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

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

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

$\mathbf{x}$	state vector
$\mathbf{x}^t$	'true' state vector
$\mathbf{x}^b$	background state vector
$\mathbf{x}^a$	optimal analysis state vector
$\mathbf{y}$	observation vector
$\mathbf{h}$	observation operator
$\mathbf{H}$	linearisation of observation operator
$\mathbf{B}$	background error covariance matrix
$\mathbf{R}$	observation error covariance matrix
$\mathbf{C}$	observation error correlation matrix
$\mathbf{D}$	observation error variance matrix
$\mathbf{S}_a$	analysis error covariance matrix
$\mathbf{m}$	nonlinear forward model
$\mathbf{M}$	linear forward model
$\mathbf{M}^T$	adjoint model
$\mathbf{K}$	Kalman gain matrix
J	cost function
J_b	background term of cost function
J_o	observation term of cost function
$\mathbf{d}_b^o$	background innovation vector
$\mathbf{d}_a^o$	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 p 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 ap92351(r)]TJ2049op92351(r)]6(a)0.0492.23498

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 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 struc-**

ational 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 u.199178(S)0.32898(E)-0.01(i)-0.250651(l)-0.250651(s)-340.59(o)0.0492351(

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

ical cost, robustness, and the optimality of the solution generated are all important 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,

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

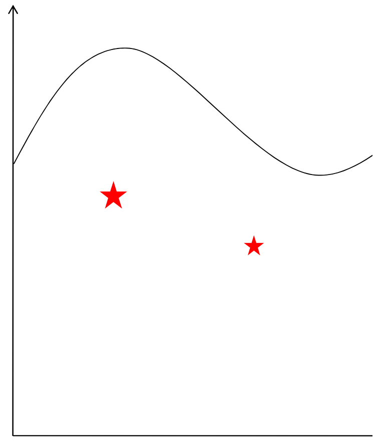
$$y = h(x^t) + \epsilon^o, \tag{2.2}$$

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 assu [illegible]

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

$$S_a = (H^T R^{-} H + B^{-})^{-} .$$

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



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

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

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

subject to

$$\delta x_{i+}^{(k)} = M(t_i, t_{i+}\ , x^{(k)})\delta x_i^{(k)},$$

where H_i is the linearisation of the observation operator h_i around the state $x_i^{(k)}$.

6. Update the guess field using

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

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 l

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

[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 av

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

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. At-

18(i)-0.250651(a)0.049...(r)-0.211487(a)0.04934.064(s)0.089106(16294]TJ41

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 $\frac{d\tau(z)}{dz}$. Therefore a radiance measurement

ance 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

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

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.e os51832351(m)00816 sity 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

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 $\mathbf{R} = \mathbb{E}[\epsilon^o(\epsilon^o)^T] \in \mathbb{R}^{m\times m}$. The Gaussian assumption does not hold in practice but the resultant pdfs make equation manipulation involving the errors algebraically simpler. The error 326742(r)-0.210368(e)-0.232748(s)0610368(.)-4260Td and y_2, respectively, and σ_2 is the error covariance of the two measurement components.

The observation errors can be classified as systematic or random, depending on whether 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

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

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

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

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

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 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

$$\mathbf{c} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T \qquad (2.29)$$

where $\mathbf{H}$ is the linearised observation operator. The i, j-th element of $\mathbf{c}$ 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

the method below.

Equations (2.1) and (2.2) show how the background state $\mathbf{x}^b$ and the observation vector $\mathbf{y}$ are approximations to the true state of the atmosphere $\mathbf{x}^t$. Assuming that the obser-

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, $\mathbb{E}\left[\epsilon^o(\epsilon^o)^T\right] = \mathbf{R}$, and the observation operator is linear. Care must therefore be taken when interpreting the results using these diagnostics.

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 covarim 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 of the observations in the analysis. The diagonal approximation is now in the form

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

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

$$\mathbf{T_m} = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \dots & t_{-(m-1)} \\ t_1 & t_0 & t_{-1} & \dots & t_{-(m-2)} \\ t_2 & t_1 & t_0 & \dots & \dots \\ \vdots & \ddots & & & \end{bmatrix}$$

plicitly, 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 $\mathbf{C}$ can be written in the form

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

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

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

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

$$F = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{-\frac{2\ i}{}} & e^{-\frac{2\ i}{}\times 2} & \dots & e^{-\frac{2\ i}{}\times(m-\ )} \\ 1 & e^{-\frac{2\ i}{}\times 2} & e^{-\frac{2\ i}{}\times 4} & \dots & e^{-\frac{2\ i}{}\times 2(} \end{bmatrix}$$

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 $\left\lceil \frac{m}{2} \right\rceil_+$ and $\left\lceil \frac{m}{2} + 1 \right\rceil_+$ where $\left\lceil \frac{m}{2} \right\rceil_+$ 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,

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

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 rs0.2J6.54297TLT[(C)-0.32677294(.)-0.248413(e)-0.2.48413(n)0.3

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 $m \rightarrow \infty$, the difference between $\mathbf{T}$ and $\mathbf{C}$ converges in the Frobenius norm, and $\mathbf{C}^{-1}$ becomes a good approximation to $\mathbf{T}^{-1}$, i.e, $\mathbf{C}^{-1}\mathbf{T} = \mathbf{I}$.

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

atmosphere, can be written in the form,

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

where k is the atmospheric level, μ is the mean of the spatial series, ρ is the autoregressive parameter, and $\square_{k+}$ 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 $\square$ as mutually independent, uncorrelated with the value of T, and Gaussian distributed with mean zero and variance $\sigma_{\square}^2$, 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 ma[illegible] 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

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

$$\mathbf{R} = \mathbf{D}^{1/2}\Big(\alpha\mathbf{I} + \sum_{k=}^{K} \big(\lambda_k - \alpha\big)$$

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

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

substituting in the Kalman gain matrix K we obtain

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

In 3D-Var data assimilation, S_a

error covariance matrix, as in [78], giving

$$S_a^{\square} = S_a + K R^{\square} K^T \tag{3.14}$$

$$K = BH^T(HBH^T + R_f)^{-} \tag{3.15}$$

$$R^{\square} = R_t - R_f \tag{3.16}$$

where $S_a^{\square}$ 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

Approaches A1 and A2

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 $\log_2$

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 $\mathbf{L}$, as in [33], such that $\mathbf{L}\mathbf{B}\mathbf{L}^T = \mathbf{I}$ and $\mathbf{L}\mathbf{S_a}\mathbf{L}^T = \hat{\mathbf{S}}_\mathbf{a}$, where $\mathbf{B}$ and $\mathbf{S_a}$ are both symmetric positive definite. This transformation is not unique as we can replace $\mathbf{L}$ by $\mathbf{X}^T\mathbf{L}$ where $\mathbf{X}$ is an orthogonal matrix. Now if we take $\mathbf{X}$ to be the matrix of eigenvectors of $\hat{\mathbf{S}}_\mathbf{a}$, then we simultaneously reduce $\mathbf{B}$ to the identity matrix and $\hat{\mathbf{S}}_\mathbf{a}$ to a diagonal matrix of its eigenvalues, $\mathbf{\Lambda}$;

$$
\begin{aligned}
\mathbf{X}^T\mathbf{L}\mathbf{B}\mathbf{L}^T\mathbf{X} &= \mathbf{X}^T\mathbf{X} = \mathbf{I}, \\
\mathbf{X}^T\mathbf{L}\mathbf{S_a}\mathbf{L}^T\mathbf{X} &= \mathbf{X}^T\hat{\mathbf{S}}_\mathbf{a}\mathbf{X} = \mathbf{\Lambda}.
\end{aligned}
$$

After this transformation, the diagonal elements (eigenvalues) of the transformed matrix $\mathbf{L}\mathbf{B}\mathbf{L}^T$ are unity and each corresponds to an individual degree of freedom. The eigenvalues of $\hat{\mathbf{S}}_\mathbf{a}$ may therefore be interpreTJ/R7310.9091Tf6.60q.853355n(59(t)0.313314(h)0.32898(e)-

n by

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 $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ 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

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

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 $f : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}$ 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,

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

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

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

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 $\mathbf{S}_{\mathbf{a}}^{\prime}$ and $\mathbf{S}_{\mathbf{a}}$ are the analysis error covariance matrices of $\mathbf{R}_{\mathbf{f}}$ and $\mathbf{R}_{\mathbf{t}}$, respectively.

3.8 Summary

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

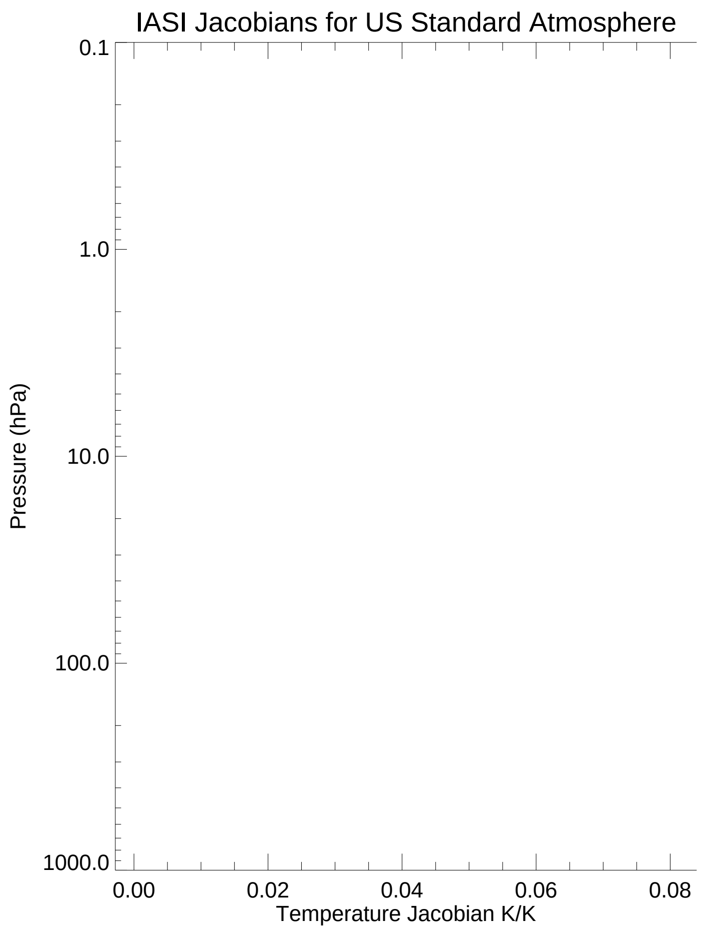
instrument was launched on the MetOp-A satellite in 2006 as part of the EUMETSAT European Polar System (EPS). Its spectral interval of 645-2760cm^{-} is divided into three bands and sampled by 8461 channels at a resolution of 0.5cm^{-}. Band one, from 645-1210cm^{-}, is used primarily for temperature and ozone sounding, band two (1210-2000cm^{-}) for water vapour sounding and the retrieval of N_2O and CH_4 column

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


IASI Jacobians for US Standard Atmosphere
0.1
1.0
10.0
100.0
1000.0
Pressure (hPa)
0.00
0.02
0.04
0.06
0.08
Temperature Jacobian K/K

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.

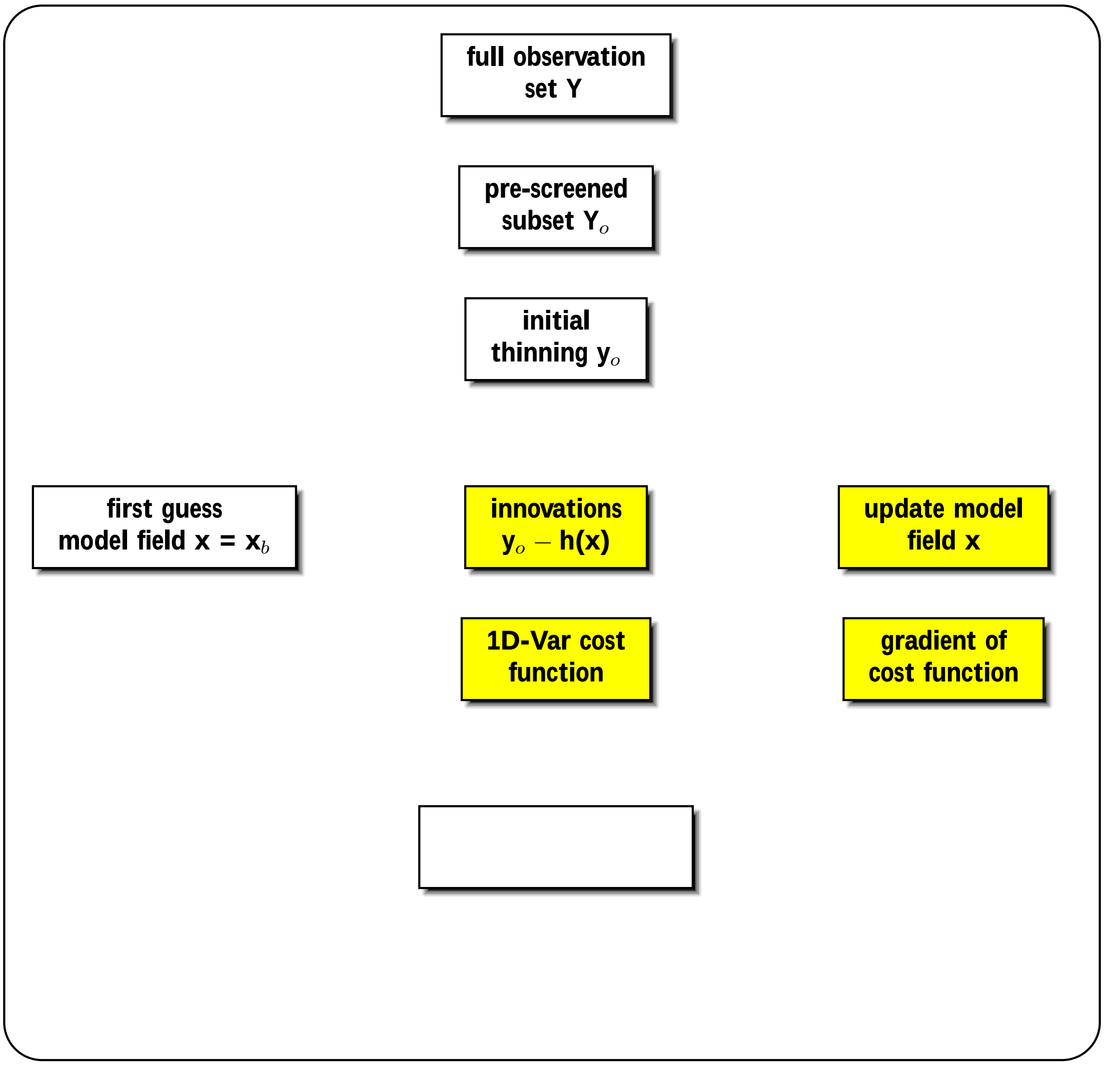
4.1.2 Processing

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,

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

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

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 O□ce 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 observat

4.2 Application of the Desroziers' diagnostic

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

$$\mathbb{E}\left[\mathbf{d}_a^o(\mathbf{d}_b^o)^T\right] \approx \mathbf{R} \tag{4.2}$$

where $\mathbf{d}_b^o = \mathbf{y} - h(\mathbf{x}^b)$ is the background innovation vector and $\mathbf{d}_a^o = \mathbf{y} - h(\mathbf{x}^a)$ is the analy-

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 $\mathbf{d}_\mathbf{a}^\mathbf{o}$ innovation statistics for the 4D-Var assimilation. This process is shown in Figure 4.3.

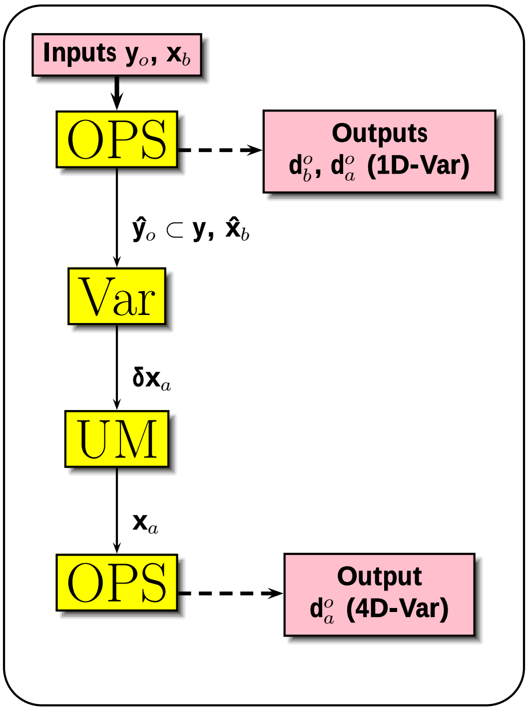


Figure 4.3: Met Office assimilation process: $\mathbf{y}_o$ is the initial observation set, $\hat{\mathbf{y}}_o$ is the quality control observation subset, x is the background, x̂ is the quality control background, δx is the analysis increment, and x is the analysis. The yellow boxes represent assimilation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

Clearly we only want to generate our statistics from those observations that are deemed

Backgro

Observation

IASI Sp

Diagı

The error correlation matrix can be determined easily from the error covariance matrix using the identity $\mathbf{R} = \mathbf{D}^{\ /2}\mathbf{C}\mathbf{D}^{\ /2}$ 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

Diagnos

IASI Sp

Di ı

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

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

Chapter 5

Information content studies in a 3D-Var framework

tions. 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 o approximate error covariance matrix, R_f, is used in the assimilation process. We will use the measures of Shannon Information Content (

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

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

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

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

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 $\mathbf{R_f}$ as the true error covariance matrix. The analysis error covariance matrix is then defined to be

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

and hence the information measures can be written as

$$\begin{aligned} \mathbf{SIC}^{(3)} &= \frac{1}{2}\ln|\mathbf{B H^T R_f^{-}\ H + I}|, && (5.13)\\ \mathrm{dof}_{\mathbf{S}}^{(3)} &= \mathbf{n} - \mathrm{trace}(\mathbf{H^T R_f^{-}\ HB + I})^{-} \end{aligned}$$

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 obser derived in [20],

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

where r is the distance between observation stations, L is the length scale and R_0

(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;

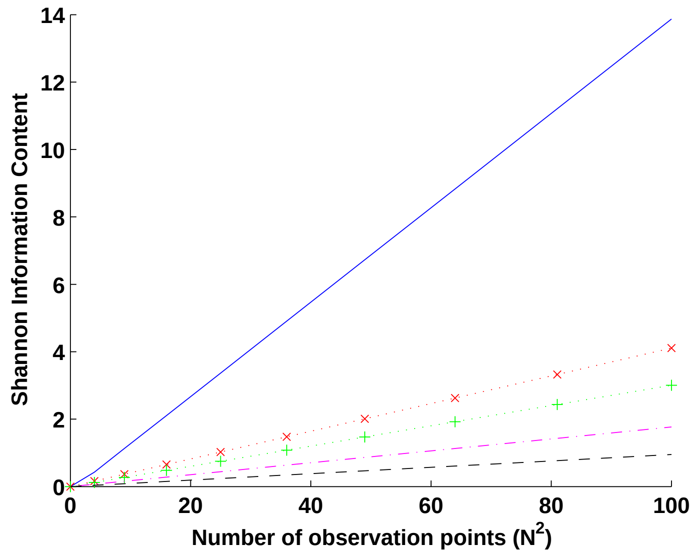


Figure 5.2: SIC for different grid sizes using R (blue line), diagonal approximation (2) (red crossed line), and diagonal approximation (3) with d = 2 (green plus line), d = 4 (pink dot-dashed line) and d = 8 (black double-dashed line).

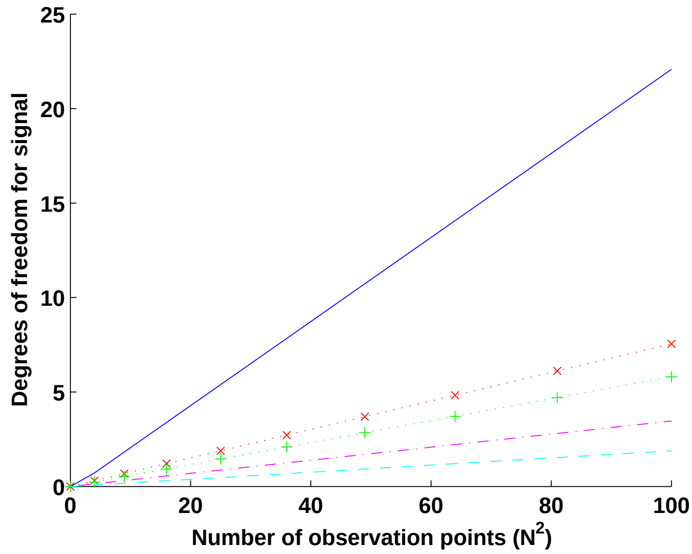
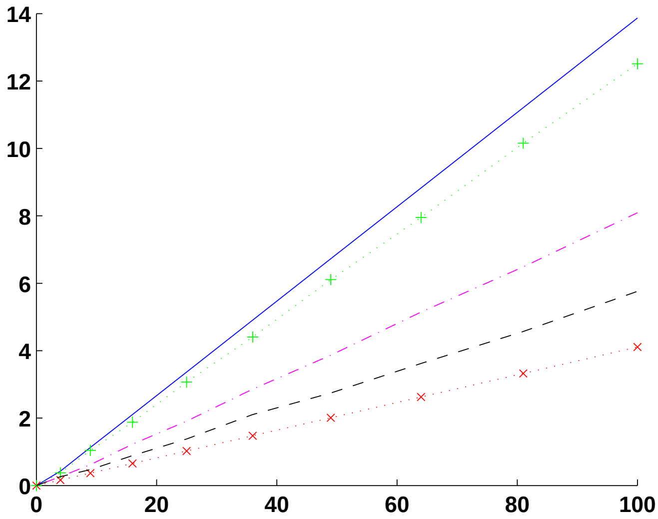
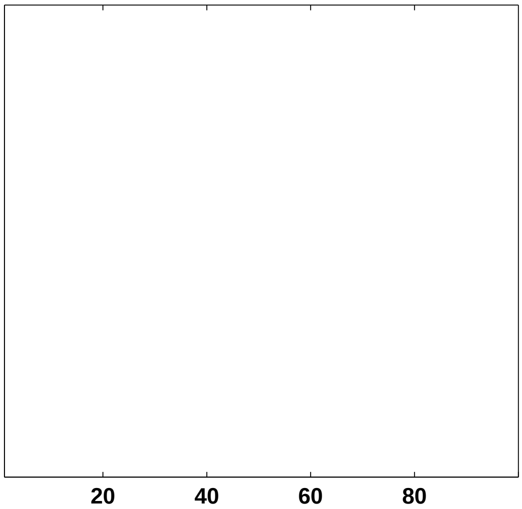
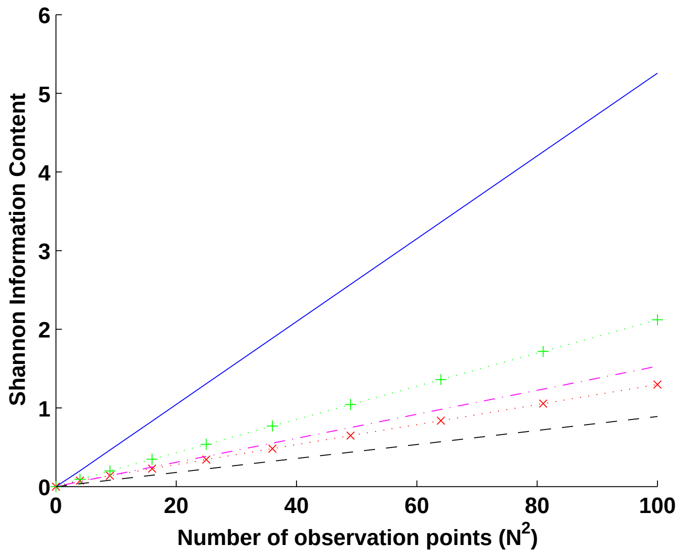


Figure 5.3: dof for different grid sizes using R (blue line), diagonal approximation i63061.3l(002(l).e)0]3362720.155202

mental 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).




20
40
60
80



Shannon Information Content
Number of observation points (N^2)
0
1
2
3
4
5
6
0
20
40
60
80
100

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.

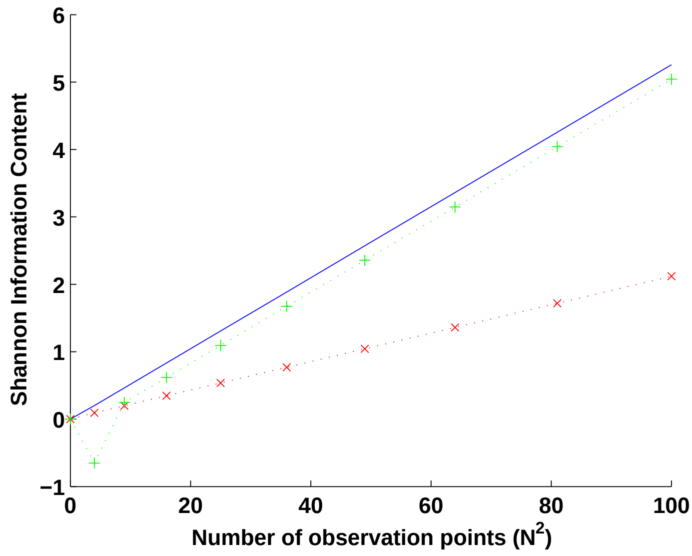


Figure 5.11: SIC for different grid sizes using R (blue line), diagonal approximation (2) (red x line) and a circulant approximation (green + line).

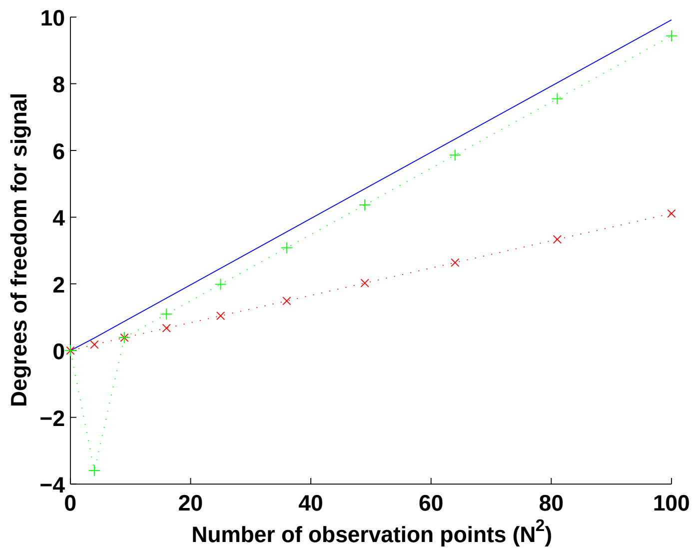
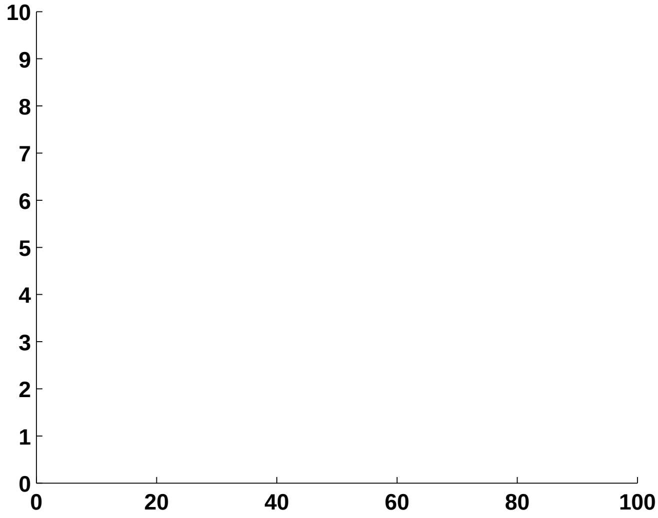


Figure 5.12: dof_s for different grid sizes using R (blue line), diagonal approximation (2) (red x line) and a circulant approximation (green + line).

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 RRR


10
9
8
7
6
5
4
3
2
1
0
0
20
40
60
80
100

can lead to misleading and inflated information content values, and subsequently incorcorrelation 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** 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

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. optseu31393o186(d)-332.945(t)]T 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

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

$$\frac{Du}{Dt} + \frac{\partial \varphi}{\partial x_D} = -g\frac{\partial h_0}{\partial x_D}, \tag{6.1}$$

$$\frac{D(\ln\varphi)}{Dt} + \frac{\partial u}{\partial x_D} = 0, \tag{6.2}$$

where

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

and $h_0 = h_0(x_D$

hence the ‘shallow’ nature of the problem. A schematic of one-dimensional shallow water

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.

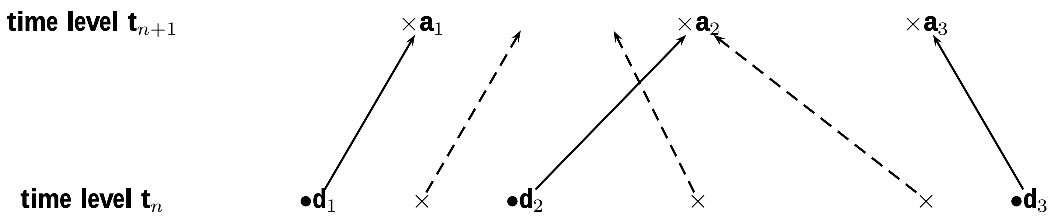


Figure 6.2: A semi-Lagrangian scheme with departure points (d , d_2, d_3) and arrival points (a , a_2, a_3). The paths taken by water parcels from the determined departure points are shown by the full lines, and the paths taken by water parcels from the regular grid points are shown by the dashed lines.

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

$$\frac{u_a^{n+} - u_d^n}{\Delta t} + (1-\alpha)\left(\frac{\partial\varphi}{\partial x_D} + g\frac{\partial h_o}{\partial x_D}\right)_d^n + \alpha\left(\frac{\partial\varphi}{\partial x_D} + g\frac{\partial h_o}{\partial x_D}\right)_a^{n+}$$

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 ver

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 erro vious 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 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 oncex 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 $(\lambda_k, v_k), k = 1, \ldots, K$ 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 $R_E^{-1} x$ and $x^T R_E^{-1} x$ needed for the calculation of the cost function and its gradient:

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

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

where R_E^{-1} is the ED matrix inverse, x is the incremental innovation vector, $D = \sigma^2 I$ is the diagonal matrix of the error variances, $s_k = v_k^T x$ is the dot product of v_k and x, and v_{ik} is the i^{th} component of the k^{th} eigenvector. As with the Markov matrix, the expression for $x^T R_E^{-1} x$ is unnecessary if $R_E^{-1} x$ 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 perfcamdsi4919e

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 tolerance $\square_o$, the solution accuracy increases. Setting $\square_o = 0.01$, we computed the normalised gradient and found it to be of order 10^{-}

which rearranged gives

$$\chi(\alpha) \equiv \frac{J(\mathbf{x}_0 + \alpha\delta\mathbf{x}_0) - J(\mathbf{x}_0)}{\alpha\delta\mathbf{x}_0^T \nabla J(\mathbf{x}_0)} = 1 + O(\alpha) \tag{6.9}$$

where α is a small scalar and $\delta\mathbf{x}_0 = \frac{\nabla J}{\|\nabla J\|}$ 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 cost. Equations specifying their use in the data assimilati

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

- **Which matrix approximation is the most robust to changes in the true error cor-**

The height of the obstacle in the fluid is given by

$$h_o(x_D) = \begin{cases} h_C \left(1 - x_D^2 \right. \\ \end{cases}$$

in preference to a Gaussian structure because its distribut covariance matrix $\mathbf{R_t}$. 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

$$\left\| \bar{\mathbf{x}}_{\mathbf{R}} - \mathbf{x}^{*} \right\|_2 \tag{7.3}$$

where $\mathbf{x}^{*}$ is the true solution of the original model run from which the observations are sampled, and $\bar{\mathbf{x}}_{\mathbf{R}}$ is the converged solution to the assimilation problem using imperfect observations when the approximation $\mathbf{R_f}$ 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

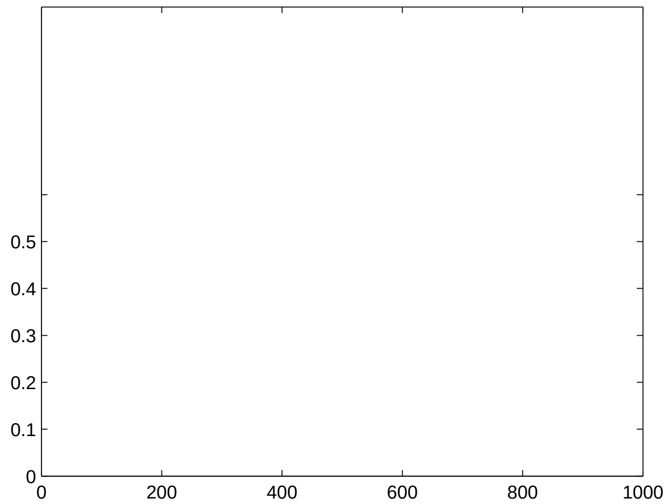
$$\left\| \bar{\mathbf{x}}_{\mathbf{R}} - \bar{} \right.$$

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 performf48.2402TLT[(4)026n331218(c)-0.230491190.04923542lrated eca90L4(i)-0440.210368(r)-0.210368(


0.5
0.4
0.3
0.2
0.1
0
0
200
400
600
800
1000

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 = 0.1m to the Markov

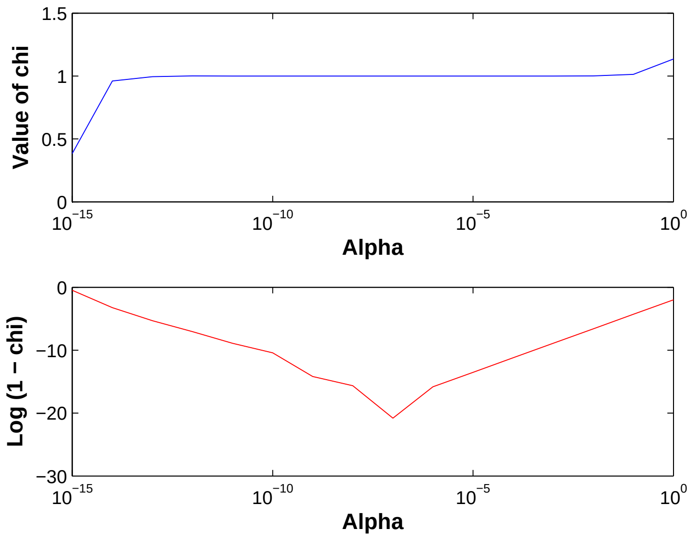


Figure 7.7: Gradient test for a Markov approximation to a Markov error covariance matrix

Approximation	E1: $\|\bar{x}_R - x^t\|_2$	$\|\bar{x}_R - \bar{x}_R\|_2$	E2 (%)
Truth	0.20	0	0
Diagonal	0.30	0.23	7.2
2 $\times$ Diagonal	0.31	0.23	7.2
4 $\times$ 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 ($\|x\|_2$ = 3.20)

ing the true error covariance matrix, i.e, a Markov matrix with length scale L_R = 0.1m, produces the smallest analysis errors; the percentage error E2 is zero for this matrix because $R_t = R_f$

Approximation	E1: □x

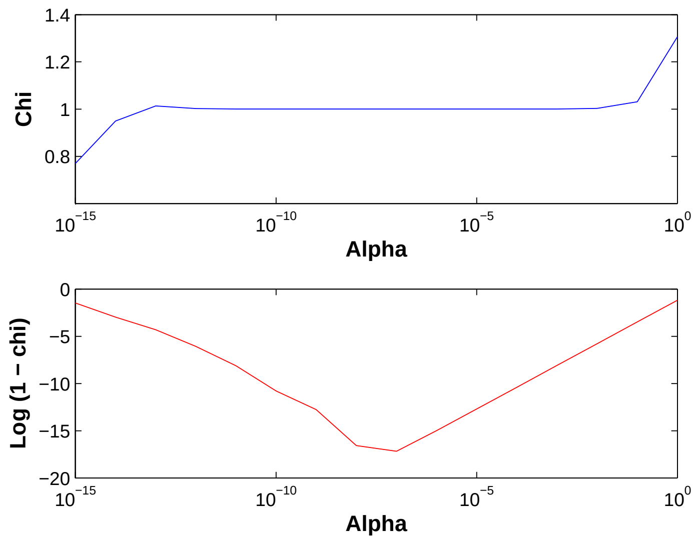


Figure 7.8: Gradient test for a diagonal approximation to a SOAR error covariance matrix

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$\times$ diagonal approximation is used. Inflated diagonal approximations perform slightly worse than a simple diagonal approximation; this is in line with the information content results in Chapter 5, when the background errors were uncorrelated.

Approximation	E1: $\left\|\bar{\mathbf{x}}_{\mathbf{R}} - \mathbf{x}^{*}\right\|_2$	$\left\|\bar{\mathbf{x}}_{\mathbf{R}} - \bar{\mathbf{x}}_{\mathbf{R}}\right\|_2$	E2 (%)
Truth	0.11	0	0
Diagonal	0.31	0.28	8.8
2 $\times$ Diagonal	0.32	0.29	9.1
4 $\times$ 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 ($\|\bar{\mathbf{x}}\|_2$ = 3.19)

Approximation	E1: $\left\|\bar{\mathbf{x}}_{\mathbf{R}} - \mathbf{x}^{*}\right\|_2$	$\left\|\bar{\mathbf{x}}_{\mathbf{R}} - \bar{\mathbf{x}}_{\mathbf{R}}\right\|_2$	E2 (%)
Truth	0.57	0	0
Diagonal	3.36	3.32	5.3
2 $\times$ Diagonal	3.59	3.55	5.7
4 $\times$ 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 ($\|\bar{\mathbf{x}}\|_2$ = 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

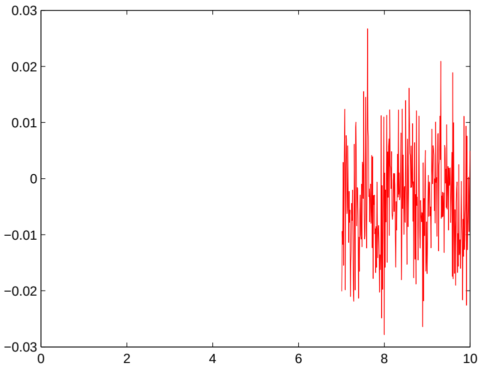
aeoupae eigenpairs represent 99% of td on o a Markov error covariance structure in Section 7.2. This is because, for a SOAR error covariance matrix, more uncertainty is represented using the same number of eigenpairs; as demonstrated in the steeper gradient in Figure 7.6.

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



0.03
0.02
0.01
0
−0.01
−0.02
−0.03
0
2
4
6
8
10

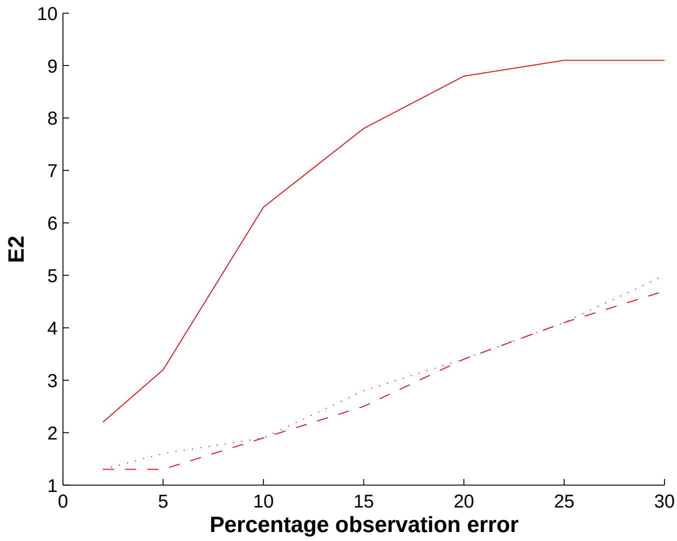


Figure 7.11: Plot of E2 against level of observation noise for u field. The solid line is for the diagonal approximation, the dashed line for the ED approximation with k = 50 and the dotted line for the Markov approximation with L = 0.05m.

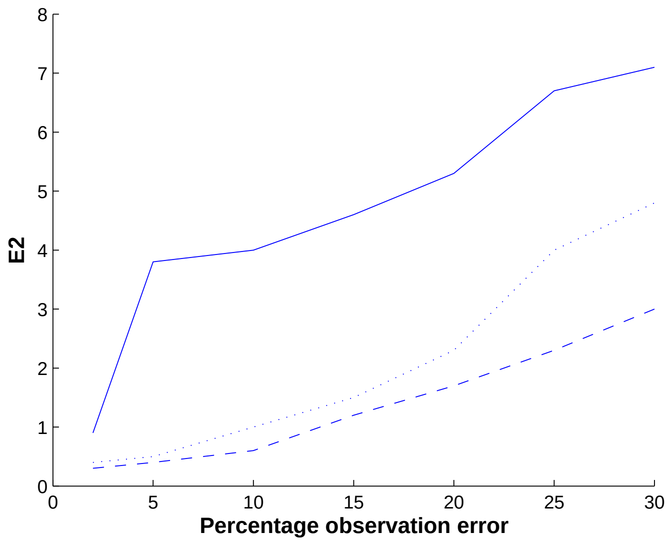


Figure 7.12: As in Figure 7.11 but for φ field.

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

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

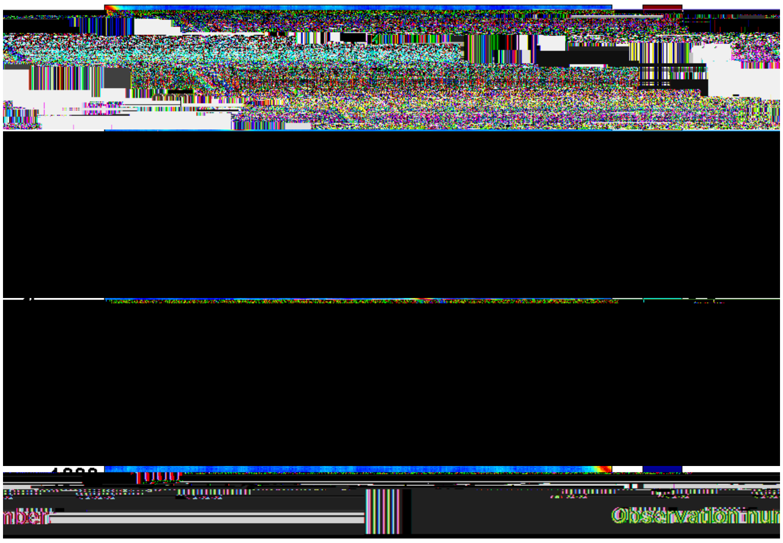
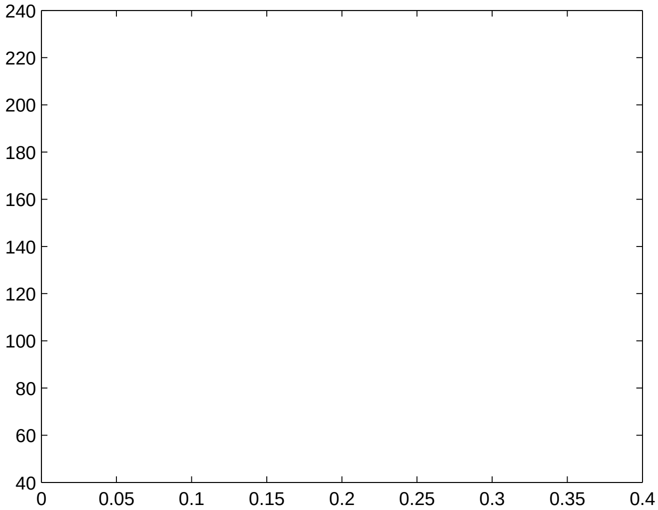


Figure 7.13: Diagnosed observation error correlation matrix.



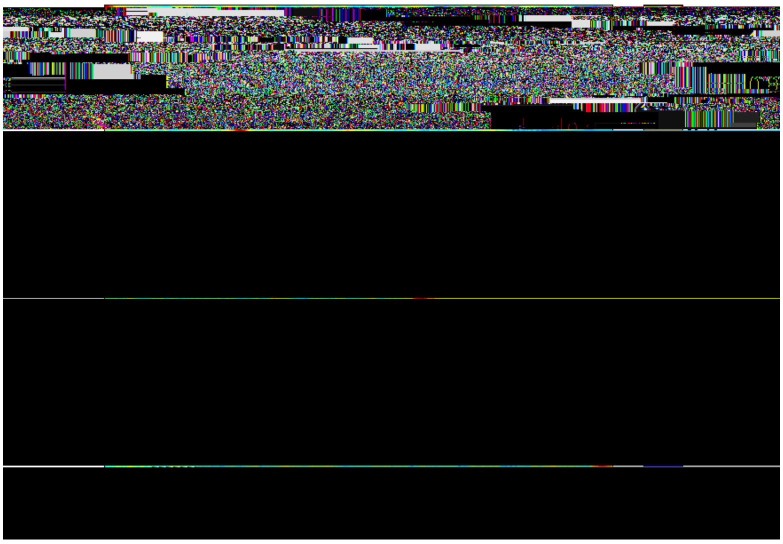


Figure 7.16: Diagnosed observation error correlation matrix using 100 observation sets.

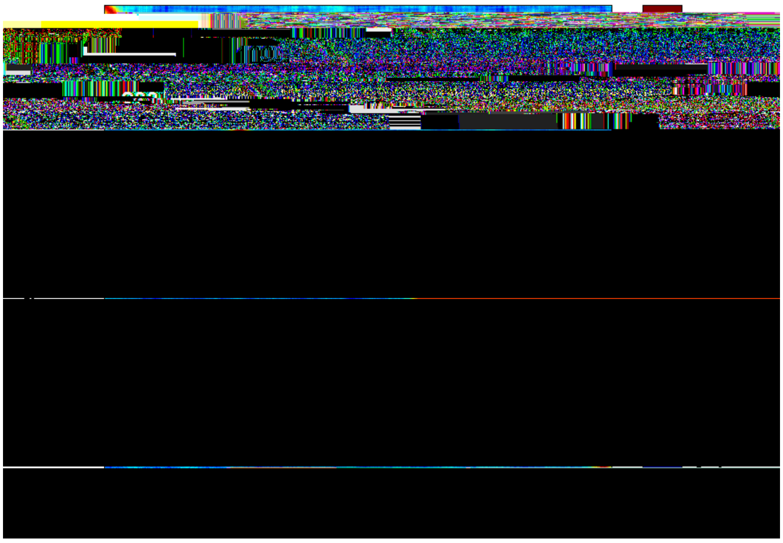


Figure 7.17: Diagnosed observation error correlation matrix using 500 observation sets.

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 $\mathbf{C_M}$. As in Figure 7.14 we vary the length scale to find the best fit to the diagnosed data. The smallest value of $\|\mathbf{C} - \mathbf{C_M}\|_F$ 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.

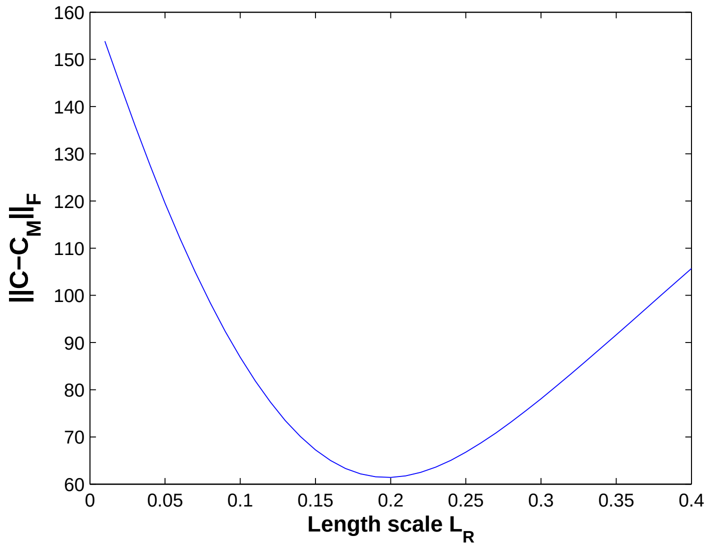


Figure 7.18: Frobenius norm of the difference between the diagnosed matrix **C** and a Markov approximation $\mathbf{C}_M$ with length scale L_R.

7.4.4 Summary

In this section we have shown new and original results on the d 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 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

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 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 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

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

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 [illegible] 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

- **By choosing a suitable matrix approximation it is feasible to cheaply include some**

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 viocovariance 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, 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 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

MetDB channel	OPS index	Var index	Central wave

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

Consider a state vector $\mathbf{x}_0$ at time 0, whose true value is $\mathbf{x}_t$ and whose background estimate is $\mathbf{x}_b$;

$$\mathbf{x}_t = \mathbf{x}_b + \epsilon^b,$$

where ϵ^b is the background error. The state vector can be evolved forward to time i under the tangent linear model $\mathbf{M}(t_i, t_0) = \mathbf{M}_i \mathbf{M}_{i-} \ldots \mathbf{M}_2 \mathbf{M}$, i.e, $\mathbf{x}_i = \mathbf{M}(t_i, t_0, \mathbf{x}_0)$. Consider m observations at different times, where the observations are related to the state vector through a forward model h,

$$
\begin{aligned}
\mathbf{y} &= h(\mathbf{x}) + \epsilon^o = h(\mathbf{M}\,\mathbf{x}_t) + \epsilon^o \\
\mathbf{y}_2 &= h(\mathbf{x}_2) + \epsilon^o_2 = h(\mathbf{M}_2\mathbf{M}\,\mathbf{x}_t) + \epsilon^o_2 \\
&\vdots \\
\mathbf{y}_m &= h(\mathbf{x}_m) + \epsilon^o_n = h(\mathbf{M}_m \ldots \mathbf{M}_2\mathbf{M}\,\mathbf{x}_t) + \epsilon^o_m
\end{aligned}
$$

where y 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

$$J(x_0$$

where

$$
\begin{aligned}
\mathbf{y} &= (\mathbf{y}^{\mathbf{T}}, \mathbf{y}_2^{\mathbf{T}}, \ldots, \mathbf{y}_{\mathbf{n}}^{\mathbf{T}})^{\mathbf{T}}, \\
\hat{\mathbf{H}} &= (\mathbf{M}^{\mathbf{T}}\mathbf{H}^{\mathbf{T}}, \mathbf{M}^{\mathbf{T}}\mathbf{M}_2^{\mathbf{T}}\mathbf{H}^{\mathbf{T}}, \ldots, \mathbf{M}^{\mathbf{T}}\mathbf{M}_2^{\mathbf{T}} \ldots \mathbf{M}_{\mathbf{n}}^{\mathbf{T}}\mathbf{H}^{\mathbf{T}})^{\mathbf{T}}, \\
\square^{\mathbf{o}} &= ((\square^{\mathbf{o}})^{\mathbf{T}}, (\square_2^{\mathbf{o}})^{\mathbf{T}}, \ldots, (\square_{\mathbf{n}}^{\mathbf{o}})^{\mathbf{T}})^{\mathbf{T}}, \\
\mathbf{R} = \mathbb{E}\left[\square^{\mathbf{o}}(\square^{\mathbf{o}})^{\mathbf{T}}\right] &= \mathbb{E}\left[\begin{pmatrix}
\square^{\mathbf{o}}(\square^{\mathbf{o}})^{\mathbf{T}} & \square^{\mathbf{o}}(\square_2^{\mathbf{o}})^{\mathbf{T}} & \ldots & \square^{\mathbf{o}}(\square_{\mathbf{n}}^{\mathbf{o}})^{\mathbf{T}} \\
\square_2^{\mathbf{o}}(\square^{\mathbf{o}})^{\mathbf{T}} & \square_2^{\mathbf{o}}(\square_2^{\mathbf{o}})^{\mathbf{T}} & \ldots & \square_2^{\mathbf{o}}(\square_{\mathbf{n}}^{\mathbf{o}})^{\mathbf{T}} \\
\vdots & \vdots & \ddots & \vdots \\
\square_{\mathbf{n}}^{\mathbf{o}}(\square^{\mathbf{o}})^{\mathbf{T}} & \square_{\mathbf{n}}^{\mathbf{o}}(\square_2^{\mathbf{o}})^{\mathbf{T}} & \ldots & \square_{\mathbf{n}}^{\mathbf{o}}(\square_{\mathbf{n}}^{\mathbf{o}})^{\mathbf{T}}
\end{pmatrix}\right] \\
&= \left(\; \mathbb{E}\left[\square_{\mathrm{n}}^{\mathbf{o}}\right.\right.
\end{aligned}
$$

Appendix C: Additional gradient tests

Additional gradient tests for Chapter 7.

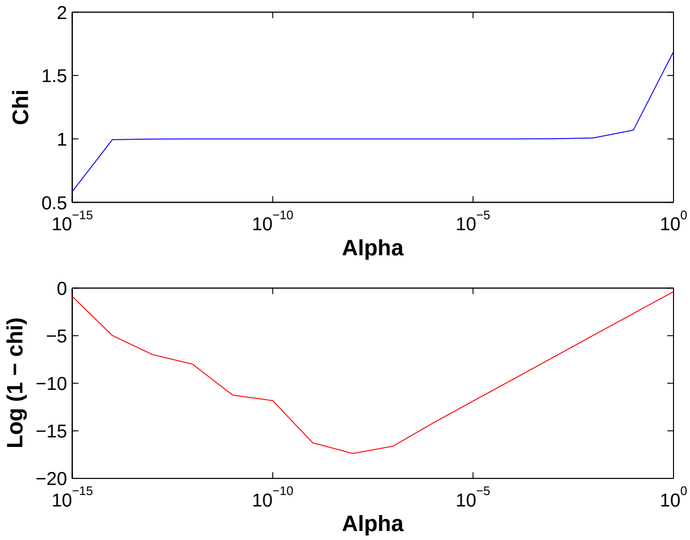


Figure 8.1: Gradient test for a diagonal approximation to a Markov error covariance matrix.

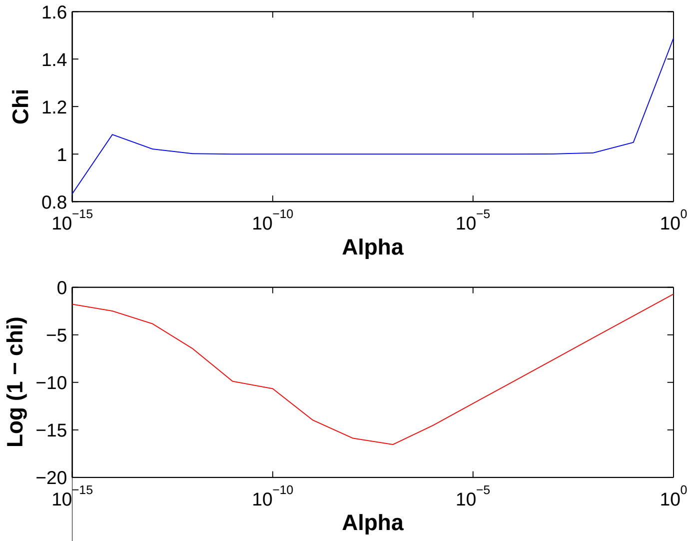
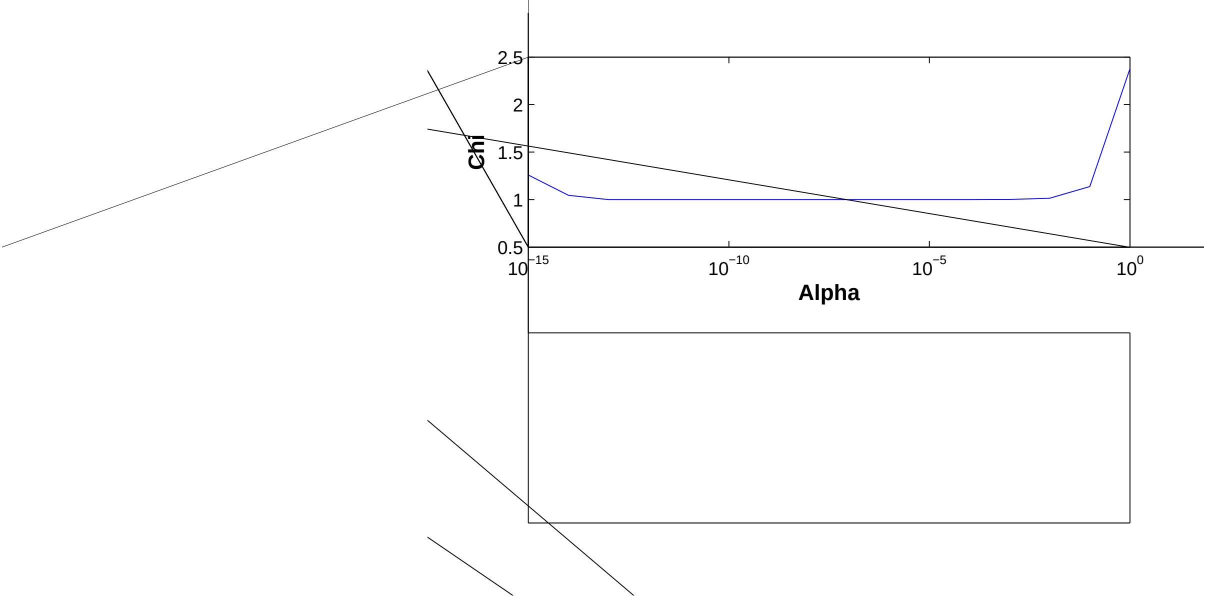


Figure 8.2: Gradient test for an ED approximation with k = 50 to a Markov error covariance matrix.



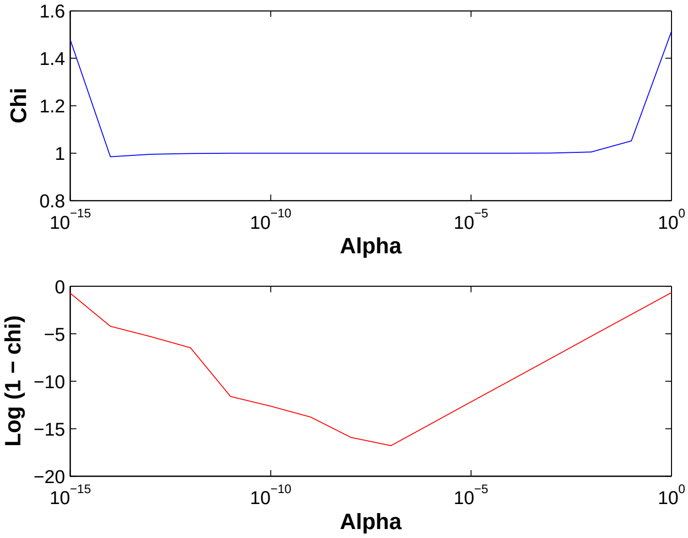


Figure 8.4: Gradient test for an ED approximation with k = 50 to a SOAR error covariance matrix.

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

[7] N. Bormann, S. Saarinen, G. Kelly, and J.-N. Thépaut. The

[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,

[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, 2001.091Tf95.88049235

[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

[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.

[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.

[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, 105:12[illegible] r d.051473(3)0.0492351(,)

[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 .