation and background data, and an analysis trajectory for the fluid potential u and geopotential ϕ is generated for each spatial and temporal step.

The maximum number of outer and inner loops performed are 20 and 200, respectively.

The outer and inner loops are terminated once the relative change in function (6.8) and the relative gradient stopping criteria (6.7) have been achieved with tolerance of 0.01 and 0.1, respectively.

7.1.1 Simulated error correlation structure

To answer the questions posed at the start of the chapter, it is important that we run the assimilation under different realisations of the true error structure, so as to draw robust conclusions. We perform three experiments examining the impact of modelling correlation structure by the proposed error covariance matrix representations discussed in Section 6.3: a diagonal approximation, a Markov approximation, and an ED approximation.

In the first experiment we compare the proposed approximate error correlation matrices against a true error correlation matrix with a Markov distribution, C_M, given by

$$C_M(i,j) = \exp\left\{ \frac{-|i-j|\Delta x_D}{L_R} \right\} \tag{7.1}$$

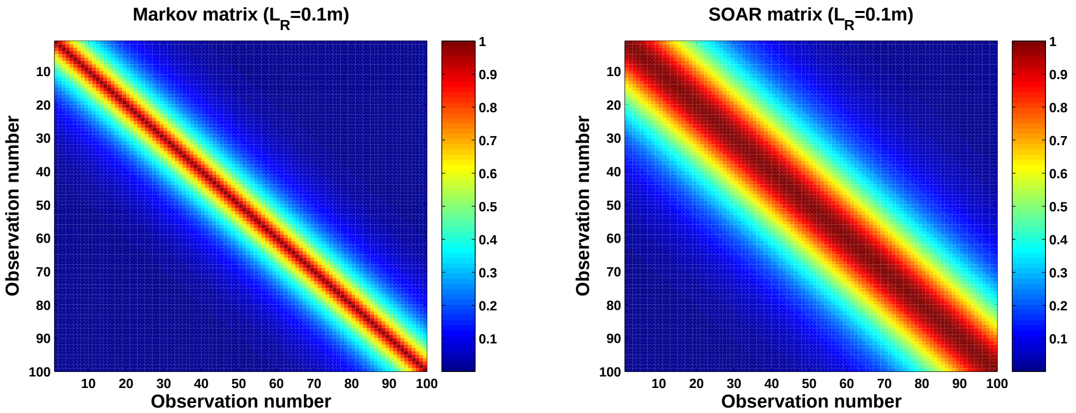
where $\Delta x_D = 0.01$m is the spatial separation and $L_R = 0.1$m is the length scale. The Markov matrix [78] is also an option for a matrix approximation (Section 7.2). Figure 7.1(a) shows a Markov error correlation matrix with length scale $L_R = 0.1$m.

The second experiment is an extension of the first, in which the true error correlation structure follows a SOAR (second-order autoregressive) distribution rather than a Markov one. The SOAR error covariance matrix is given by

$$C_S(i,j) = \left(1 + \frac{|i-j|\Delta x_D}{L_R}\right) \exp\left\{ \frac{-|i-j|\Delta x_D}{L_R} \right\}. \tag{7.2}$$

The SOAR matrix is an idealised correlation structure that is used at the Met Office to model background error correlation structures in the horizontal [3]. It is often used in preference to a Gaussian structure because its distribution has longer tails, better at matching empirical estimates. Figure 7.1(b) shows a SOAR error correlation matrix with length scale $L_R = 0.1$m. Comparing this to Figure 7.1(a), we see that the SOAR correlation matrix has a wider band of non zero correlations.

Finally in the third experiment we consider the case when the true error correlation structure is unknown. Using the Desroziers' method [25] that was successfully applied in Chapter 4, we aim to diagnose the true correlation structure and an optimal Markov approximating structure. All three experiments are run with uncorrelated background errors where the background error variances are half those of the observation errors, i.e, $B_t = \frac{1}{2}\text{diag}(R_t)$.



(a)

(b)

Figure 7.1: Observation error correlation structure of (a) a 100 × 100 Markov matrix with length scale $L_R = 0.1$m and (b) a 100 × 100 SOAR matrix with length scale $L_R = 0.1$m. Note that we plot a 100 × 100 matrix as opposed to the 1001 × 1001 matrix used in the problem because the intrinsic correlation structures are more clearly shown.

7.1.2 Analysis error

We now describe the retrieval properties 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

$$\left\|\overline{x}_{R_f} - x^*\right\|_2 \tag{7.3}$$

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 using imperfect observations when the approximation R_f is used;

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

$$\frac{\left\|\overline{x}_{R_f} - \overline{x}_{R_t}\right\|_2}{\left\|\overline{x}_{R_t}\right\|_2} \times 100 \tag{7.4}$$

where $\overline{x}_{R_t}$ is the true converged solution to the assimilation problem using imperfect observations when the true error covariance matrix R_t is used.

Errors E1 and E2 provide us with information on the closeness of different analyses at the beginning and the end of the assimilation time window, and are calculated at the end of each outer loop. Since the magnitude of the ϕ field is an order larger than that of the u field, we produce separate error norms (7.3) and (7.4) for u and ϕ to avoid changes in the u field being overshadowed by changes in the ϕ field. Following the discussion of solution accuracy in Section 6.4, we can expect the accuracy in the converged solution error E1 to be two decimal places. The error E2 is constructed by the difference of converged solutions and therefore we expect this also to be accurate to approximately two decimal places.

7.2 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. First we will give some motivation for the different realisations of the matrix approximations used, and demonstrate their correct coding in the algorithm. The retrieval properties described in Section 7.1.2 are then calculated for each matrix approximation.

7.2.1 Matrix representations

Many different realisations of the proposed matrix approximations could be used to model the simulated error correlation structure. The choices we use and the motivation for them are given in this section. Firstly the diagonal matrix representations will be 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 information content results in Chapter 5 and from the results given in [14]. These showed that a 2-4 times variance inflation was preferable to a simple diagonal approximation when observation and background error correlations were present; but under correlated observation errors and uncorrelated background errors, a simple diagonal approximation performed better.

The Markov matrix representations will be 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. Markov error covariance matrices generated using these length scales are plotted in Figure 7.1(a) and Figures 7.2 - 7.4. Note that as the length scale decreases, the thickness of the central correlation band decreases. This is also demonstrated in Figure 7.5, where the central row of each Markov matrix is plotted. We also test the Markov matrix representation for when L_R is small enough so 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.

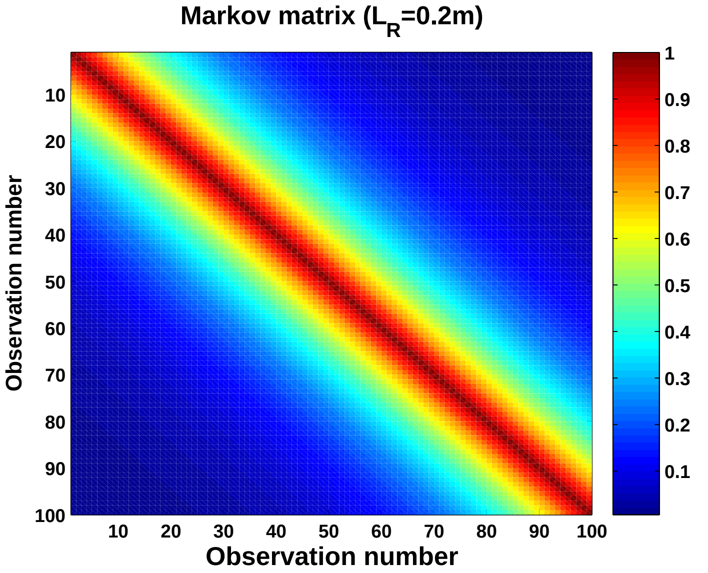


Figure 7.2: Observation error correlation structure of a 100 × 100 Markov matrix with length scale $L_R = 0.2$m.

The ED matrix representations will be truncated eigendecompositions of the true error correlation matrix, using a reduced number of eigenpairs. The formula used in their calculation is given in Chapter 3 (3.10) and the specific equations in Chapter 6 (6.6). 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 with length scale $L_R = 0.1$m, are plotted in Figure 7.6. The plots show that the eigenvalue size declines sharply as the eigenvalue number increases. After 100 eigenvalues, the eigenvalue size is less than two for the Markov matrix, and less than one for the SOAR matrix, and 80% and 99% of the overall un-

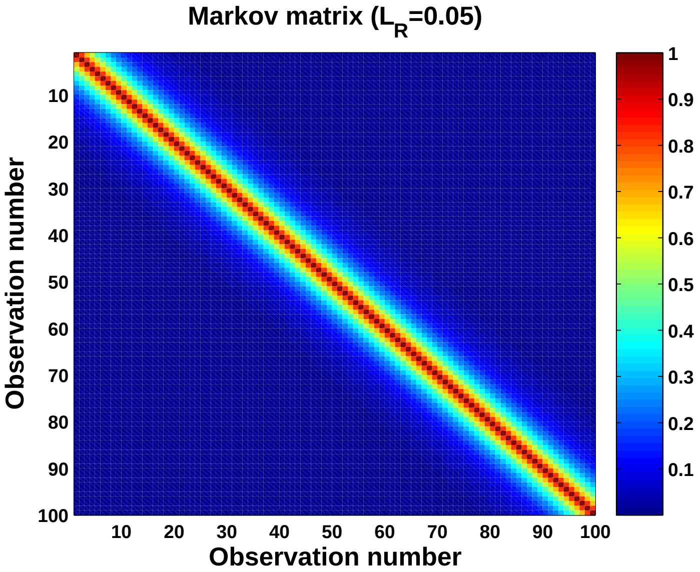


Figure 7.3: Observation error correlation structure of a 100 × 100 Markov matrix with length scale $L_R = 0.05$m.

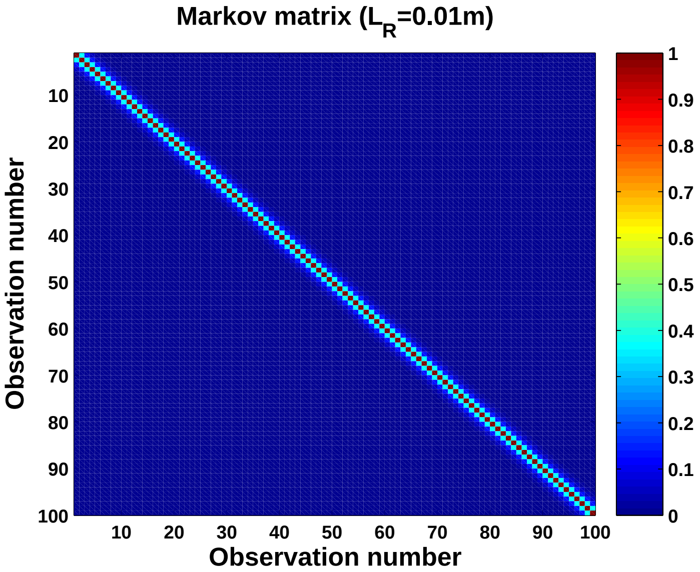


Figure 7.4: Observation error correlation structure of a 100 × 100 Markov matrix with length scale $L_R = 0.01$m.

certainty is represented, respectively. Therefore we use 100 eigenpairs as an empirical upper limit to the number of eigenpairs used in the assimilation.

The number of eigenpairs we will 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.

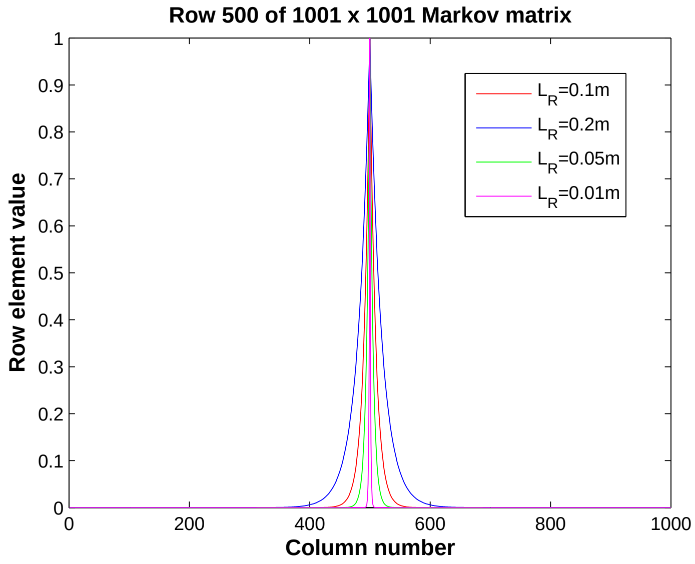
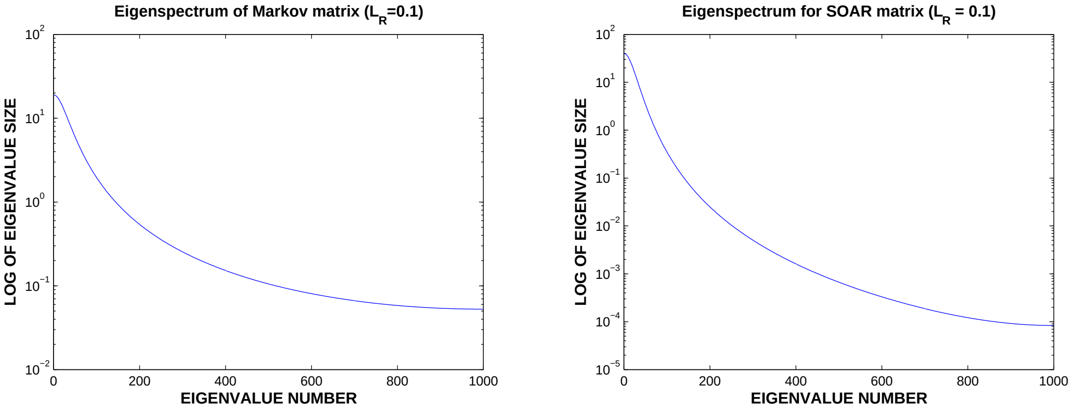


Figure 7.5: Middle row of a 1001 × 1001 Markov matrix



(a) (b)

Figure 7.6: (a)Eigenspectrum of a 1001 × 1001 Markov error correlation matrix; (b)Eigenspectrum of a 1001 × 1001 SOAR error correlation matrix

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, i.e, error E2 in Section 7.1.2. We therefore also run the assimilation using the ED approximation with the full number of eigenpairs.

7.2.2 Model tests

In Chapter 6 we described several tests used to ensure the validity of the model. In the experiments performed in this chapter we modify the code used in the calculation of the cost function and its gradient to allow for different approximations to the error covariance matrix. Therefore in order to ensure the true gradient of the cost function is being calculated by the modified adjoint code, we perform the gradient test described in Section 6.2.5 under different specifications of the observation error covariance matrix. In Figure 8.4 we plot $\chi(\alpha)$ versus α and $\log(|\chi(\alpha)-1|)$ versus α, where χ is defined by (6.9), for the case when a Markov approximation with length scale $L_R = 0.1$m to the Markov error covariance matrix is used. These figures compare well with those illustrated in [58] and therefore we can conclude that our adjoint model does provide the true gradient for the tangent linear model. Additional plots for other matrix approximations can be found in Appendix C.

7.2.3 Numerical results

We now present the analysis error diagnostics generated from the assimilation of different observation error covariance structures. 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 7.1 and 7.2.

We can see that in all cases the approximation results in an improvement to the background field: $\left\|x^b - x^*\right\|_2 = 0.32$ for the u field and $\left\|x^b - x^*\right\|_2 = 6.32$ for the ϕ field. Us-

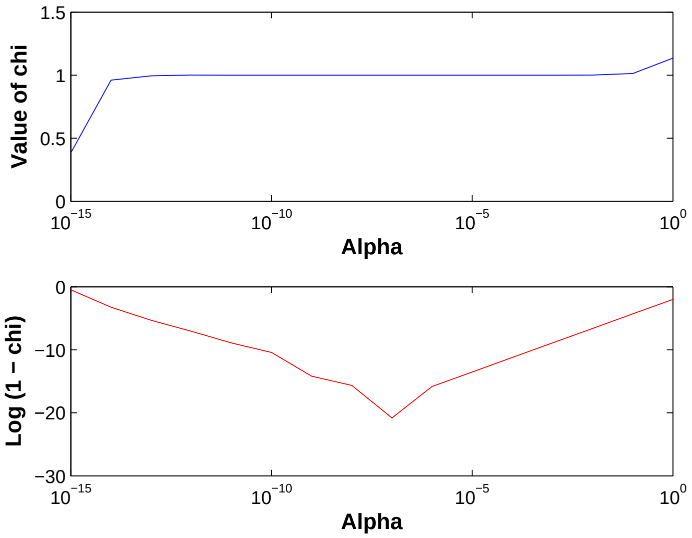


Figure 7.7: Gradient test for a Markov approximation to a Markov error covariance matrix

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

ing 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

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

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 densities. We found that using more 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 number of observations had the biggest impact on the diagonal approximation. Even when there was an observation 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.

7.2.4 Summary

We have investigated the impact of approximating a Markov error covariance matrix with diagonal, Markov, and ED matrices. We motivated our choices for the approximating structures and showed that the adjoint model was coded properly for their inclusion. The conclusions from these initial data assimilation experiments using a Markov observation error correlation structure can be summarised as:

- All approximations improve on the background field but a Markov approximation is the cheapest and most effective;
- A Markov approximation is robust under choice of length scale, even using a very short length scale is an improvement on a diagonal approximation;
- It is often better to use some correlation structure than none at all.

In the next sections we will extend the experiments performed here to different realisations of the true error correlation structure. We are interested to see if similar conclusions will be drawn.

7.3 Experiment 2: SOAR error correlation structure

In this section we consider the effect of our choice of the true observation error correlation structure. In Section 7.2 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 Section 7.2. 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.

7.3.1 Model tests

We must first ensure that the model is valid under the assumptions we are using. We follow the same procedure described in Section 7.2.2 and apply the gradient test under different approximations to the true error covariance matrix. Figure 7.8 shows that the adjoint model generates the true gradient when a diagonal approximation is used in the assimilation. Additional plots for other matrix approximations can be found in Appendix C.

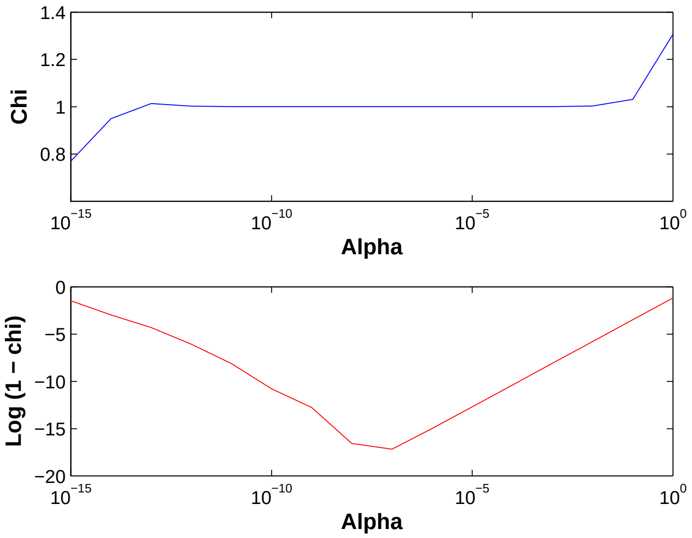


Figure 7.8: Gradient test for a diagonal approximation to a SOAR error covariance matrix

7.3.2 Numerical results

The analysis errors E1 and E2 at $t = 0$ for the different approximations to the SOAR error covariance matrix are given in Tables 7.3 and 7.4. Comparing the results to Table 7.1 and 7.2, 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 wider spread of correlations in a SOAR matrix (Figure 7.1). 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× 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 Chapter 5, when the background errors were uncorrelated.

Approximation	E1: $\left\lVert \overline{x}_{R_f} - x^* \right\rVert_2$	$\left\lVert \overline{x}_{R_f} - \overline{x}_{R_t} \right\rVert_2$	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 7.3: Analysis errors in u field at $t = 0$ for different approximations to a SOAR error covariance matrix ($\lVert \overline{x}_R \rVert_2 = 3.19$)

Approximation	E1: $\left\lVert \overline{x}_{R_f} - x^* \right\rVert_2$	$\left\lVert \overline{x}_{R_f} - \overline{x}_{R_t} \right\rVert_2$	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 7.4: Analysis errors in ϕ field at $t = 0$ for different diagonal approximations to a SOAR error covariance matrix ($\lVert \overline{x}_R \rVert_2 = 62.54$)

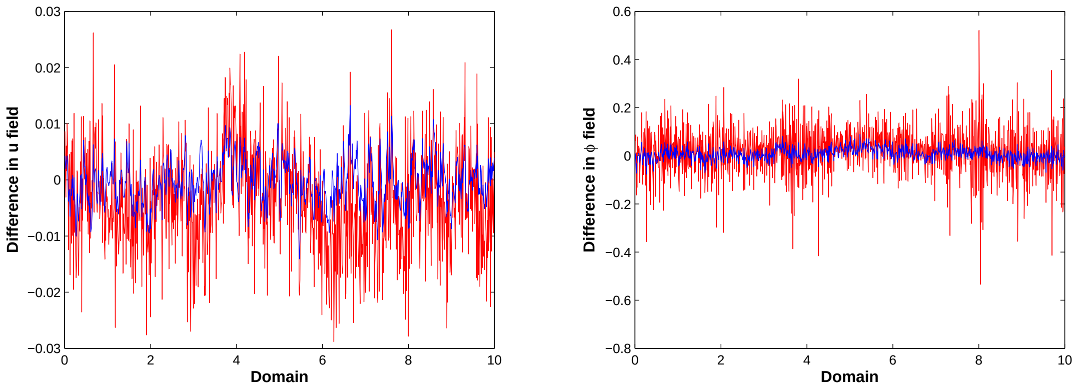
It is also expected that an ED matrix using 100 eigenpairs results in a very small analysis error relative to the converged solution, because as we observed in Section 7.2.2, 100 eigenpairs represent 99% of the overall uncertainty in the matrix. It is encouraging that an ED approximation using even fewer 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 Section 7.2. 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 7.6.

In conclusion, the results when assimilating different matrix approximations to a SOAR error covariance matrix have

- demonstrated the robustness of a Markov matrix as a desirable approximation to modelling observation error correlation structure, but a larger length scale is needed;
- shown that an ED approximation with as few as 50 eigenpairs is an improvement on ignoring observation error correlations entirely.

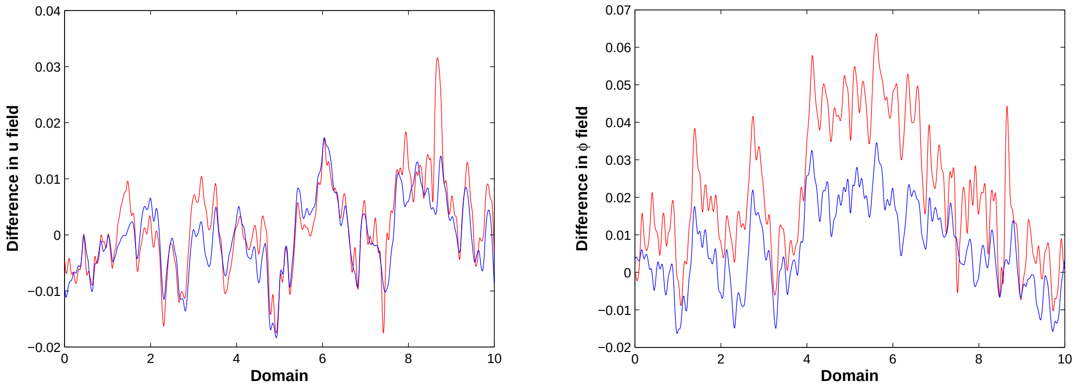
It is also interesting to look at individual analysis errors over the domain. At each grid point the analysis error is given by the difference between the true analysis and the analysis resulting from the assimilation. Figures 7.9 and 7.10 show the analysis errors in the u and ϕ fields at $t = 0$ and $t = 50$, 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 7.9 to 7.10 we observe that as the forecast evolves the analysis errors become smoother. At the centre of the time window, the errors in the u field for a Markov and a diagonal approximation are very similar compared to at the start of the



(a) (b)

Figure 7.9: Analysis errors in (a) u field and (b) ϕ field at the start of the time window. The red line is for a diagonal approximation and the blue line is for a Markov approximation with $L_R = 0.2$m.



(a) (b)

Figure 7.10: Analysis errors in (a) u field and (b) ϕ field at the centre of the time window. The red line is for a diagonal approximation and the blue line is for a Markov approximation with $L_R = 0.2$m.

time window, where the diagonal approximation was noticeably poorer. The same is true of the ϕ field although 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.

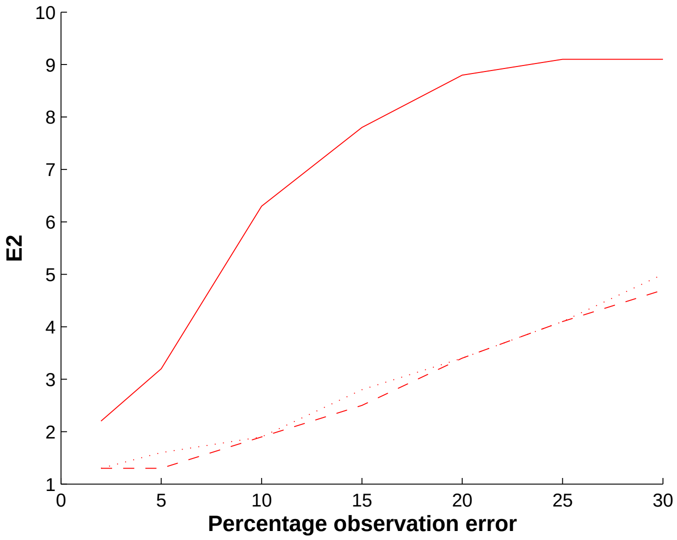


Figure 7.11: Plot of E2 against level of observation noise for u field. The solid line is for the diagonal approximation, the dashed line for the ED approximation with $k = 50$ and the dotted line for the Markov approximation with $L_R = 0.05$m.

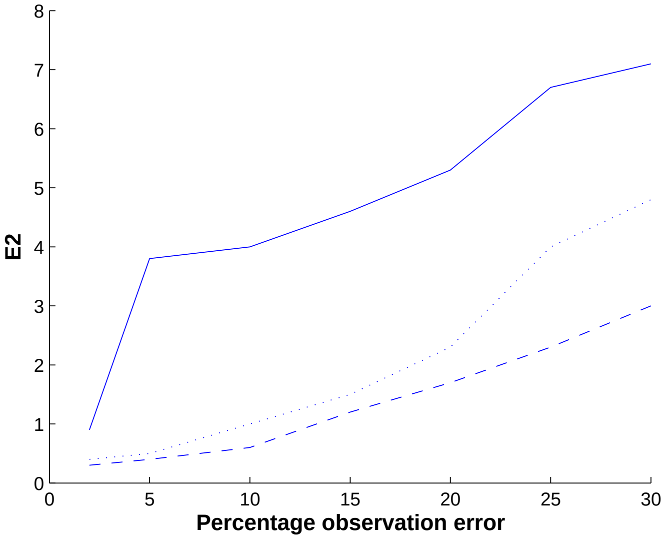


Figure 7.12: As in Figure 7.11 but for ϕ field.

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, i.e, 0.02ms^{-1} for the u field; 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 (7.4). A plot of this error measure versus the percentage observation error in the u and ϕ field is shown in Figures 7.11 and 7.12, 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.

7.3.3 Summary

We have examined different matrix approximations to an observation error correlation matrix with a SOAR distribution. The aim of this section was to reinforce and extend on the conclusions of Section 7.2 where the true error correlation matrix had a Markov structure. Using the same experimental framework we found that the Markov matrix approximation still produced the smallest analysis error when assimilated instead of the true error covariance matrix. An ED approximation with 100 eigenpairs also performed well, but would be more expensive to implement. Noticeably most matrix approximations that included some level of correlation structure outperformed uncorrelated diagonal approximations; exceptions were ED approximations with very few eigenpairs. The findings were in line with those in Section 7.2, and demonstrated the Markov matrix as a robust choice for modelling error correlation structure.

We also studied the spread of analysis errors over the model domain at the start and the middle of the assimilation time window. The differences between assimilations using the diagonal and Markov matrix approximations were not uniform over the model domain, and were significantly larger in places. Finally we looked at the behaviour of the analysis error E2 under different levels of observation noise. We concluded that regardless of noise level, the difference in the analysis error using a correlated and non-correlated approximation was significant. Also, as observation error increased, the difference in analysis error when using a diagonal approximation became larger. This reinforced the conclusion that including some correlation structure is both feasible and beneficial under a variety of assimilation conditions.

7.4 Experiment 3: Unknown error correlation structure

In the previous two experiments we assumed that the true observation error correlation structure was known. In reality this is often not the case. In this section we investigate the choice of approximating error structure when the error covariance matrix is assumed unknown and is subsequently derived using the post analysis diagnostic (2.33) described in Section 2.6.2. The aim of this final experiment is to determine (a) if it is possible to derive the true error correlation structure using the diagnostic; and (b) if we can successfully fit a Markov matrix approximation to the derived structure.

7.4.1 Application of Desroziers' method

Recall from Chapter 2, the post-analysis diagnostic used to determine observation error covariance structure is given by

$$\mathbb{E}\left[d_a^o (d_b^o)^T\right] \approx R \tag{7.5}$$

where $d_b^o = y - h(x^b)$ is the background innovation vector and $d_a^o = y - h(x^a)$ is the analysis innovation vector. In this experiment, all observations are taken directly so the observation operator is the identity matrix, i.e, $h = H = I$.

In the experiments performed in Chapter 4, the diagnostic (7.5) was shown to successfully quantify observation error correlation structure for IASI channel data. We apply the diagnostic using a similar approach to that described in Section 4.2. The experimental set up is the same as when we knew the true error correlation structure in Section 7.3, with the exception that the observations are taken at every timestep. We use a SOAR correlation matrix (7.2) with length scale $L_R = 0.1$m to generate the observation error correlations. The assimilation is then run under the assumption that the error structure is unknown and a diagonal approximation of the true error variances is used as the observation error covariance matrix. We then calculate the innovation vectors d_b^o and d_a^o at each of the 100 timesteps and write them to a MATLAB file. This procedure is performed a total of 10 times, to produce a sample set of 1000 background and analysis innovation vectors from which to calculate the error correlation statistics.

The diagnosed observation error covariance matrix is then computed as follows. For each spatial point i, we compute the observation error covariance with point j by averaging the product of the background and analysis innovations over the total number of observations N,

$$\begin{aligned} R(i,j) &= \frac{1}{N}\sum_{k=1}^{N}\{(d_a^o)_i\,(d_b^o)_j\}_k - \left(\frac{1}{N}\sum_{k=1}^{N}\{(d_a^o)_i\}_k\right)\left(\frac{1}{N}\sum_{k=1}^{N}\{(d_b^o)_j\}_k\right) \\ &= \frac{1}{N}\sum_{k=1}^{N}\{y_i - x_i^a\}_k\{y_j - x_j^b\}_k \\ &\quad - \left(\frac{1}{N}\sum_{k=1}^{N}\{y_i - x_i^a\}_k\right)\left(\frac{1}{N}\sum_{k=1}^{N}\{y_j - x_j^b\}_k\right), \end{aligned} \tag{7.6}$$

where y_i is the observation value at point i, and x_i^b and x_i^a are the background and analysis values at point i, respectively.

7.4.2 Diagnosing the true error correlation structure

The resultant diagnosed error correlation matrix is shown in Figure 7.13. The matrix is more symmetric than the IASI error correlation matrices diagnosed in Chapter 4. This is expected since the ignored SOAR correlation structure is weaker than that present in the IASI observation errors; hence we are deviating less from the assumption of correctly specified errors used in creating the diagnostic (7.5). Using this matrix we can approximate the true correlation structure of the errors, and determine the best Markov approximation to the derived errors. First we consider how well this matrix approximates the true correlation matrix used to generate the observation errors. In Figure 7.14 we plot the difference in the Frobenius norm between the diagnosed error correlation matrix and a SOAR matrix approximation. The Frobenius norm (3.30), as described in Section 3.5, provides an elementwise evaluation of the closeness of two matrices $\|C - C_S\|_F$ where C is the diagnosed error correlation matrix and C_S is the SOAR matrix approximation. By varying the length scales of the approximations, we can determine which SOAR matrix best fits the data.

In Figure 7.14 we see that the value $\|C - C_S\|_F$ is a minimum when the length scale of the SOAR matrix approximation C_S is $L_R = 0.1$m. This is the length scale of the matrix used in generating the error correlations, and we can therefore conclude that the diagnostic is successful in deriving the true error correlation structure.

In Figure 7.15 a typical row of the diagnosed error correlation matrix is plotted against the derived SOAR matrix approximation. The plot shows that although the diagnosed error correlation matrix fits well to a SOAR distribution for the main band of correlation structure, spurious correlations are present on the off-diagonals. If we wished to use

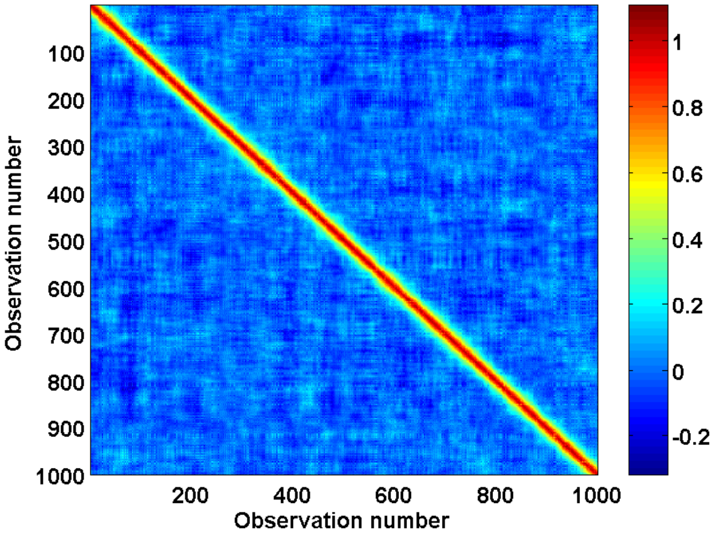


Figure 7.13: Diagnosed observation error correlation matrix.

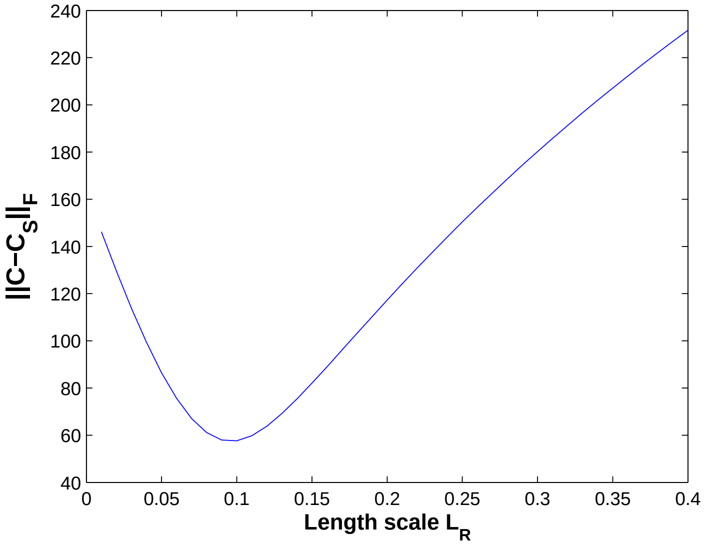


Figure 7.14: Frobenius norm of the difference between the diagnosed matrix C and a SOAR approximation C_S with length scale $L_R = 0.1$m.

the diagnosed matrix directly in a data assimilation algorithm, these spurious error correlations could possibly be removed using covariance localisation [42]. However, this procedure can be problematic for vertical errors, or for inter-channel error correlations.

Further experiments demonstrate how the success of the reconstruction is dependent on

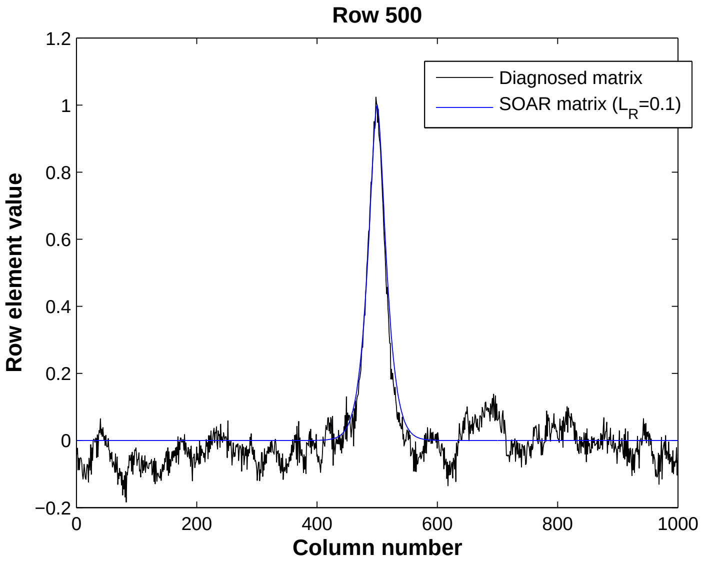


Figure 7.15: Row 500 of the diagnosed C matrix (black line) and a SOAR matrix with $L_R = 0.1$ (blue line)

the number of observations used in retrieving the diagnostics. In Figures 7.16 and 7.17, respectively, the observation error correlation matrices derived using 100 and 500 observations are shown. When using only 100 observations, spurious correlations are heavily present; whereas using 500 observations produces a matrix structure very similar to that in Figure 7.13. This suggests that it is possible to obtain a good reconstruction with fewer observations, but there must be a sufficient number to avoid spurious long range correlations. Similar conclusions were drawn for the construction of an ED approximation using a subset of eigenpairs in Chapter 5.

It is also important to note that the success of the matrix reconstruction was obtained for a correlation matrix with only one free parameter ρ. Comparing this to an IASI error correlation matrix where the correlation structure can be influenced by several parameters, we conclude that we are limited in the extent to which we can generalise these findings. However, the results do demonstrate the applicability of the method and motivate further study.

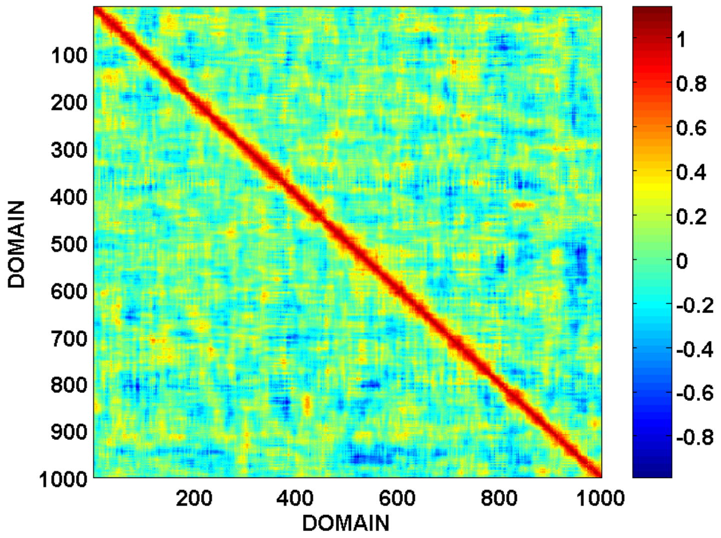


Figure 7.16: Diagnosed observation error correlation matrix using 100 observation sets.

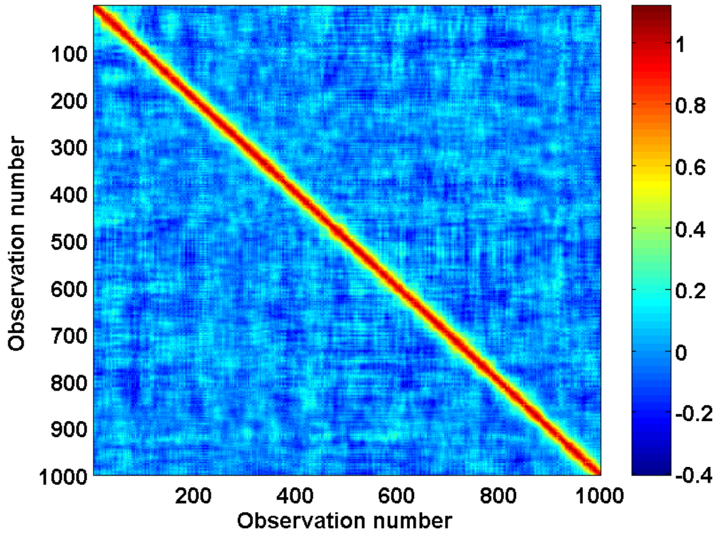


Figure 7.17: Diagnosed observation error correlation matrix using 500 observation sets.

7.4.3 Diagnosing an approximate error correlation structure

Now we address the second aim of this final results section: can the diagnosed error correlation structure be used to derive an optimal Markov approximation. We use a Markov approximation because it has been shown in the previous two experimental sections to be a robust and efficient way of modelling error correlation structure. Figure 7.18 shows the difference in the Frobenius norm between the diagnosed matrix error correlation matrix C and a Markov matrix approximation C_M. As in Figure 7.14 we vary the length scale to find the best fit to the diagnosed data. The smallest value of $\|C - C_M\|_F$ occurs when the length scale of the Markov matrix is $L_R = 0.2$m. This was the length scale found to generate the most successful Markov approximation in the previous tests using a known SOAR error correlation matrix. However, these results demonstrate that such an approximation can be diagnosed without prior knowledge of the error correlation distribution. This is encouraging for situations when calculating the true error correlation structure may be difficult.

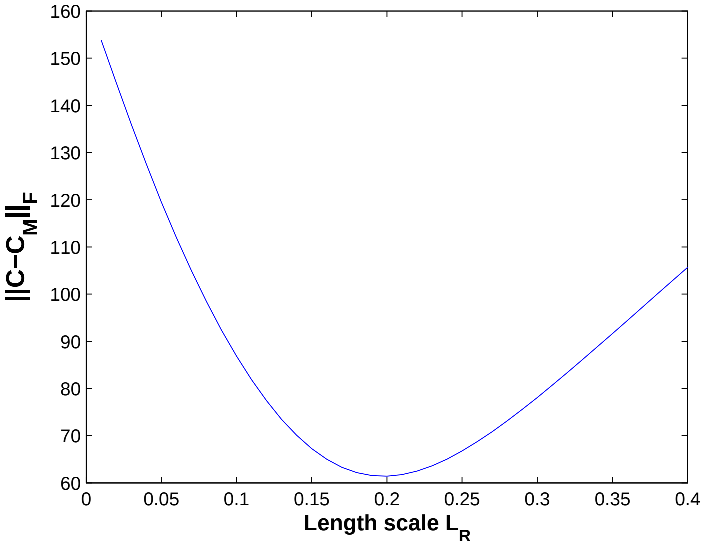


Figure 7.18: Frobenius norm of the difference between the diagnosed matrix C and a Markov approximation C_M with length scale L_R.

7.4.4 Summary

In this section we have shown new and original results on the derivation of true and approximate error correlation structures using post-analysis diagnostics. We have successfully derived the true SOAR error correlation structure when the assumption of uncorrelated errors was used in the assimilation. For a good approximation, a sufficient number of observations were needed; using too few observations resulted in long-range spurious error correlations.

We were also able to fit a Markov matrix approximation to the derived structure using the Frobenius norm as a measure of the difference between matrices. The Markov matrix diagnosed to be the best fit was also shown to be the best matrix approximation in Section 7.3, where the experiment conditions were very similar. We can therefore conclude that it is possible to diagnose a successful Markov approximation to a simple correlation matrix without prior knowledge of the error distribution.

7.5 Conclusions

In this chapter we investigated the inclusion of observation error correlation structure in an incremental 4D-Var algorithm using a 1D shallow water model. The work extended on the findings in Chapter 5 using the techniques described in Chapter 6. We ran the assimilation using three different approximate error correlation structures: diagonal matrices, Markov matrices and ED matrices. In experiments 1 and 2, these matrix approximations were tested against a simulated error correlation structure following a Markov and a SOAR distribution, respectively.

The Markov matrix approximating structures were found to generate the smallest analysis errors for both distributions. The structure was also shown to be robust under choice of length scale. Diagonal approximations performed poorly, and both a Markov matrix with very small length scale and an ED approximation with 5% of the available eigenpairs produced a more accurate analysis. The results reinforced conclusions made in Chapter 5, and demonstrated that including some correlation structure, even a basic approximation, is often better than incorrectly assuming error independence. The findings also support the work in [43] where Healy and White showed that using an approximate error correlation structure gave clear benefits over using no observation error correlations.

In the final section of this chapter we examined the choice of an approximate error correlation structure when the true error distribution was assumed unknown. We used a Markov matrix as the approximating matrix based on its successful performance in the previous two experiments. The observation error correlations were sampled from a SOAR distribution but were treated as uncorrelated in the assimilation, i.e, a diagonal observation error covariance matrix was used. Using the post-analysis diagnostic shown in Chapter 4 to accurately quantify IASI error correlations and in [25] to accurately estimate mis-specified observation error variances, we successfully diagnosed the true observation error correlation structure. The derived matrix was however subject to spurious long-range error correlations. We then used matrix differences in the Frobenius norm to ascertain the optimal Markov matrix approximation to the derived error correlation matrix. This was found to be the same matrix as that which generated the smallest analysis error in Section 7.3. We therefore concluded that even when the true error corelation structure is unknown, it is possible to derive cheaply an approximating structure that performs well in the assimilation.

The results in this chapter addressed the final thesis question posed in Chapter 1: how well do the proposed matrix approximations perform in a data assimilation algorithm? We have shown that correlated approximations can reduce the analysis error when used over simplistic diagonal approximations. The final section also demonstrated how to choose an approximating correlation structure when the true error correlation structure was unknown. In conclusion we can deduce that (a) it is often better to model error correlation structure incorrectly than not at all, and (b) including error correlation structure in data assimilation algorithms is both feasible and beneficial.

Chapter 8

Conclusions and future work

In numerical weather prediction (NWP), an accurate, high-resolution representation of the current state of the atmosphere is needed as an initial condition for the propagation of a weather forecast. Data assimilation techniques combine observations of atmospheric variables with *a priori* knowledge of the atmosphere to obtain a consistent representation. The weighted importance of each is determined by the size of their associated errors, so it is crucial to the accuracy of the forecast that these errors be specified correctly.

Satellite radiance observations account for approximately 90% of the total data used in operational assimilations [4], and are a contributing factor in the success of data assimilation algorithms such as four-dimensional variational assimilation (4D-Var). The correct treatment of radiance observation errors is a dual 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 spatial and horizontal 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 infeasible.

To avoid the issues involving observation error correlations, operational weather centres treat most observation errors as independent. Often for satellite observations, the lack of correlation is compensated for by inflating the error variances so that the observations have a more appropriate weighting in the analysis [46]. The assumption of zero correlations is often also used in conjunction with data thinning methods such as superobbing [5], in which data in a region are reduced to a single representative observation. Under such conditions, increasing observation density beyond some threshold value has been shown to yield little or no improvement in analysis accuracy [60], [21]. With the advent of high-resolution nowcasting, in which all available data is required to provide details on finer scales, such assumptions will not be viable and an alternative approach to dealing with observation error correlations is needed.

In this thesis we expanded on the existing body of work on observation error correlation structure, and addressed the dual problem of correlation specification and modelling. In Chapter 1 we posed three questions which we answered through the subsequent experiments and analysis:

- What is the true structure of the observation error correlations?
- What approximations are available to model error correlation structure? What is their impact on data assimilation diagnostics?
- How well do these approximations perform in a data assimilation experiment? Is

it better to model observation error correlation structure incorrectly than not at all?

We began in Chapter 2 by introducing the concepts of data assimilation and satellite remote sensing. The role of observation error correlations was explained in this context; we gave an overview of their possible origins and discussed current issues in their treatment. Finally we described two statistical methods used to diagnose error correlations; the Desroziers' statistical approximation [25] was later applied in Chapter 4 and in Chapter 7.

In Chapter 3 we addressed the second question posed in Chapter 1, and examined the different approximating structures that can be used to represent error covariance matrices. Three different types of approximating structure were described: diagonal [14], circulant [43], [78] and eigendecomposition [34] approximations. We focused on the unique properties of the approximating matrices that make them suitable for use in variational data assimilation algorithms. Details on several retrieval measures used to evaluate the success of these approximations were also given.

In order to generate a good approximation, we must first have an accurate estimate of the true error correlation structure. In Chapter 4 we returned to the first question posed in Chapter 1, and successfully used the Desroziers' post-analysis diagnostic described in Chapter 2 to quantify cross-channel error correlations for IASI observations. The statistics used in the construction of the diagnostic were generated from the Met Office operational systems. The observation error covariance matrix was derived for both the one-dimensional retrieval procedure and the incremenatal 4D-Var assimilation. It is a new approach to use the Desroziers' technique to estimate the full error covariance matrix and not just the error variances.

We presented more new results in Chapter 5. Here we quantified the success of each of the matrix approximations described in Chapter 3 in modelling an empirically derived observation error correlation structure. The experiments were performed for independent and correlated background errors using a three-dimensional variational assimilation framework. Using the information content measures described in Chapter 3, we calculated the information provided by each approximation relative to the truth. The work in this chapter addressed the second thesis question.

Finally we chose to investigate modelling observation error correlation structure in the framework of the one-dimensional SWEs. This is a relatively simple model that retains key dynamics similar to those of the full atmosphere. The penultimate two chapters addressed the third and final question posed in Chapter 1. In Chapter 6, we developed an incremental 4D-Var data assimilation system for the 1D SWEs which models observation error correlation structure using diagonal, Markov and eigendecomposition matrix approximations. The important matrix-vector products for implementing these approximations efficiently in the data assimilation process were provided. In Chapter 7 we performed assimilation experiments extending the findings in Chapter 5 using the framework described in Chapter 6. The assimilation accuracy was evaluated for each approximation under different realisations of the true observation error distribution. We now discuss the conclusions we are able to draw from the new results of the work in this thesis.

8.1 Summary

We separate our conclusions under the three questions posed in Chapter 1.

Question 1: What is the true structure of the observation error correlations?

In Chapter 4 we showed that the cross-channel observation error correlation structure can be derived for IASI data using a post-analysis diagnostic [25]. Using statistics generated from the Met Office operational systems we deduced the following conclusions from the numerical experiments:

- There exist significant error correlations between neighbouring channels with similar properties, such as sensitivity to water vapour and typical brightness temperature measurements. This results in a block diagonal structure in the error covariance and correlation matrices;
- Error variances are being overestimated in the 1D-Var retrieval procedure and the 4D-Var assimilation process. This inflation is needed because of the current mis-treatment of off-diagonal error covariances;
- The largest errors of representativity are present in channels highly sensitive to water vapour. This suggests that fine-scale water vapour structures are observed by the IASI instrument but are not represented at the current model resolution.

Question 2: What approximations are available to model error correlation structure? What is their impact on data assimilation diagnostics?

In Chapter 3 we described three approximating structures suitable for modelling observation error covariance matrices. We focused on the features of these matrices which made them suitable for inclusion in variational data assimilation algorithms, namely having a cheaply generated and easy to store inverse. In Chapter 5 we made a quantitative comparison of the information content from a simulated set of observations under each approximating structure. The empirical conclusions were:

- Information content is severely degraded under the incorrect assumption of independent observation errors. This supports the results seen in [14] and [43];
- Retaining some error correlation structure shows clear benefits in terms of information content. A circulant approximation was shown to retain the most information content of all the approximations. An eigendecomposition approximation retained more information than a diagonal approximation but sufficient eigenpairs must be used to avoid spurious long range error correlations as suggested in [34];
- The diagnosed information content was sensitive to the specification of the analysis error covariance matrix. If the approximating observation error covariance matrix was assumed to be correct, then the resultant information content values were inflated and misleading. This highlighted the importance of knowing accurately the correct error correlation structure for an observation type, even if an approximation to this structure is to be made.

Question 3: How well do these approximations perform in a data assimilation experiment? Is it better to model error correlation structure incorrectly than not at all?

In Chapter 6 and 7 we developed an incremental 4D-Var data assimilation algorithm that used correlated approximations to model a simulated error correlation structure. This was applied to one-dimensional SWEs, and the impact of each approximation on analysis accuracy was determined. In this final new results section 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, it is often 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;
- A Markov matrix approximation is an effective and robust approximation to modelling error correlation structure;
- It is possible to derive a suitable Markov matrix approximation when the observation errors are assumed incorrectly to be independent and the Desroziers' diagnostic is used to derive the true error correlation structure. Care must be taken to use a sufficient number of statistics in the diagnostic to avoid spurious long-range error correlations in the derived matrix.

The three sets of new results presented in this work have answered the questions posed at the start of the thesis. We have demonstrated that is it both feasible and beneficial to model observation error correlation structure under a variety of assimilation conditions. Current operational systems require methods of incorporating observation error correlation structure; the correlated approximations described in this work have shown promise and warrent further study. However, the results generated also have some limitations. In the next section we will comment on the constraints of our studies and describe the possible extensions to our work that will generalise our results in a wider framework.

8.2 Future work

In Chapter 4 we used a post-analysis diagnostic derived from variational data assimilation theory to quantify cross-channel error correlations for IASI observations. The diagnostic proved successful in generating a feasible observation error covariance matrix; however the matrix was not entirely symmetric. We can attribute the asymmetry to violations in the assumptions used in constructing the diagnostic. To derive the diagnostic from [25], we assumed that all error covariances were specified correctly in the analysis. In the assimilations performed, the observation errors are incorrectly treated as independent, and the background errors may be poorly specified. Therefore this assumption does not hold entirely, and by construction the diagnosed matrix is not expected to be symmetric.

For an observation error covariance matrix to be used in operational applications it must be symmetric. If we wished to use our diagnosed matrix, R, directly in an assimilation system, we could use its symmetric part, R_{sym}, as the observation error covariance matrix:

$$R_{\text{sym}} = \frac{1}{2}(R + R^T).$$

We would hope that the closer we get to the correct specification of the observation and background error covariances, the more symmetric the diagnosed matrix would be.

However, even if a symmetric observation error covariance matrix was diagnosed, the current Met Office incremental 4D-Var system does not support a non-diagonal observation error covariance matrix, and so we cannot directly test the impact of using this diagnosed matrix in the assimilation. However, the 1D-Var retrieval procedure does allow a correlated observation error covariance matrix. In Chapter 4 the diagnosed error covariance matrix for the 1D-Var assimilation was shown to be very weakly correlated, implying that we would see little impact from including correlation structure. However, if we were to use reconstructed radiances in the 1D-Var procedure, we would expect a more strongly correlated matrix in 1D-Var. It would then be possible to assimilate the IASI reconstructed radiances in 1D-Var using a diagnosed error covariance matrix, and to evaluate the impact on the accuracy of the subsequent 4D-Var assimilation. Positive results would further motivate the inclusion of observation error correlation structure in the main assimilation.

By using a simple one-dimensional SWM in Chapters 6 and 7, we were able to evaluate the impact of modelling error correlation structure against using the true distribution of the observation errors. We observed improvements in analysis accuracy when Markov and ED correlated approximations were used over uncorrelated diagonal approximations. These results were encouraging but before the matrix approximations can be considered for operational applications they also need to be tested in a framework more representative of the full atmospheric model.

In the model used in Chapter 7, 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 success of the correlated approximations does however motivate an extension of this work in which an observation operator incorporating the typical integrated nature of satellite measurements is used in the assimilation algorithm. Also the assumption of uncorrelated background errors is unrealistic. We have seen in Chapter 5 how the background error structure can influence the choice of the observation error covariance matrix approximation. By coding a correlated background error covariance matrix, the interaction between observation and background errors could be studied further.

Additional methods to assess the quality of the analysis and the performance of the data assimilation algorithm could also be used. For operational interest it would be useful to compare the convergence properties and computational efficiency of the assimilation using each matrix approximation. Techniques to study the assimilation convergence rates are already available in the SWM code. Also, the conditioning of the minimisation could be studied by generating the Hessian matrix of the incremental cost function. The Hessian matrix can be described as the inverse of the analysis error covariance matrix, therefore from the Hessian we would also be able to calculate the information content available from the observations and compare with our 3D-Var results from Chapter 5.

In Chapter 7 we concluded that a Markov matrix was a robust and effective approximation to modelling error correlation structure. We also diagnosed an 'optimal' Markov matrix approximation when the true error correlation structure was unknown. A useful extension to this work would be a technique to derive the optimal Markov approximation to an observation error correlation matrix without needing to evaluate the Frobenius norm metric for several different matrices. Similar problems have been studied in financial research. In [87] the problem of finding the nearest positive semidefinite Toeplitz matrix (in the Frobenius norm) to an arbitrary matrix was considered. In [44] the nearest correlation matrix to a given symmetric matrix was determined by minimising the distance between the two matrices in a weighted Frobenius norm. By extending the ideas in [44] and [87] we could solve the minimisation problem of determining the optimal Markov matrix approximation to a given correlation matrix.

Such a technique would also be applicable for the current treatment of observation error correlations. The inflation of observation error variances performed in many operational centres is done using educated guess-work. By finding the diagonal approximation to a true error correlation matrix which minimised the matrix difference in a weighted Frobenius norm, we would have a more accurate representation of the observations in the analysis. In a situation where it was unavoidable to use the assumption of uncorrelated errors, we could at least be confident that the observations were being weighted correctly.

Appendix A: IASI channel information

All the details in this appendix are provided by Fiona Hilton (personal communication). The tables below contain information on the 314 IASI channels stored in the Met Office database (MetDB). The column entries are described below:

1. MetDB channel number: the channel number out of 314 stored in the MetDB
2. OPS index number: the index of the MetDB channel, out of 183, used in the OPS (starting at 0)
3. Var index number: the index of the MetDB channel, out of 139, used in 4D-Var (starting at 0)
4. Central wave number of the channel
5. Q jac peak (hPa): the pressure level at which the water vapour mixing ratio Jacobian peaks [30]
6. Summed Q jac peak: the sum over all model pressure levels of the absolute value of the water vapour mixing ratio Jacobian, normalised by the maximum of the totals for the 314 MetDB channels (out of 1)

MetDB channel number	OPS index number	Var index number	Central wave number	Q jac peak (hPa)	Summed Q jac
1	0		648.50	0.36	0
2	1	0	654.25	0.36	0
3	2		657.00	0.45	0
4	3	1	657.50	0.45	0
5	4		658.50	2.06	0
6	5	2	659.00	1.36	0.001
7	6		659.50	1.36	0.003
8	7		660.60	1.09	0
9	8	3	660.50	0.36	0
10	9		661.25	0.29	0
11	10		662.25	2.04	0
12	11		662.75	0.87	0
13	12		663.25	0.87	0
14	13		664.50	1.66	0
15	14		665.00	0.29	0
16	15		665.50	0.70	0
17	16		666.00	1.09	0
18	17		666.50	2.06	0
19	18		667.00	0.45	0
21	19		668.50	0.70	0
22	20		669.00	0.70	0
23	21		669.50	0.56	0
24	22		670.00	0.87	0
25	23		670.75	14.81	0
26	24		671.25	0.87	0
27	25	4	672.00	1.09	0
28	26		672.50	0.45	0
29	27		673.00	2.51	0
30	28	5	673.75	0.87	0
31	29		674.50	0.22	0
32	30	6	675.25	0.56	0
33	31		676.00	1.36	0
34	32	7	676.75	1.36	0
35	33		677.50	1.09	0
36	34		678.00	0.36	0
37	35	8	678.50	0.87	0
38	36		679.25	1.36	0
39	37	9	680.00	1.36	0
40	38		680.75	0.29	0
41	39		681.25	1.09	0
42	40	10	681.75	0.22	0
43	41		682.50	0.45	0
44	42	11	683.25	0.17	0
45	43		684.00	0.29	0
46	44		684.50	0.87	0
47	45	12	685.00	0.70	0
48	46		685.50	0.56	0

MetDB channel number	OPS index number	Var index number	Central wave number	Q jac peak (hPa)	Summed Q jac
49	47	13	686.50	0.45	0
50	48		687.25	0.87	0
51	49	14	688.00	0.87	0.002
52	50		688.75	1.36	0.001
53	51	15	689.50	1.09	0.001
54	52	16	689.75	0.87	0
55	53	17	691.00	0.17	0
56	54	18	691.50	0.70	0
57	55	19	693.00	0.87	0
58	56	20	694.50	321.50	0.001
59	57	21	696.00	269.65	0.001
60	58	22	696.50	269.65	0.001
61	59	23	697.25	339.39	0
62	60	24	697.75	269.65	0.006
63	61	25	698.25	286.60	0.002
64	62	26	699.00	416.40	0
65	63	27	699.50	396.81	0
66	64	28	700.25	321.50	0
67	65	29	700.75	339.39	0
68	66	30	701.25	436.95	0
69	67	31	702.25	321.50	0
70	68	32	702.75	358.28	0
71	69	33	703.75	303.55	0.009
72	70	34	704.50	478.54	0.003
73	71	35	705.25	339.39	0.047
74	72	36	705.50	339.39	0.053
75	73	37	706.25	436.95	0.006
76	74	38	707.00	358.28	0.016
77	75	39	707.75	416.40	0.025
78	76	40	708.25	478.54	0.004
79	77	41	709.75	457.27	0.004
80	78	42	710.25	457.27	0.004
81	79	43	711.00	610.60	0.016
82	80	44	711.50	610.60	0.005
83	81	45	712.00	610.60	0.005
84	82	46	713.50	358.28	0.103
85	83	47	714.75	638.60	0.022
86	84	48	715.25	610.60	0.009
87	85	49	718.25	457.27	0.004
88	86	50	718.75	457.27	0.002
89	87	51	719.50	377.05	0
90	88		720.50	0.45	0
91	89	52	721.25	543.05	0.004
92	90	53	725.50	727.44	0.038
93	91	54	726.50	759.16	0.123
94	92	55	727.00	727.44	0.026
95	93	56	728.50	696.97	0.032
96	94	57	731.00	478.54	0.478

MetDB channel number	OPS index number	Var index number	Central wave number	Q jac peak (hPa)	Summed Q jac
97	95	58	731.50	610.60	0.092
98	96	59	732.25	727.44	0.113
99	97	60	733.25	727.44	0.022
100	98	61	733.75	759.16	0.096
101	99	62	734.75	727.44	0.023
102	100	63	736.25	727.44	0.022
103	101	64	737.50	727.44	0.062
104	102	65	738.00	727.44	0.048
105	103	66	738.50	759.16	0.121
106	104	67	739.00	727.44	0.076
107	105	68	739.50	727.44	0.054
108	106	69	740.00	727.44	0.175
109	107	70	740.50	416.40	0.223
110	108	71	741.25	543.05	0.008
111	109	72	742.00	696.97	0.08
112	110	73	744.25	610.60	0.426
113	111	74	745.00	610.60	0.413
114	112	75	745.75	759.16	0.215
115	113	76	746.50	759.16	0.178
116	114	77	747.25	759.16	0.237
117	115	78	748.25	610.60	0.224
118	116	79	748.75	759.16	0.344
119	117	80	751.25	759.16	0.132
120	118	81	751.75	759.16	0.288
121	119	82	752.75	727.44	0.188
122	120	83	753.25	727.44	0.188
123	121	84	754.50	610.60	0.496
124	122	85	756.00	696.97	0.298
125	123	86	759.00	792.18	0.229
126	124	87	773.50	792.18	0.360
127	125	88	781.25	792.18	0.354
128	126	89	782.75	792.18	0.380
130	127	90	786.25	792.18	0.367
131	128	91	787.50	792.18	0.359
132	129	92	788.00	792.18	0.348
133	130		806.25	792.18	0.349
134	131	93	810.25	792.18	0.305
135	132	94	811.75	792.18	0.301
136	133	95	833.75	792.18	0.265
137	134	96	861.50	792.18	0.234
138	135	97	871.25	727.44	0.668
139	136	98	875.00	792.18	0.222
140	137	99	901.50	792.18	0.198
141	138		906.25	792.18	0.326
142	139		925.00	759.16	0.521
143	140	100	928.00	792.18	0.178
144	141		942.50	792.18	0.157
145	142	101	943.25	792.18	0.167

MetDB channel number	OPS index number	Var index number	Central wave number	Q jac peak (hPa)	Summed Q jac
146	143	102	962.50	792.18	0.150
162	144		1091.25	696.97	0.528
163	145	103	1096.00	792.18	0.087
164	146	104	1115.75	826.58	0.088
165	147	105	1142.50	826.58	0.086
166	148		1149.50	610.60	0.694
167	149	106	1168.25	792.18	0.099
168	150		1174.50	377.05	0.896
170	151	107	1204.50	826.58	0.106
171	152	108	1206.00	727.44	0.485
176	153	109	1330.00	457.27	0.907
178	154	110	1367.00	610.60	0.947
179	155	111	1371.50	478.54	0.914
183	156	112	1380.75	610.60	0.951
184	157	113	1381.75	499.54	0.900
185	158	114	1382.50	610.60	0.937
186	159	115	1384.25	499.54	0.942
189	160	116	1391.75	457.27	0.901
195	161	117	1401.50	436.95	0.899
196	162	118	1402.00	457.27	0.909
200	163	119	1408.00	478.54	0.942
201	164	120	1409.25	478.54	0.925
202	165	121	1410.75	478.54	0.911
215	166	122	1436.75	208.16	0.276
221	167	123	1456.75	208.16	0.231
251	168	124	1521.25	208.16	0.299
259	169	125	1539.00	208.16	0.257
261	170	126	1540.25	208.16	0.255
263	171	127	1542.00	208.16	0.269
270	172	128	1927.25	727.44	0.979
271	173	129	1986.75	727.44	0.887
272	174	130	1987.50	610.60	1.000
273	175	131	1989.50	638.60	0.967
274	176	132	1990.00	638.60	0.984
275	177	133	1990.50	610.60	0.973
276	178	134	1994.00	638.60	0.975
277	179	135	1994.50	727.44	0.912
278	180	136	1995.00	727.44	0.822
279	181	137	1995.50	759.16	0.748
280	182	138	1996.00	759.16	0.679

Appendix B: Application of the Desroziers' diagnostic to 4D-Var assimilation

Consider a state vector x_0 at time 0, whose true value is x_t and whose background estimate is x_b;

$$x_t = x_b + \epsilon^b,$$

where ϵ^b is the background error. The state vector can be evolved forward to time i under the tangent linear model $M(t_i, t_0) = M_i M_{i-1} \dots M_2 M_1$, i.e, $x_i = M(t_i, t_0, x_0)$. Consider m observations at different times, where the observations are related to the state vector through a forward model h,

$$\begin{aligned}
y_1 &= h(x_1) + \epsilon_1^o = h(M_1 x_t) + \epsilon_1^o \\
y_2 &= h(x_2) + \epsilon_2^o = h(M_2 M_1 x_t) + \epsilon_2^o \\
&\vdots \\
y_m &= h(x_m) + \epsilon_n^o = h(M_m \dots M_2 M_1 x_t) + \epsilon_m^o
\end{aligned}$$

where y_1 is an observation at time 1, y_2 is an observation at time 2, etc, and ϵ_i^o is the observation error for y_i.

In 4D-Var assimilation, the observations are combined with the background estimate, x_b, to produce an optimal analysis x_a, which minimises the cost function

$$J(x_0) = \frac{1}{2}(x_0 - x^b)^T B^{-1}(x_0 - x^b) + \frac{1}{2}\sum_{i=0}^{m}(h(x_i) - y_i)^T R_i^{-1}(h(x_i) - y_i) \tag{8.1}$$

where $R_i = \mathbb{E}\left[\epsilon_i^o(\epsilon_i^o)^T\right]$.

Assuming that the observation and background errors are uncorrelated, the cost function (8.1) can be approximated in matrix form by

$$J(x_0) = \frac{1}{2}(x_0 - x_b)^T B^{-1}(x_0 - x_b) + \frac{1}{2}(y - \hat{H}x_0)^T R^{-1}(y - hx_0)$$

where

$$
\begin{aligned}
y &= (y_1^T, y_2^T, \ldots, y_n^T)^T, \\
\hat{H} &= (M_1^T H^T, M_1^T M_2^T H^T, \ldots, M_1^T M_2^T \ldots M_n^T H^T)^T, \\
\epsilon^o &= ((\epsilon_1^o)^T, (\epsilon_2^o)^T, \ldots, (\epsilon_n^o)^T)^T, \\
R = \mathbb{E}\left[\epsilon^o(\epsilon^o)^T\right] &= \mathbb{E}\left[\begin{pmatrix} \epsilon_1^o(\epsilon_1^o)^T & \epsilon_1^o(\epsilon_2^o)^T & \ldots & \epsilon_1^o(\epsilon_n^o)^T \\ \epsilon_2^o(\epsilon_1^o)^T & \epsilon_2^o(\epsilon_2^o)^T & \ldots & \epsilon_2^o(\epsilon_n^o)^T \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_n^o(\epsilon_1^o)^T & \epsilon_n^o(\epsilon_2^o)^T & \ldots & \epsilon_n^o(\epsilon_n^o)^T \end{pmatrix}\right] \\
&= \begin{pmatrix} \mathbb{E}\left[\epsilon_1^o(\epsilon_1^o)^T\right] & 0 & \ldots & 0 \\ 0 & \mathbb{E}\left[\epsilon_2^o(\epsilon_2^o)^T\right] & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathbb{E}\left[\epsilon_n^o(\epsilon_n^o)^T\right] \end{pmatrix} \\
&= \begin{pmatrix} R_1 & 0 & \ldots & 0 \\ 0 & R_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & R_n \end{pmatrix},
\end{aligned}
$$

and H is the linearised observation operator.

This is the form of the cost function from 3D-Var data assimilation, and can be solved using the same approach, i.e, minimising the cost function. The solution to the 4D-Var assimilation problem can therefore be approximated by

$$x_a = x_b + B\hat{H}^T(\hat{H}B\hat{H}^T + R)^{-1}(y - \hat{H}x_b).$$

Appendix C: Additional gradient tests

Additional gradient tests for Chapter 7.

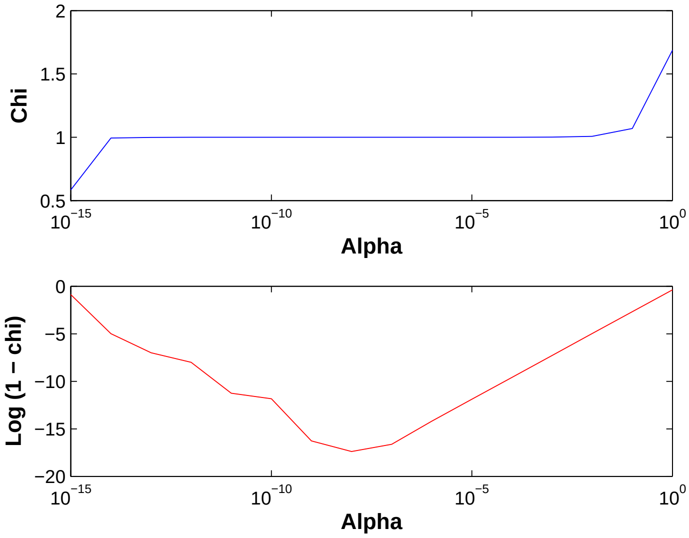


Figure 8.1: Gradient test for a diagonal approximation to a Markov error covariance matrix.

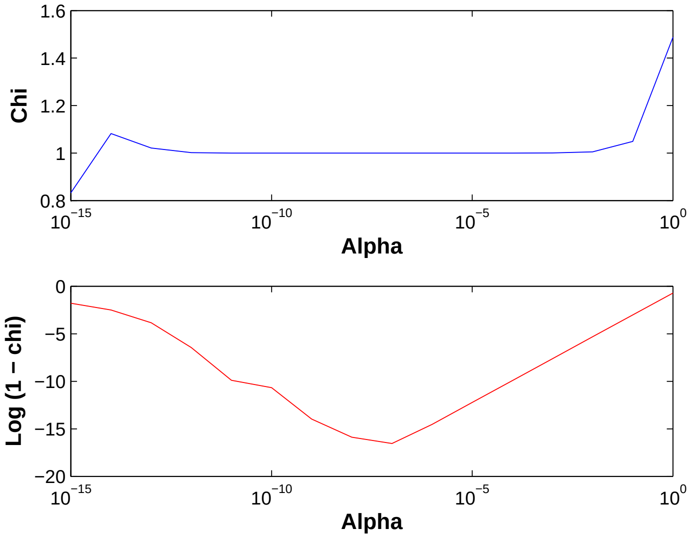


Figure 8.2: Gradient test for an ED approximation with $k = 50$ to a Markov error covariance matrix.

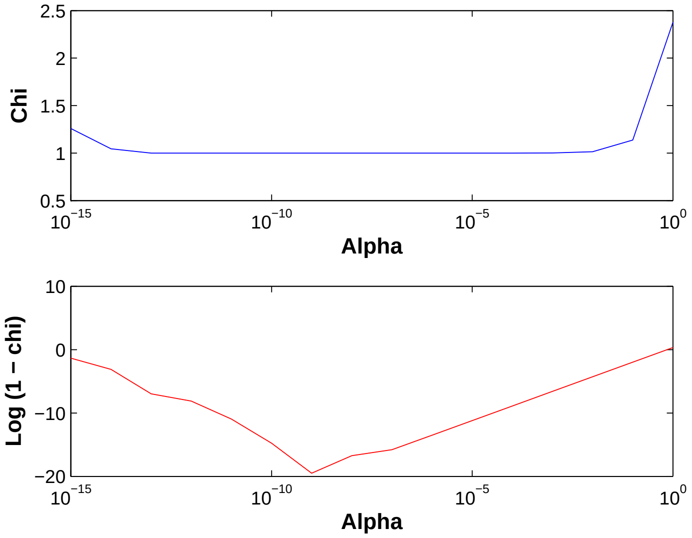


Figure 8.3: Gradient test for a Markov approximation with $L_R = 0.1$m to a SOAR error covariance matrix.

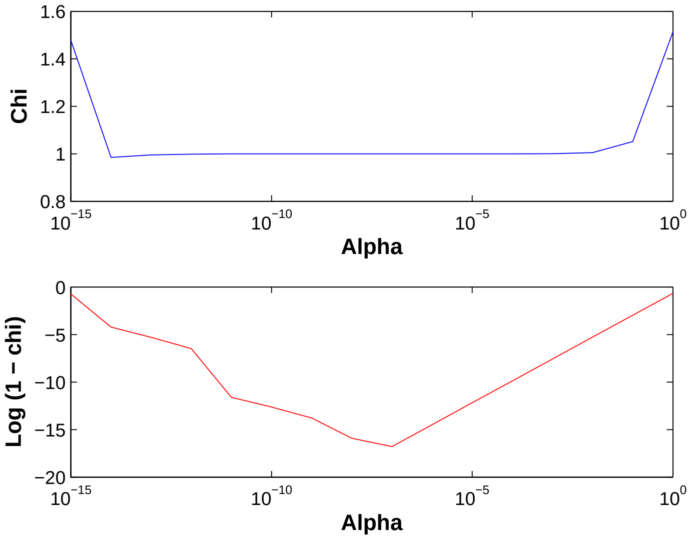


Figure 8.4: Gradient test for an ED approximation with $k = 50$ to a SOAR error covariance matrix.

Bibliography

[1] Met Office UK: Operational numerical modelling. In *http://www.metoffice.com/research/nwp/numerical/operational/index.html*, 2009.

[2] P. Antonelli, H.E. Revercomb, and L.A. Sromovsky. A principal component noise filter for high spectral resolution infrared measurements. *J. Geophys. Res.*, 109, 2004.

[3] R. Bannister. On control variable transforms in the Met Office 3D and 4D Var., and a description of the proposed waveband summation transformation. In *DARC internal report no. 5. http://www.met.rdg.ac.uk/ ross/DARC/WS/Waveband.pdf*, 2006.

[4] P. Bauer. The global observing system. In *Meteorological Training Course Lecture Series, ECMWF, Reading*, 2009.

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

[6] P. Bergthorsson and B. Doos. Numerical weather map analysis. *Tellus*, 7:329–340, 1955.

[7] 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.

[8] F. Bouttier and P. Courtier. Data assimilation concepts and methods. *ECMWF Meteorological Training Course Lecture Series*, 1999.

[9] G. Chalon, F. Cayla, and D. Diebel. IASI: An advanced sounder for operational meteorology. In *Proceedings of IAF, Toulouse, France, 1-5 October*, 2001.

[10] C. Chatfield. *Statistics for technology.* Chapman and Hall, London, third edition, 1983.

[11] M. H. Chaudhry. *Applied hydraulic transients.* Van Nostrand Reinhold Co., New York, second edition, 1987.

[12] W.C. Choa and L.P. Chang. Development of a four dimensional variational analysis system using the adjoint method at GLA. part 1: Dynamics. *Monthly Weather Review*, 120:1661–1674, 1992.

[13] A. Collard and A.P. McNally. The assimilation of Infrared Atmospheric Sounding Interferometer radiances at ECMWF. *Q.J.R.Meteorol.Soc.*, 135:1044–1058, 2009.

[14] 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.

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

[16] J. Conan, L.M. Mugnier, T. Fusco, V. Michau, and G. Rousset. Myopic deconvolution of adaptive optics images by use of object and point-spread function power spectra. *Appl. Optics*, 37:4614–4622, 1998.

[17] P. Courtier, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A.Holingsworth, F. Rabier, and M. Fisher. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part 1: formulation. *Q.J.R.Meteorol.Soc.*, 124:1783–1807, 1998.

[18] 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.

[19] M.J.P. Cullen. The unified/forecast climate model. *Meteorol. Mag.*, 122:81–94, 1993.

[20] R. Daley. *Atmospheric Data Analysis.* Cambridge University Press, Cambridge, 1993.

[21] 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.

[22] T. Davies, M.J.P. Cullen, A.J. Malcolm, M.H. Mawson, A. Staniforth, A.A. White, and N. Wood. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q.J.R.Meteorol.Soc.*, 131:1759–1781, 2005.

[23] D.P. Dee. Bias and data assimilation. *Q.J.R.Meteorol.Soc.*, 131:3323–3343, 2005.

[24] D.P. Dee and A.M. da Silva. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part 1: Methodology. *Monthly Weather Review*, 127:1822–1834, 1999.

[25] G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Q.J.R.Meteorol.Soc.*, 131:3385–3396, 2005.

[26] G. Desroziers and S. Ivanov. Diagnosis and adaptive tuning of observation-error parameters in variational assimilation. *Q.J.R.Meteorol.Soc.*, 127:1433–1452, 2001.

[27] R.M. Errico. What is an adjoint model? *Bull. Am. Met. Soc*, 78:2577–2591, 1997.

[28] R.M. Errico and T. Vukicevic. Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. *Mon. Wea. Rev.*, 120:1644–1660, 1992.

[29] J. Eyre. Planet Earth seen from space: Basic concepts. In *Exploitation of the new generation of satellite instruments for numerical weather prediction, seminar proceedings, pages 5-20, ECMWF, Reading*, 2000.

[30] J. Eyre. Inversion methods for satellite sounding data. In *Meteorological training course lecture series, ECMWF, Reading. http://www.ecmwf.int/newsevents/training/rcourse_notes/data_assimilation/index.html*, 2002.

[31] J. Eyre, S. English, V. Casses, and J. Pailleux. Benefits expected from the MTG-IRS mission. In *EUMETSAT 3rd Post-MSG User Consultation Workshop, Darmstadt. http://www.eumetsat.int/home/main/what_we_do/satellite/future_satellites/meteosat_third_generation/index.htm*, 2007.

[32] J. Eyre, G.A. Kelly, A.P. McNally, E. Andersson, and A. Persson. Assimilation of TOVS radiance information through one-dimensional variational analysis. *Q.J.R.Meteorol.Soc.*, 119:1427–1463, 1993.

[33] M. Fisher. Estimation of entropy reduction and degrees of freedom for signal for large variational analysis systems. *ECMWF Technical Memoranda*, 397, 2003.

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

[35] N. Fourrié and J.-N. Thépaut. Evaluation of the AIRS near-real-time channel selection for application to numerical weather prediction. *Q.J.R.Meteorol.Soc.*, 129:2425–2439, 2003.

[36] G.Golub and C.F.Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, third edition, 1996.

[37] R. Giering and T. Kaminski. Recipes for adjoint code construction. *Max-Planck Institute for Meteorology Technical Report*, 212, 1996.

[38] B. Gilchrist and G. Cressman. An experiment in objective analysis. *Tellus*, 6:309–318, 1954.

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

[40] M.D. Goldberg, Y. Qu, L.M. McMillin, W. Wolf, L. Zhou, and M. Divakarla. AIRS near real-time products and algorithms in support of operational numerical weather prediction. *IEEE Trans. Geosci. Remote. Sens.*, 41:379–389, 2003.

[41] A. Gray. *Toeplitz and circulant matrices: A review (Foundations and Trends in Communication and Information Theory)*. now publishers inc., 2006.

[42] T.M. Hamill, J.S. Whitaker, and C. Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129:2776–2790, 2001.

[43] 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.

[44] N.J. Higham. Computing the nearest correlation matrix - a problem from finance. *IMA.J.Numer.Anal.*, 22:329–343, 2002.

[45] F. Hilton, N.C. Atkinson, S.J. English, and J.R. Eyre. Assimilation of IASI at the Met Office and assessment of its impact through observing system experiments. *Q.J.R.Meteorol.Soc.*, 135:495–505, 2009.

[46] 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.

[47] A. Hollingsworth and P. Lonnberg. The statistical structure of short-range forecast errors as determined from radiosonde data. Part 1: The wind field. *Tellus*, 38A:111–136, 1986.

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

[49] E. Kalnay. *Atmospheric Modelling, Data Assimilation and Predictability.* Cambridge University Press, United Kingdom, 2003.

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

[51] A. Lawless. Irrotational shallow water model. In *http://darc.nerc.ac.uk*, 2005.

[52] 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.

[53] 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.

[54] 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.

[55] H.W. Lean, P.A. Clark, M. Dixon, N.M. Roberts, A. Fitch, R. Forbes, and C. Halliwell. Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Monthly Weather Review*, 136:3408–3424, 2008.

[56] J.M. Lewis, S. Lakshmivarahan, and S.K. Dhall. *Dynamic data assimilation: A least squares approach.* Cambridge University Press, New York, 2006.

[57] J. Li, C.-Y. Liu, H.-L. Huang, T.J. Schmidt, X. Wu, W.P. Menzel, and J.J. Gurka. Optimal cloud-clearing for AIRS radiances using MODIS. *IEEE Trans. Geosci. Remote. Sens.*, 43:1266–1278, 2005.

[58] Y. Li, I.M.Navon, P.Courtier, and P.Gauthier. Variational data assimilation with a semi-Lagrangian semi-implicit global shallow water equation model and its adjoint. *Monthly Weather Review*, 121:1759–1769, 1993.

[59] K.-N. Liou. *An Introduction to Atmospheric Radiation.* Academic Press, Inc, New York, 1980.

[60] 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.

[61] A. Lorenc. A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, 109:701–721, 1981.

[62] A. Lorenc. Analysis methods for numerical weather prediction. *Q.J.R.Meteorol.Soc.*, 112:1177–1194, 1986.

[63] A. Lorenc, S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Baker, J.R. Bray, A.M. Clayton, T. Dalby, D. Li, T.J. Payne, and F.W. Saunders. The Met Office global three-dimensional variational data assimilation scheme. *Q.J.R.Meteorol.Soc.*, 126:2991–3012, 2000.

[64] E.N. Lorenz. *The essence of chaos.* UCL Press, 1995.

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

[66] M. Matricardi, F. Chevallier, G. Kelly, and J.-N. Thépaut. An improved general fast radiative transfer model for the assimilation of radiance observations. *Q.J.R.Meteorol.Soc.*, 130:153–173, 2004.

[67] A.P. McNally. Analyis of satellite data. In *Meteorological Training Course Lecture Series, ECMWF, Reading*, 2009.

[68] A.P. McNally. The direct assimilation of cloud-affected satellite radiances in the ECMWF 4D-Var. *Q.J.R.Meteorol.Soc.*, 135:1214–1229, 2009.

[69] M.K. Ng. *Iterative Methods for Toeplitz Systems.* Oxford University Press, New York, 2004.

[70] E. Pavelin, S.J. English, and J.R. Eyre. The assimilation of cloud-affected infrared radiances for numerical weather prediction. *Q.J.R.Meteorol.Soc.*, 134:737–749, 2008.

[71] J. Pedlosky. *Geophysical Fluid Dynamics*. Springer-Verlag, New York, 1979.

[72] F. Rabier, A. Bouchard, C. Faccani, N. Fourrié, E. Gerard, V. Guidard, F. Guillaume, F. Karbou, P. Moll, C. Payan, P.Poli, and D. Puech. Global impact studies at Meteo-France. In *wwww.crnm.meteo.fr/gmap/sat/Global-Studies-MF.pdf*, 2009.

[73] F. Rabier and P. Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Q.J.R.Meteorol.Soc.*, 118:649–672, 1992.

[74] F. Rabier, N. Fourrié, D. Chafai, and P. Prunet. Channel selection methods for infrared atmospheric sounding interferometer radiances. *Q.J.R.Meteorol.Soc.*, 128:1011–1027, 2002.

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

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

[77] C.D. Rodgers. Information content and optimisation of high spectral resolution measurements. *Proc. SPIE*, 2830:136–147, 1996.

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

[79] Y. Sasaki. Some basic formulisms in numerical variational analysis. *Monthly Weather Review*, 98:875–883, 1970.

[80] R. Saunders, P. Rayer, T. Blackmore, M. Matricardi, P. Bauer, and D. Salmond. A new fast radiative transfer model - RTTOV-9. In *Joint 2007 EUMETSAT Meteorological Satellite Conference and the 15th Satellite Meteorology and Oceanography Conference of the American Meteorological Society, Amsterdam, The Netherlands*, 2007.

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

[82] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, 1949.

[83] 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.

[84] A. Staniforth and J. Coté. Semi-Lagrangian integration schemes for atmospheric models - a review. *Monthly Weather Review*, 119:2206–2223, 1991.

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

[86] L. Stewart, S. 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.

[87] T.J. Suffridge and T.L. Hayden. Approximation by a Hermitian positive semidefinite Toeplitz matrix. *SIAM J.Matrix.Anal.Appl.*, 14:721–734, 1993.

[88] O. Talagrand. A posteriori verification of analysis and assimilation algorithms. In *Proceedings of Workshop on diagnosis of data assimilation systems, ECMWF, Reading, UK, 2-4 November 1998*, pages 17–28, 1999.

[89] J.-N. Thépaut. Satellite data assimilation in numerical weather prediction: an overview. In *Meteorological Training Course Lecture Series, ECMWF, Reading. http://www.ecmwf.int/newsevents/training/rcourse_notes/data_assimilation/index.html*, 2003.

[90] J.-N. Thépaut and P. Courtier. Four dimensional data assimilation using the adjoint of a multi-level primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, 117:1225–1254, 1991.

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

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

[93] D.P. Wylie and W.P. Menzel. Eight years of high cloud statistics using HIRS. *J. Climate*, 12:170–184, 1999.