UNIVERSITY OF READING

DEPARTMENT OF MATHEMATICS AND STATISTICS

Using Observations at Different Spatial Scales in Data Assimilation for Environmental Prediction

Joanne A. Waller

Thesis submitted for the degree of

Doctor of Philosophy

April 2013

Abstract

Observations used in combination with model predictions for data assimilation can contain information at smaller scales than the model can resolve. Errors of representativity are errors that arise when the observations can resolve scales that the model cannot. Little is known about representativity errors, and consequently they are currently not correctly included in assimilation schemes. The aim of this thesis is to understand the structure of representativity error, and investigate if the assimilation can be improved by correctly accounting for representativity error.

The first approach is to use an existing method that assumes that the model state is a truncation of a high resolution truth. Using the Kuramoto-Sivishinky equation as the model, it is shown that representativity error is correlated. It is also shown that the correlation structure depends not on the number of observations but the distance between them. The representativity error is also affected by the observation type and model resolution. Using the same method representativity error is calculated for temperature and specific humidity fields from the Met Office high resolution model. This shows that representativity error is more significant for specific humidity than temperature and that representativity error is state and time dependent.

This provides motivation to combine an ensemble filter with a method that uses statistical averages of background and analysis innovations to provide an estimate of the observation error covariance matrix. Using this method it is possible to estimate a time varying observation error covariance matrix that when included in the assimilation scheme improves the analysis.

With further development of these methods it is possible that representativity errors could be correctly included in the assimilation in the context of numerical weather prediction.

Declaration

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

Joanne Waller

Acknowledgments

Firstly I would like to thank my supervisors Dr Sarah Dance, Dr Amos Lawless and Prof. Nancy Nichols. Their support and guidance had been invaluable during this research. Thanks also go to my Met Office supervisor Dr John Eyre. I would also like to thank the members of the Department of Mathematics and Statistics, Data Assimilation Research Centre and the student data assimilation group, with whom I have had many useful discussions.

Thanks must also go to my husband, parents and brother, their encouragement and timely distractions have been a great help throughout the past few years. Without their support I could not have completed this work.

Finally I acknowledge the financial support of the National Environmental Research Council (NERC) and the Met Office.

Contents

List of Figures

List of Tables

List of Symbols

Data Assimilation notation

$\mathbf{B}_n$	background error covariance matrix at time t_n
$\mathbf{C}_L$	Localised covariance matrix
$\mathbf{C}$	Covariance matrix
$\mathbf{d}^b$	Background innovation
$\mathbf{d}^a$	Analysis innovation
$E[\cdot]$	Expectation operator
$\mathbf{F}^t$	Fourier transform for the true state
$\mathbf{F}^p$	Fourier matrix for the true observations
$\mathbf{F}^p_m$	Fourier matrix for the model observations
$\mathbf{F}^m$	Fourier transform matrix for thr model state
$\mathbf{H}_n$	Linearized observation operator
$\mathcal{H}$	Non-linear observation operator
$\mathbf{I}$	Identity matrix
$\mathcal{J}$	Cost function
K^m	Highest wave number resolved by the model
K^t	Highest wave number resolved by the truth
$\mathbf{K}_n$	Kalman gain matrix
$\mathbf{L}$	Loacalisation matrix
l	Domain length
M^m	Number of model spectral coefficients

M^t	Number of true spectral coefficients
$\mathcal{M}$	non-linear forward model operator
M	Linearized forward model
N^t	Length of exact state vector
N^p	length of observation vector
N	Number of ensemble members
N^m	length of state vector
$\mathbf{P}_n^a$	Analysis covariance matrix
$\mathbf{P}_n^f$	forecast error covariance matrix
$\mathbf{R}_n$	Observation error covariance matrix at time t_n
$\mathbf{R}^H$	Forward model error covariance matrix
$\mathbf{R}^I$	Instrument error covariance matrix
$\hat{\mathbf{S}}$	spectral covariance matrix for the high resolution state
$\mathbf{S}$	covariance matrix of the high resolution state in physical space.
$\mathbf{T}$	Truncation matrix
$\boldsymbol{\Upsilon}$	Square root matrix
$\mathbf{U}$	Matrix of singular vectors
$\mathbf{W}^m$	Fourier matrix for the model observations
$\mathbf{W}^t$	True weighting matrix
$\hat{w}_k$,	k^{th} spectral coefficient of the weighting function
$\mathbf{x}_n$	state vector at time t_n
$\hat{\mathbf{x}}$	Spectral coefficients of the model state
$\hat{\mathbf{x}}^t$	Spectral coefficients of the true state
$\bar{\mathbf{x}}_n^f$	Forecast ensemble mean
$\bar{\mathbf{x}}_n^a$	Analysis ensemble mean
$\bar{\mathbf{X}}$	Ensemble mean matrix
$\mathbf{X}'^f$	Forecast ensemble perturbation matrix
$\mathbf{X}'^a$	Analysis ensemble perturbation matrix
$\mathbf{X}$	State ensemble matrix

$\mathbf{x}_n^{f,i}$	Forecast state of i^{th} ensemble member
$\mathbf{x}_n^{a,i}$	Analysis state of i^{th} ensemble member
$\mathbf{x}_n^b$	background at time t_n
$\mathbf{x}_n^a$	analysis
$\mathbf{x}_n^t$	exact state of the system at time t_n
$\mathbf{Y'}_n^f$	matrix holding the measurements of the ensemble perturbations
$\mathbf{y}^m$	Model representation of the observations
$\mathbf{y}_n$	observation vector at time t_n
$\mathbf{y}_n^t$	noise free observation vector at time t_n
$\mathbf{\Lambda}$	Matrix of singular values

γ	Inflation factor
$\boldsymbol{\epsilon}_n^I$	instrument error at time t_n
$\boldsymbol{\epsilon}_n^H$	forward mapping error at time t_n
$\boldsymbol{\epsilon}_n^o$	observation error at time t_n
$\boldsymbol{\epsilon}_n^b$	background error
$\boldsymbol{\epsilon}_n^m$	model error at time t_n

Kuramoto-Sivashinsky equation notation

a_n	Coefficient for the ETDRK4 scheme
b_n	Coefficient for the ETDRK4 scheme
c_n	Coefficient for the ETDRK4 scheme
c	localistion lengthscale.
F	Fourier transform
h	space step
I_N	Approximate integral
k	Wave number
$\mathcal{L}$	linear operator

$\mathbf{L}$	linear operator
L	Localisation function
L_∞	L infinity norm
$\mathcal{N}$	non-linear operator
$\mathbf{N}$	non-linear operator
t	time
t_n	time
$\mathbf{u}_n$	Solution to the KS equation at time t_n
$\hat{\mathbf{u}}_t$	Spectral solution to the KS equation at time t_n
u_t	First derivative wrt t
u	Solution to the KS equation
u_x	First derivative wrt x
u_{xx}	Second derivative wrt x
u_{xxxx}	Fourth derivative wrt x
v_n	Spectral solution to the KS equation at time t_n
x	Spatial domain
z	Complex number
Γ	Contour in the complex plane
Δt	Time step

Abbreviations

BLUE	Best linear unbiased estimator
DA	data assimilation
EnKF	Ensemble Kalman filter
ETKF	Ensemble transform Kalman filter
KS	Kuromoto-Sivashinsky (non-linear PDE)
KF	Kalman Filter
NWP	Numerical weather prediction
NAE	North Atlantic and Europe (Met Office NWP model)
PDE	Partial differential equation
pdf	probability density function
SVD	Singular value decomposition
UKV	UK variational (Met Office NWP model)
3D-Var	3D Variational data assimilation
4D-Var	4D Variational data assimilation

Chapter 1

Introduction

Data assimilation is the incorporation of observational data into a numerical model to produce a model state that accurately describes the observed reality [Kalnay, 2002]. It is applicable to many situations as it provides a complete set of accurate initial conditions for input into a numerical model. One situation where data assimilation is used is numerical weather prediction (NWP) [UK Met Office, 2013]. In NWP observations are either measured in-situ or are remotely sensed. Data can be remotely sensed by a network of satellites or from ground based instruments such as radar. Whether remotely sensed or measured in-situ these observations contain errors. The statistics associated with the errors in the observations are included in the data assimilation scheme in the observation error covariance matrix. Different types of error contribute to the observation error, including instrument or measurement error, errors associated with the preprocessing of observational data, errors in the observation operators that map the model space into observation space and representativity error. Representativity error [Daley, 1991], also known as representativeness error [Liu and Rabier, 2002], representation error [Oke and Sakov, 2007] or representivity error [Swinbank et al., 2003], is the error that arises when the observations can resolve scales that the model cannot [Daley, 1991]. Representativity error combined with the errors in the observation operator are known as forward model error or forward interpolation error [Lorenc, 1986].

Little is known about forward model and representativity error or how they may affect the assimilation scheme if accounted for. Previous work [Stewart et al., 2009, 2012b, Bormann et al., 2002, Bormann and Bauer, 2010, Bormann et al., 2010] has shown that for certain observation types the observation error covariance matrix contains significant correlations. It has been suggested that part of the correlation comes from representativity error rather than the instrument error or errors in the observation operator [Stewart, 2010, Weston, 2011]. Accounting for these correlated errors in the assimilation scheme may be computationally costly as the number of observations available is $O(10^7)$, so to reduce the cost it is assumed that the observation error covariance matrix is block diagonal, with each block corresponding to a different observation type. Methods such as variance inflation [Hilton et al., 2009, Whitaker et al., 2008], observation thinning [Lahoz et al., 2010] and superobbing [Daley, 1991] are used to account for the unknown and unrepresented correlation structure. Efforts are being made to find methods of reducing the cost of using correlated observation error matrices [Stewart et al., 2012a, Stewart, 2010, Healy and White, 2005]. Once these methods are in place it will be important to have accurate estimates of the covariance matrices, as these are required to obtain the optimal estimate from any data assimilation system [Houtekamer and Mitchell, 2005, Stewart et al., 2008]. It is therefore important to understand forward model and representativity error. The requirement of a better understanding of forward model error, how it can be calculated, its structure and its effect on a data assimilation system provides the motivation for this thesis. We now present in the next two sections the main aims and principal results of the thesis. We then give an overview of each chapter in section 1.3.

1.1 Aims

In this thesis we aim to use existing methods and develop our own schemes to investigate forward model error and representativity error. In particular we wish to:

- Understand what representativity error is and how it can be calculated and included in the data assimilation scheme.

- Understand the structure of representativity error and see if it may be a cause of correlations in the observation error covariance matrix.
- Understand if representativity error is a significant error.
- Determine if the inclusion of representativity error in the data assimilation scheme can improve the analysis.
- Determine if it is possible to calculate a time dependent estimate of forward model error.

1.2 Principal results

The principal new results of this thesis are:

- Representativity error is correlated and case dependent. The representativity error variance is independent of the number of available observations. The correlation structure of representativity error is dependent not on the number of observations, but the distance between them.
- Representativity error can be reduced by increasing the model resolution or increasing the observation lengthscales.
- Representativity error is more significant for humidity than temperature and varies throughout the atmosphere.
- Including representativity error in an assimilation scheme may improve the analysis.
- Using a method developed in this thesis it is possible to estimate time varying observation error covariance matrices within an ensemble transform Kalman filter.

1.3 Outline

In Chapter 2 we introduce the concept of data assimilation and the notation used throughout this thesis. We give a brief description of both variational and sequential data as-

similation. We then describe a sequential method known as the best linear unbiased estimate. The concept of ensemble data assimilation is introduced and the ensemble transform Kalman filter is discussed. We also discuss two techniques that can be used to overcome problems associated with ensemble filtering. We conclude the chapter by describing two diagnostics that can be used to assess the performance of assimilation schemes.

We introduce forward model error and representativity error in Chapter 3. We review previous methods that have been used for calculating the observation error covariance matrix and consider the results relating to the structure of these matrices. We discuss existing methods for calculating forward model error and representativity error. One of these methods developed by Daley [1993] and then by Liu and Rabier [2002] is discussed in more detail. We also discuss a method used to calculate the observation error covariance matrix, the Desroziers et al. [2005] diagnostic, in more detail.

In Chapter 4 we describe how the matrices used in the Daley [1993] method can be defined. We then present some new theoretical results that relate to the Daley [1993] method.

In Chapter 5 we calculate representativity error for the Kuramoto-Sivashinsky (KS) equation [Kuramoto, 1978, Sivashinsky, 1977]. First the KS equation is introduced. We then discuss how the KS equation can be solved numerically using the ETDRK4 method and show convergence of this method. We consider solutions of the KS equation at different resolutions. Next we use the Daley [1993] method to calculate representativity error and forward model error for the KS equation. We compare these numerical results to the theoretical results presented in Chapter 4 and draw conclusions about forward model and representativity errors.

In Chapter 6 we calculate representativity error for temperature and specific humidity using data from the Met Office high resolution model [Tang et al., 2012]. We compare data from two cases to show how case dependent representativity error is. We also show how representativity error differs between temperature and specific humidity. We repeat the experiments in Chapter 5 with the data from the Met Office model and use these results to support the results of previous chapters. We then consider how representativity

error changes at different atmospheric levels.

In Chapter 7 we develop a new approach for calculating the observation error covariance matrix and forward model error covariance matrix. We present a method that combines the Desroziers et al. [2005] diagnostic with the Ensemble Transform Kalman Filter (ETKF). First we show that a correlated observation error covariance matrix can be calculated using the Desroziers et al. [2005] diagnostic after an ETKF assimilation cycle has finished. We then introduce a rolling time window that is used to provided the samples that are required by the Desroziers et al. [2005] diagnostic to calculate the observation error covariance matrix. We then show that it is possible to calculate an observation error covariance matrix within the ETKF cycle that can then be fed back into the scheme to improve the analysis.

We summarise the work in this thesis in Chapter 8. We draw conclusions from the results seen throughout the thesis and suggest future work that may be carried out.

Chapter 2

Data Assimilation

In this chapter we describe the mathematics of data assimilation (DA) and the notation and terminology used in this thesis. We start by considering the Bayesian approach to DA and how, with some approximations, the Bayesian approach gives the equations for the variational DA method known as 3D-Var. We then describe a sequential DA system known as the Kalman filter (KF), and its ensemble version, the ensemble Kalman filter (EnKF). We then consider the ensemble transform Kalman Filter (ETKF), a deterministic ensemble Kalman filter. We also consider some of the techniques that can be used to overcome some of the problems associated with ensemble filtering. We also introduce some diagnostics that allow us to determine how well an assimilation system is performing.

2.1 Notation

We consider the non-linear dynamical system,

$$\mathbf{x}_{n+1} = \mathcal{M}_n(\mathbf{x}_n) + \boldsymbol{\epsilon}_n^m, \tag{2.1}$$

where $\mathbf{x}_n$ is the model state vector of length N^m at time t_n, $\mathcal{M}_n$ is the non-linear model operator that evolves the model state at time t_n to the model state at time t_{n+1}. The model error $\boldsymbol{\epsilon}_n^m$ at time t_n is a random vector of length N^m. Often data assimilation methods

make a perfect model assumption. Under this assumption it is assumed that the model error $\epsilon_{\mathbf{n}}^{\mathbf{m}}$ is zero.

We assume we have observations $\mathbf{y}_n$ at time t_n. These observations are related to the true state of the system, $\mathbf{x}_n^t$ by,

$$\mathbf{y}_n = \mathcal{H}_n(\mathbf{x}_n^t) + \boldsymbol{\epsilon}_n^o, \tag{2.2}$$

where $\mathbf{y}_n$ is the observation vector of length N^p at time t_n. $\mathcal{H}_n$ is the possibly non-linear observation operator, a mapping to $\mathbb{R}^{N^p}$, that maps the true state into observation space. The observation error $\boldsymbol{\epsilon}_n^o$ is a random vector of size N^p at time t_n. The mean value of $\boldsymbol{\epsilon}_n^o$ is assumed to be zero. In practice it may be necessary to achieve this by using a preprocessing step to bias correct the data. The observation error covariance matrix is $\mathbf{R}_n = E[\boldsymbol{\epsilon}_n^o \boldsymbol{\epsilon}_n^{oT}]$ where E[$\cdot$] denotes expected value.

As well as observations, we have available from the numerical model the forecast at the current time. This model prediction of the state $\mathbf{x}_n^b$ is known as the background. The background state is an approximation to the true state of the atmosphere $\mathbf{x}_n^t$ such that

$$\mathbf{x}_n^b = \mathbf{x}_n^t + \boldsymbol{\epsilon}_n^b, \tag{2.3}$$

where the random vector $\boldsymbol{\epsilon}_n^b$ is known as the background error. The expected value of this background error is assumed to be zero. The covariance of the background error $\mathbf{B}_n = E[\boldsymbol{\epsilon}_n^b \boldsymbol{\epsilon}_n^{bT}]$ is the background error covariance matrix. This matrix can be static and reflect a climatological error variance. However, in some assimilation schemes it is assumed that this matrix is flow dependent.

When flow dependence is assumed, the background error variance is often denoted as $\mathbf{P}_n^f$ and known as the forecast error covariance matrix. It is assumed that the background and observation errors are mutually uncorrelated.

2.2 Data assimilation

Data assimilation techniques combine observations $\mathbf{y}_n$ at time t_n with a model prediction of the state, the background $\mathbf{x}_n^b$, weighted by their respective errors, to provide a best estimate of the state $\mathbf{x}_n^a$, known as the analysis.

There are many types of data assimilation but in general they are classified as either sequential or variational [Talagrand, 1997]. Sequential methods solve explicitly the equations that give the best state estimate, whereas variational methods minimise a cost function to implicitly solve the problem. We shall consider both types of method. However, we first derive the state estimation problem using Bayes' theorem [Lorenc, 1986]. Under the assumption of linearity this can be used to derive a sequential scheme to find the best analysis.

Theorem 2.2.1. - Bayes Theorem

The probability of A given B is,

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}. \tag{2.4}$$

If A is the model state and B the observations then Bayes Theorem [Bolstad, 2007] tells us that the probability $p(.)$ of the model state given the observations is equal to the probability of the observations given the model state multiplied by the probability of the model state divided by the probability of the observation. Here, $p(B)$ is a normalisation factor and is independent of the model state.

For most DA schemes it is assumed that the probability density functions (pdfs) are Gaussian. Following this assumption allows the prior pdf of the model state to be written as,

$$p(\mathbf{x}_n) \propto \exp\{-\frac{1}{2}(\mathbf{x}_n - \mathbf{x}_n^b)^T \mathbf{B}_n^{-1}(\mathbf{x}_n - \mathbf{x}_n^b)\}, \tag{2.5}$$

and the likelihood (the probability of the observation occurring given the model state at time t_n) is written as,

$$p(\mathbf{y}_n|\mathbf{x}_n) \propto \exp\{-\frac{1}{2}(\mathbf{y}_n - \mathcal{H}_n(\mathbf{x}_n))^T \mathbf{R}_n^{-1}(\mathbf{y}_n - \mathcal{H}_n(\mathbf{x}_n))\}. \tag{2.6}$$

Now substituting this into Bayes' Theorem and assuming that errors in observations and background are independent we obtain an expression for the posterior pdf of the state given the observations,

$$p(\mathbf{x}_n|\mathbf{y}_n) = \exp\{-\frac{1}{2}(\mathbf{x}_n - \mathbf{x}_n^b)^T \mathbf{B}_n^{-1}(\mathbf{x}_n - \mathbf{x}_n^b) - \frac{1}{2}(\mathbf{y}_n - \mathcal{H}_n(\mathbf{x}_n))^T \mathbf{R}_n^{-1}(\mathbf{y}_n - \mathcal{H}_n(\mathbf{x}_n))\}. \tag{2.7}$$

The estimate for the analysis is determined by maximising the value of $p(\mathbf{x}_n|\mathbf{y}_n)$, known as the maximum a posteriori (MAP) estimate. This is equivalent to finding the minimum variance by minimising the cost function

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

This gives an estimate of the analysis that is based on the distance between the solution and the background weighted by the error in the background and the distance between the observations and the solutions weighted by the error in the observations.

2.2.1 The best linear unbiased estimator and 3D-Var

One of the simplest forms of data assimilation aims to solve equation (2.8) explicitly. The analysis at a given time t_n is given by,

$$\mathbf{x}_n^a = \mathbf{x}_n^b + \mathbf{K}_n(\mathbf{y}_n - \mathcal{H}_n(\mathbf{x}_n^b)), \tag{2.9}$$

where $\mathbf{K}_n = \mathbf{B}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{B}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}$, a matrix of size $N^m \times N^p$, is known as the Kalman gain matrix. $\mathbf{H}_n$ is the observation operator linearised about the background state. The analysis is then forecast using the model given in equation (2.1) to provide a background

The BLUE Algorithm [Swinbank et al., 2003]
1. Calculate the analysis: $$\mathbf{x}_n^a = \mathbf{x}_n^b + \mathbf{K}_n(\mathbf{y}_n - \mathcal{H}_n\mathbf{x}_n^b),$$ where $\mathbf{K}_n = \mathbf{B}_n\mathbf{H}_n^T(\mathbf{H}_n\mathbf{B}_n\mathbf{H}_n^T + \mathbf{R}_n)^{-1}$. 2. Forecast the analysis using $$\mathbf{x}_{n+1}^b = \mathcal{M}_n(\mathbf{x}_n^a)$$ to obtain a background state at the next time.

Table 2.1: A simple BLUE algorithm.

at the new time. These analysis and forecast steps can be applied sequentially to give a scheme known as the best linear unbiased estimator (BLUE) [Lewis et al., 2006]. If $\mathcal{H}$ is linear the exact minimiser of equation (2.8) is found. However, if the observation operator $\mathcal{H}$ is non-linear it is necessary to define a linear version $\mathbf{H}$ and the solution obtained is only an approximate solution. The BLUE algorithm is summarised in Table 2.1.

For low dimensional systems the BLUE is a useful method to use as a starting point to help understand data assimilation. Before using more complex assimilation schemes operational NWP centres used an approximate BLUE in the form of optimal interpolation (OI) [Lorenc, 1981] and analysis correction (AC) [Lorenc et al., 1991].

Rather than solving equation (2.8) explicitly it is possible to use a numerical approach. The cost function can be minimised over several iterations using a gradient descent algorithm to obtain an estimate for the analysis. This analysis approach is known as 3D variational data assimilation (3D-Var). 3D-Var is a more effective method for large systems and has been used in operational NWP centres [Lorenc et al., 2000]. However, another assumption in 3D-Var is that all observations are valid at one time, rather than over a time window around the assimilation time. To take into account the time dependence of the observations the equations of 3D-Var must be extended. The extended method is known as 4D variational assimilation (4D-Var) [Sasaki, 1970]. 4D var makes use of the dynamical model to allow observations to be assimilated at the correct time.

2.2.2 The Kalman filter

So far we have encountered both variational and sequential methods. All the methods considered so far assume the the error in the background and observations are fixed in time. In many cases this is a poor assumption as the background error is related to the model state and therefore will evolve as the model evolves. We now consider a method known as the Kalman filter introduced by Kalman [1960] and Kalman and Bucy [1961], which as well as updating the state also updates the error covariance matrix $\mathbf{P}^f$. The Kalman filter requires a linear dynamical model $\mathbf{M}_n$ and linear observation operator $\mathbf{H}_n$. The Kalman filter begins with the same analysis update as the BLUE given in equation (2.9). This is followed by an update to the analysis error covariance,

$$\mathbf{P}_n^a = (\mathbf{I} - \mathbf{K}_n\mathbf{H}_n)\mathbf{P}_n^f. \tag{2.10}$$

Both the analysis and background covariance are then forecast. The analysis is forecast using equation (2.1) and the covariance is updated using

$$\mathbf{P}_{n+1}^f = \mathbf{M}_n\mathbf{P}_n^a\mathbf{M}_n^T + \mathbf{Q}_n, \tag{2.11}$$

where $\mathbf{Q}_n$ is the model error covariance matrix. If the model is assumed perfect the matrix $\mathbf{Q}$ can be omitted. This scheme is also applied sequentially to give analysis and background error covariances at times when observations are available. We summarise the Kalman filter algorithm in Table 2.2.

The Kalman filter is a useful data assimilation scheme as it uses observations at the time they are available. It is the optimal linear filter in terms of minimum variance. It provides an unbiased estimate of the forecast and analysis as well as estimates of a flow dependent background matrix [Jazwinski, 1970]. However, it is restricted to use with linear dynamical systems. The theory of the Kalman filter can be extended to take account of non-linear models. This leads to methods such as the extended Kalman filter [Gelb, 1974] and the ensemble Kalman filter [Evensen, 1994], a form of ensemble data assimilation.

The Kalman Filter Algorithm [Kalman and Bucy, 1961]
1. Calculate the analysis: $$\mathbf{x}_n^a = \mathbf{x}_n^b + \mathbf{K}_n(\mathbf{y}_n - \mathbf{H}_n\mathbf{x}_n^b),$$ where $\mathbf{K}_n = \mathbf{P}_n^f\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n^f\mathbf{H}_n^T + \mathbf{R}_n)^{-1}$.
2. Update the background error: $$\mathbf{P}_n^a = (\mathbf{I} - \mathbf{K}_n\mathbf{H}_n)\mathbf{P}_n^f.$$
3. Forecast the analysis to obtain a background state at the next time: $$\mathbf{x}_{n+1}^b = \mathcal{M}_n\mathbf{x}_n^a + \boldsymbol{\epsilon}_n^m.$$
4. Forecast the background error: $$\mathbf{P}_{n+1}^f = \mathbf{M}_n\mathbf{P}_n^a\mathbf{M}_n^T + \mathbf{Q}_n.$$

Table 2.2: The Kalman filter algorithm.

2.3 Ensemble data assimilation

The exact state of the atmosphere cannot be accurately determined and therefore it is likely that initial conditions supplied for a forecast will contain errors. Even small perturbations in the initial conditions can lead to a large change in the forecast, so it is important that uncertainty in the initial conditions is represented. One way to represent the uncertainty in the initial conditions is to represent the prior distribution of the initial state with a number of different initial conditions. Each of these different states is known as an ensemble member. Forecasting each of the ensemble members results in an ensemble of forecasts that must be combined with observations. This has led to the development of ensemble data assimilation schemes. One ensemble data assimilation scheme is known as the ensemble Kalman filter (EnKF).

The ensemble Kalman filter is an ensemble data assimilation scheme based on the Kalman filter summarised in Table 2.2. It was first introduced by Evensen [1994] and many forms of ensemble Kalman filter have been developed, for example Tippett et al. [2003], Houtekamer and Mitchell [1998], Evensen [2003], Burgers et al. [1998], Anderson [2001]. These methods can be split into two categories; deterministic filters and stochastic filters. Stochastic filters make use of a set of perturbed observations which are required to maintain the correct

statistics of the filter [Burgers et al., 1998, Lewis et al., 2006]. Deterministic filters do not require these perturbed observations, therefore no extra errors in the observations are introduced. One set of deterministic filters are known as square root filters. We now introduce the general form of square root filters. We then describe the ensemble transform Kalman filter (ETKF).

2.3.1 Notation

We first define the notation used and then describe the method. At time t_n we have an ensemble, a statistical sample of N state estimates $\left\{x_n^i\right\}$ for $i = 1 \dots N$. These ensemble members can be stored in a state ensemble matrix $\mathbf{X} \in \mathbb{R}^{N^m \times N}$ where each column of the matrix is a state estimate for an individual ensemble member,

$$\mathbf{X}_n = \begin{pmatrix} \mathbf{x}_n^1 & \mathbf{x}_n^2 & \dots & \mathbf{x}_n^N \end{pmatrix} \tag{2.12}$$

From this ensemble it is possible to calculate the ensemble mean,

$$\bar{\mathbf{x}}_n = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_n^i, \tag{2.13}$$

which can be stored in each column of the ensemble mean matrix $\bar{\mathbf{X}}$. Subtracting the ensemble mean matrix from the state ensemble matrix gives the ensemble perturbation matrix,

$$\begin{aligned} \mathbf{X}'_n &= \mathbf{X}_n - \bar{\mathbf{X}}_n & (2.14) \\ &= \begin{pmatrix} \mathbf{x}_n^1 - \bar{\mathbf{x}}_n & \mathbf{x}_n^2 - \bar{\mathbf{x}}_n & \dots & \mathbf{x}_n^N - \bar{\mathbf{x}}_n \end{pmatrix} & (2.15) \end{aligned}$$

This allows us to write the ensemble covariance matrix as

$$\mathbf{P}_n = \frac{1}{N-1} \mathbf{X}'_n {\mathbf{X}'_n}^T. \tag{2.16}$$

Here we divide by $N-1$ rather than N to give the unbiased covariance estimate [Reichmann, 1962].

We now consider how these ensembles members are used in the assimilation algorithm. Given an ensemble at time t_n, the first step (Table 2.3, step 1) is to forecast each ensemble member using the full non-linear model,

$$\mathbf{x}_n^{f,i} = \mathcal{M}_n(\mathbf{x}_n^{a,i}) \tag{2.17}$$

Using equations (2.13) and (2.16) we can calculate the the ensemble mean and forecast error covariance matrix, Table 2.3, step 2. We then use these in the update steps, Table 2.3, steps 3 - 6.

We update the ensembles by first updating the ensemble mean. We define the matrix containing the mapping of the ensemble perturbations into observation space, of size $N^p \times N^m$, as

$$\mathbf{Y'}_n^f = \mathbf{H}_n \mathbf{X'}_n^f, \tag{2.18}$$

if $\mathbf{H}_n$ is linear and

$$\mathbf{Y'}_n^f = \mathbf{Y}_n^f - \bar{\mathbf{Y}}_n^f, \tag{2.19}$$

where $\mathbf{Y}_n^f = \mathcal{H}_n(\mathbf{X'}_n^f)$ if $\mathcal{H}_\mathbf{n}$ is non-linear. It is also possible to restrict the observation operator to be linear and deal with the non-linearity using an augmented state [Evensen, 2003]. Using $\mathbf{Y'}_n^f$ and the invertible matrix $\mathbf{S}_n = \mathbf{Y'}_n^f \mathbf{Y'}_n^{fT} + \mathbf{R}_n$ of size $(N^p \times N^p)$ we can update the ensemble mean using,

$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^f + \mathbf{K}_n(\mathbf{y}_n - \mathbf{H}_n \bar{\mathbf{x}}_n^f), \tag{2.20}$$

where $\mathbf{K}_n$ is the Kalman gain $\mathbf{K}_n = \mathbf{X'}_n^f \mathbf{Y'}_n^{fT} \mathbf{S}_n^{-1}$ of size $N^m \times N^p$.

We now move onto the ensemble perturbation matrix update, which also gives information on the analysis error covariance matrix. If the observation operator is linear we wish to

update the covariance matrix as the Kalman filter covariance update. That is

$$\mathbf{X'}^a_n\mathbf{X'}^{aT}_n = \mathbf{X'}^f_n(\mathbf{I} - \mathbf{Y'}^{fT}_n\mathbf{S}_n^{-1}\mathbf{Y'}^f_n)\mathbf{X'}^{fT}_n. \tag{2.21}$$

Rather than calculate this explicitly the analysis perturbations are calculated as

$$\mathbf{X'}^a_n = \mathbf{X'}^f_n\boldsymbol{\Upsilon}_n, \tag{2.22}$$

where $\boldsymbol{\Upsilon}_n$ is the square root of $(\mathbf{I} - \mathbf{Y'}^{fT}_n\mathbf{S}_n^{-1}\mathbf{Y'}^f_n)$. The choice of this square root is not unique. We now consider the ETKF [Bishop et al., 2001] which describes one method of calculating this square root.

2.3.2 The ensemble transform Kalman filter

The Ensemble transform Kalman filter (ETKF) makes use of the identity

$$\mathbf{I} - \mathbf{Y'}^{fT}_n\mathbf{S}_n^{-1}\mathbf{Y'}^f_n = (\mathbf{I} + \mathbf{Y'}^{fT}_n\mathbf{R}_n^{-1}\mathbf{Y'}^f_n)^{-1}, \tag{2.23}$$

which can be verified by multiplying both sides by $\mathbf{I} + \mathbf{Y'}^{fT}_n\mathbf{R}_n^{-1}\mathbf{Y'}^f_n$ [Tippett et al., 2003]. The ETKF can be applied by taking the singular value decomposition (SVD) of $\mathbf{Y'}^{fT}_n\mathbf{R}_n^{-1}\mathbf{Y'}^f_n = \mathbf{U}_n\boldsymbol{\Lambda}_n\mathbf{U}_n^T$ [Livings, 2005], where $\mathbf{U}_n$ is an orthogonal matrix and $\boldsymbol{\Lambda}_n$ is a diagonal matrix both of size $N^m \times N^m$. Substituting into (2.23) gives

$$\mathbf{I} - \mathbf{Y'}^{fT}_n\mathbf{S}_n^{-1}\mathbf{Y'}^f_n = \mathbf{U}_n(\mathbf{I} + \boldsymbol{\Lambda}_n)^{-1}\mathbf{U}_n^T. \tag{2.24}$$

This allows us to write the square root matrix as $\boldsymbol{\Upsilon}_n = \mathbf{U}_n(\mathbf{I} + \boldsymbol{\Lambda}_n)^{-\frac{1}{2}}\mathbf{U}_n^T$. This is the square root form that gives the unbiased ETKF [Livings et al., 2008]. This calculation of $\boldsymbol{\Upsilon}$ is used in the analysis perturbation update. Updated ensemble members are obtained by adding the perturbations onto the ensemble analysis mean as in step 6 in Table 2.3. These ensemble members are forecast and the steps of the scheme repeated for each observation time. We summarise the ETKF algorithm in Table 2.3.

The ETKF Algorithm [Bishop et al., 2001]

1. Forecast the analysis ensemble to obtain a background state at the next time.

$$\mathbf{x}_{n+1}^{f,i} = \mathcal{M}_n \mathbf{x}_n^{a,i}$$

2. Calculate the ensemble forecast mean

$$\bar{\mathbf{x}}_n = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_n^i$$

and covariance

$$\mathbf{P}_n^f = \frac{1}{N-1}\mathbf{X'}_n^f \mathbf{X'}_n^{fT},$$

where $\mathbf{X'}_n^f$ is the ensemble forecast perturbation matrix.

3. Update the ensemble mean

$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^f + \mathbf{K}_n(\mathbf{y}_n - \mathbf{H}_n \bar{\mathbf{x}}_n^f).$$

4. Calculate the square root matrix

$$\mathbf{\Upsilon}_n = \mathbf{U}_n(\mathbf{I} + \mathbf{\Lambda}_n)^{-\frac{1}{2}}\mathbf{U}_n^T.$$

5. Update the ensemble perturbation matrix,

$$\mathbf{X'}_n^a = \mathbf{X'}_n^f \mathbf{\Upsilon}_n.$$

6. Add the ensemble perturbations to the ensemble mean,

$$\mathbf{X}_n^a = \mathbf{X'}_n^a + \bar{\mathbf{X}}_n^a,$$

to obtain the analysis ensemble.

Table 2.3: The ETKF algorithm.

2.3.3 Discussion

The main benefit from using any form of ensemble Kalman filter is that the scheme provides estimates of the pdfs associated with the analysis. When forecast, the ensembles also give information on the uncertainty in the forecast as they provide a Monte Carlo estimate of the evolution of the pdf using the forecast model. The ensembles can also be used to determine the background error covariance $\mathbf{P}_n^f$ at each time, and this helps reduce the costly computations of the error covariance forecast and the error covariance update. However, to obtain a good estimate for this flow dependent background matrix it is necessary to use a sufficient number of ensembles otherwise the estimate of $\mathbf{P}^f$ is contaminated with sampling

error. The number of ensemble members required increases as the size of the state increases. In small dynamical systems it is possible to run the assimilation scheme with enough ensemble members. However, for high dimensional systems it is too computationally costly to run the assimilation scheme with all the required ensemble members. If too few samples are used then it is likely that the ensemble will not be statistically representative of the background error; this is known as undersampling. Undersampling can introduce a number of other problems into the ensemble filtering process. To overcome undersampling and the problems it introduces it is common to use the techniques of covariance inflation [Anderson and Anderson, 1999] and localisation [Hamill et al., 2001]. Using these techniques allows the EnKF to be used in large dimensional systems by reducing the number of ensemble members required. In sections 2.3.4 and 2.3.5 we introduce the ideas of covariance inflation and localisation.

Until recently most operational centres [Rabier et al., 2000, Rawlins et al., 2007, Gauthier et al., 2007] have used 4D-Var. With the development of ensemble data assimilation the potential to use the ensemble Kalman filter in NWP was shown [Lorenc, 2003]. As the techniques have been developed that allow the EnKF to be used in large dimensional systems their use operationally has emerged. Some operational centres [Buehner et al., 2010, Miyoshi et al., 2010] are running both variational and ensemble methods as this allows the ensemble methods to be developed and compared to current operational methods. Hybrid ensemble-variational methods combining the best of the variational and ensemble techniques are also being implemented [Clayton et al., 2012]. The use of ensemble data assimilation would not have been possible, however, without the techniques of covariance inflation and localisation, which are introduced in the following sections.

2.3.4 Covariance inflation

Two of the main problems introduced by undersampling are inbreeding and filter divergence. Inbreeding is a term used describe the underestimate of the analysis error covariance [Furrer and Bengtsson, 2007]. Filter divergence occurs when the analysis error distribution moves away from the truth and this can be caused by underestimating the forecast error

covariance. Both inbreeding and filter divergence can be overcome by inflating the variance of the forecast error covariance matrix. The method was introduced by Anderson and Anderson [1999] and involves multiplying the forecast error covariance by a factor $\gamma > 1$. Multiplying by this factor should correct for the underestimate of the covariance and as $\gamma \mathbf{P}^f$, rather than $\mathbf{P}^f$, is used in the assimilation scheme there is more chance that the analysis can be corrected using the observations as less weight is given to the background. The size of the inflation factor will depend on a number of factors including the type of filter and dynamical system used [Hamill et al., 2001]. However, covariance inflation does not overcome all the problems associated with undersampling. We must also consider covariance localisation.

2.3.5 Covariance localisation

Spurious long range correlations in the forecast error covariance matrix can also be caused by undersampling. These spurious correlations can be removed using covariance localisation [Hamill et al., 2001, Buehner and Charron, 2007]. Correlation localisation makes use of the fact that generally the correlation between two points decreases as the distance between the points increases. It is achieved by taking the Schur product of the correlation matrix and a localizing function matrix. The localizing function matrix is a symmetric positive definite correlation matrix with local support. The Schur product [Schur, 1911], denoted $A \circ B$, is the element wise multiplication ($(A \circ B)_{ij} = A_{ij}B_{ij}$) of two matrices of the same dimensions. Before applying covariance localalisation it is necessary to determine a suitable localisation matrix. The localisation function must be chosen with an appropriate lengthscale such that nearby correlations are preserved. As the separation distance increases the localisation function should decrease until it reaches zero.

The localised covariance matrix is determined by taking the Schur product of the localisation matrix, L, and general covariance matrix, C,

$$C_L = C \circ L. \tag{2.25}$$

The localised covariance matrix will be a symmetric positive definite matrix and the variance will remain unchanged. Due to the zero correlation in the localising function at long range, the localised covariance matrix will contain no spurious long range correlations as they are eliminated by the application of the Schur product.

Although we have introduced covariance inflation and localisation as techniques to overcome problems in ensemble filters they can also be used in other areas of data assimilation. There is also no restriction to applying these techniques only to $\mathbf{P}^f$. It is possible to apply covariance localisation any time a set of samples are used to generate a covariance matrix, this is demonstrated later in Chapter 5.

2.4 Data assimilation diagnostics

We have introduced a number of different assimilation schemes. We now present two diagnostics that an be used to show how well the schemes are performing.

2.4.1 Root mean squared error

To show how well an assimilation scheme is performing we consider the root mean squared error (RMSE); this is directly linked to the 2-norm. The RMSE measures the average magnitude of the error, again it requires knowledge of the truth and analysis state at time t. In general the truth is not known, but it is available when twin experiments are carried out. In this situation the RMSE is a useful diagnostic. When an ensemble DA method is used, the ensemble mean is used to calculate the RMSE.

If the true solution is the vector $\mathbf{x}^t$ of length N^m and the analysis state is the vector $\mathbf{x}^a$ then the RMSE is calculated by summing over the differences of the components in $\mathbf{x}^t$ and $\mathbf{x}^a$. That is

$$RMSE = \sqrt{\frac{\sum_{j=1}^{N_m}(\mathbf{x}_j^a - \mathbf{x}_j^t)^2}{N_m}}. \tag{2.26}$$

The RMSE is a good diagnostic that allows us to determine how an assimilation scheme is

performing. However, for ensemble methods it only gives us information on the ensemble mean. For ensemble schemes the ensemble spread can be understood by considering the rank histogram.

2.4.2 Rank histograms

As well as considering the RMSE we use another diagnostic known as the Rank Histogram [Hamill, 2000]. Rank histograms are useful for diagnosing errors in the mean and spread of the ensemble members and determining the reliability of the ensemble forecast.

Rank histograms are created by considering where the value of the observation at time t_n falls compared to the values of the forecast ensemble members at the same time. For a particular component of the model state, the values of ensemble members at time t_n are sorted into ascending order so that if $\tilde{x}^i$ are the ensemble members we have $\tilde{x}^1 \leq \tilde{x}^2 \leq ... \leq \tilde{x}^N$. Using this ordering a set of bins covering the intervals $(-\infty, \tilde{x}^1](\tilde{x}^1, \tilde{x}^2] \ldots (\tilde{x}^{N-1}, \tilde{x}^N](\tilde{x}^N, \infty)$ are created. The bin, also known as rank, in which the observation at the particular point at time t_n falls is noted. This value is tallied over all times. The resulting tally is plotted and this is known as the rank histogram.

If the ensemble is statistically representative there is an equal probability that the observation, y, will fall into any of the $N + 1$ ranks. This implies that the resulting rank histogram should have bars of equal height (Figure 2.1 (a)). Histograms that are not uniform can give us information on the quality of the ensemble. Many of the common histograms that are produced from ensemble forecasts are discussed in Hamill [2000]. Four of the most common histograms obtained are shown in Figure 2.1. A U-shaped histogram (Figure 2.1 (b)) suggests a possible lack of variability in the particle sample, whereas an excess of variability overpopulates the central ranks (Figure 2.1 (c)). Having the left half of the histogram overpopulated (Figure 2.1 (d)) suggests that particles are positively biased, overpopulation in the right half implies the particles are negatively biased. However, these are not the only possible reasons for these types of behaviour.

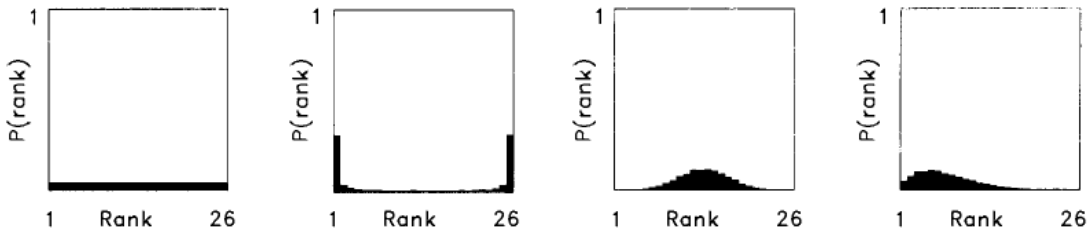


Figure 2.1: Common rank histograms obtained. The rank histograms show the number of times the observation falls in each bin. Plots from Hamill [2000]

2.5 Summary

In this chapter we have introduced the concept of data assimilation; a method that combines models and observations to provide a best guess of the true state. The notation for dynamical systems and data assimilation that will be used throughout this thesis has been introduced. We have then given a brief overview of some different types of sequential and variational data assimilation discussing both their benefits and problems. The ensemble transform Kalman filter has been discussed in greater detail as we use this method in Chapter 7 in the thesis. Ensemble filters suffer from a number of problems and we have discussed two methods, covariance inflation and localisation, that can be used to overcome these. Some diagnostics for data assimilation are also considered. We have seen how to calculate the RMSE and the rank histogram. How to interpret the rank histogram has also been discussed. In this chapter we have seen that there are various errors that must be accounted for in data assimilation. One of these errors is forward model error. We now discuss this important error in greater detail in Chapter 3.

Chapter 3

Forward Model and Representativity Error

In this chapter we give a mathematical definition of forward model error. We consider how forward model and representativity errors are currently treated in data assimilation schemes. We then go on to consider existing methods for calculating the observation error covariance matrix and for estimating representativity error. We describe two of these methods, which we will use in the thesis, in detail.

3.1 Defining representativity error

For a data assimilation scheme to produce an optimal estimate of the state the error covariances must be well understood and correctly specified. Although in practice many assumptions are violated and the analysis provided by the assimilation is far from optimal, it is still desirable to have good estimates of the covariance matrices. Therefore it is important that we understand the sources of error that are represented in the observation error covariance matrix.

One form of error, $\boldsymbol{\epsilon}_n^I$, contains information on the instrument error, one describes the error in the observation operator, and a third contains information on the representativity

error, also known as representativeness or representivity error. We follow Daley [1993, 1991] and define the representativity error as the error that arises when the observations resolve spatial scales that the model cannot. The representativity error and the error in the observation operator can be combined as a single error known as the forward model or forward interpolation error [Lorenc, 1986] $\boldsymbol{\epsilon}_n^H$. In more recent literature [Cohn, 1997, Liu and Rabier, 2002, Janjic and Cohn, 2006], the term representativity error has been used to describe the forward model error. However, we use the definition given by Daley [1991] as in this work we focus on the error caused by the misrepresentation of small scales. The instrument error is determined for specific instruments under a set of test conditions by the instrument manufacturer or from in-orbit calibration data for satellite measurements. Less is known about the forward model error; however, it has been suggested that the forward model error is larger than the instrument error for observations of fields that are highly variable such as moisture and wind [Mènard et al., 2000]. As it is important to have good estimates of the covariance matrices, it is important that we understand forward model error.

Forward model error,

$$\boldsymbol{\epsilon}^H = \mathbf{y}^t - \mathcal{H}(\mathbf{x}^t), \tag{3.1}$$

is the difference between the noise free observation vector, $\mathbf{y}^t$, of length N^p and the mapping of the true state, $\mathbf{x}^t$, into observation space using the possibly non-linear observation operator $\mathcal{H}$. The noise free observation vector is a theoretical construct that represents an observation measured by a perfect observing instrument, i.e. with no instrument error. It is related to the actual measurement via the equation

$$\mathbf{y} = \mathbf{y}^t + \boldsymbol{\epsilon}^I, \tag{3.2}$$

where $\mathbf{y}$ is the observation vector and $\boldsymbol{\epsilon}^I$ is the instrument error. The forward model error contains contributions from two error sources. Errors can be introduced due to the mis-specification of $\mathcal{H}$; these include modeling errors such as the misrepresentation of gaseous constituents in radiative transfer model, parameter errors caused by preprocessing of the

data such as cloud clearance for radiances, and errors due to the approximation of a continuous function as a discrete function. The second contributor to forward model errors are the errors of representativity, the errors introduced when the observations resolve spatial scales that the model cannot. These errors can be accounted for in a data assimilation scheme by considering the covariance of the forward model error $E[\boldsymbol{\epsilon}^H \boldsymbol{\epsilon}^{H^T}] = \mathbf{R}^H$. We see that it makes sense to include the statistics of forward model error in the observation error covariance matrix $\mathbf{R} = \mathbf{R}^H + \mathbf{R}^I$, where $\mathbf{R}^I = E[\boldsymbol{\epsilon}^I \boldsymbol{\epsilon}^{I^T}]$ is the instrument error covariance matrix. Not all methods that take account of representativity error assume that this is where the forward model statistics should be included and we will briefly discuss these later in this chapter in section 3.3. However, we do assume that forward model errors are included in $\mathbf{R}$ in the work carried out in this thesis. Now we have an understanding of what representativity errors are, we consider how they are currently treated in assimilation schemes.

3.2 Current treatment of forward model error

Currently forward model error is rarely treated explicitly within data assimilation systems. Comparatively little is known about the structure of forward model error, however, work from Stewart [2010] and Weston [2011] has shown that the forward model errors may be correlated. Until recently it has been assumed that it is too expensive to include correlated observation error matrices in assimilation schemes. Due to the cost of using correlated errors and lack of understanding about forward model errors, a diagonal observation error covariance matrix is often used. Currently with a diagonal matrix $\mathbf{R}$ and a correlated matrix $\mathbf{B}$, all the scale dependent filtering and spreading of the observation information is accomplished by the matrix $\mathbf{B}$ [Seaman, 1977]. However, when using a correlated matrix $\mathbf{R}$ the correlations between the model space and observations are still determined by $\mathbf{BH}^T$, but the innovation vector will be rescaled and rotated and scaled differently, resulting in a different analysis. While it has been too costly to use correlated $\mathbf{R}$ matrices it has been necessary to account for the unknown forward model in different ways. Variance inflation [Hilton et al., 2009, Whitaker et al., 2008] is used to inflate the variance of the instrument

error in an attempt to account for some of the unknown forward model error and absent correlation structure. The effect of correlated error is also reduced by using techniques such as observation thinning [Lahoz et al., 2010] or superobbing [Daley, 1991]. Observation thinning is a method where the number of observations in a given area is reduced. In its most basic form observations are thinned by discarding observations until the remaining observations are at a desired density. However, thinning can be more sophisticated so more observational data is used, one such method is superobbing. Superobbing thins the data by averaging a group of observations in a given area and using the average of these observations as a single observation. The idea behind reducing the density of observations is that observations further apart are less likely to have correlated errors, therefore with fewer observations the assumption of a diagonal observation error covariance is more valid [Liu and Rabier, 2002, 2003, Dando et al., 2007]. These methods help reduce the correlations in the observation errors; however, they also reduce the amount of useful information that can be extracted from the observations. To make the most of the observational data available the correlated errors must be accounted for in data assimilation. Efforts are being made to find methods of reducing the cost of using correlated observation error matrices [Stewart et al., 2009, Stewart, 2010, Healy and White, 2005]. The full observation error correlation matrix may also be poorly conditioned and this may affect the minimisation in the assimilation, so the preconditioning of these correlated error matrices is also being considered [Weston, 2011]. Once these methods are in place it will be important to have accurate estimates of the covariance matrices, as these are required to obtain the optimal estimate from any data assimilation system [Houtekamer and Mitchell, 2005]. It is therefore important to understand how to estimate forward model error. Once forward model error can be estimated, it will need to be included in the data assimilation scheme. We will now consider how forward model error could be accounted for in a data assimilation scheme.

3.3 Accounting for representativity error

We showed in section 3.1 that one way to include forward model error in a data assimilation scheme is in the observation error covariance matrix $\mathbf{R}$, but there are also other suggestions on how forward model error could be accounted for. We shall now discuss some of the methods that exist for calculating and accounting for forward model error. Most of the methods that exist produce time independent estimates of forward model error; however, methods that produce time dependent estimates do exist. Some of the methods we describe do not directly calculate the forward model error, but give an estimate of the observation error covariance matrix $\mathbf{R}$. As $\mathbf{R} = \mathbf{R}^H + \mathbf{R}^I$, if the instrument error is known, then an estimate of the forward model error covariance can be calculated. The instrument error is often supplied by the instrument manufacturer, or in the case of satellite instruments may be calibrated by the instrument itself.

Many of the methods that exist for estimating forward model error give a time independent estimate. The forward model error covariance, $\mathbf{R}^H$, or the observation error covariance matrix $\mathbf{R}$ can be calculated outside of the data assimilation scheme. This calculated observation error covariance matrix $\mathbf{R} = \mathbf{R}^H + \mathbf{R}^I$ is then used as a fixed matrix in the assimilation scheme. The statistical samples used in the calculation are taken over a large period of time. Therefore there is an implicit assumption that the errors are not changing over time. This assumption is invalid as in theory it has been shown that the representativity error is time and state dependent [Janjic and Cohn, 2006]. We show later in Chapter 6 that representativity error is dependent on synoptic situation. It will be larger when the field contains more smaller scale features, and smaller when the field is dominated by large scale features. It is likely that samples taken over a period of time will contain samples where the representativity error is large and some where the representativity error is small. Using these samples gives a time averaged forward model error which when used in an assimilation scheme will overestimate the representativity error when there are many large scale features, therefore not trusting the observations enough. When the field contains small scale features the time independent forward model error will be an underestimate, therefore trusting the observations too much. The assumption that representativity error

is time independent is invalid; however, these methods have the benefit that they are calculated outside the data assimilation scheme. This means that no extra computational cost is added to the assimilation, and as much computation as is required can be used to calculate the forward model error. We now discuss some of the methods that exist for estimating time independent forward model error.

We now consider methods that have been used to calculate the observation error covariance matrix $\mathbf{R}$. One of the first, and most commonly used methods used to calculate $\mathbf{R}$ is the Hollingsworth-Lönnberg method [Hollingsworth and Lönnberg, 1986]. This method was originally introduced for estimating the background error assuming a known observation error but has since been used for the estimation of the observation error covariance matrix. The method makes use of the expected value of the innovation statistics and the assumption that background errors are correlated while observation errors are not. Given that the background errors (equation (2.3)) and observation errors (equation (2.2)) are uncorrelated the expected value of the background innovations, $\mathbf{d}^b = \mathbf{y} - \mathcal{H}(\mathbf{x}^b)$, is

$$\begin{aligned} E[\mathbf{d}^b \mathbf{d}^{bT}] &= E[(\mathbf{y} - \mathcal{H}(\mathbf{x}^b))(\mathbf{y} - \mathcal{H}(\mathbf{x}^b))^T], \\ &\approx \mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}. \end{aligned} \tag{3.3}$$

The result of this can be plotted and from this the error can be split into the correlated background error covariance matrix $\mathbf{B}$ and the uncorrelated observation error covariance matrix $\mathbf{R}$ [Bormann and Bauer, 2010]. This was an appropriate method to use when uncorrelated observation error covariances were required. But as efforts are being made to account for correlated errors a method that produces correlated estimates is required. Such a method was proposed by Desroziers et al. [2005]. The Desroziers diagnostic, like the Hollingsworth-Lönnberg method, makes use of the innovation statistics. However, rather than just using the background innovation, the Desroziers diagnostic also uses the innovation of the analysis states $\mathbf{d}^a = \mathbf{y} - \mathcal{H}(\mathbf{x}^a)$. Under the assumption that the covariance matrices $\mathbf{R}$ and $\mathbf{B}$ used to calculate the analysis are correct, taking the the expectation of the cross product of the analysis and background innovations gives an expression for calculating the observation error covariance matrix, $E[\mathbf{d}^a \mathbf{d}^{bT}] = \mathbf{R}$. The derivation of

this is given later in Section 3.3.2. It is suggested that the diagnostic can be used as a consistency check after the analysis, but more recently it has been used to estimate a matrix $\mathbf{R}$ that is then fed back into a scheme to iteratively improve the analysis. We use this method in the thesis, so further details are given in section 3.3.2.

As well as estimating the full observation error covariance matrix, $\mathbf{R}$, there are also methods that have been developed that allow the forward model error covariance $\mathbf{R}^H$ to be calculated individually. One of these methods was first defined by Daley [1993] and then used by Liu and Rabier [2002]. In this method it is assumed that the observations can be written as the mapping of a high resolution state into observation space, and that the model state is a spectral truncation of this high resolution state. This method is used within the thesis and is explained in more detail in section 3.3.1.

Other methods have attempted to calculate just the representativity error using a set of observations. The work by Oke and Sakov [2007] takes a set of high resolution observations, and then averages these observations to the resolution of the model. This averaged data is then interpolated back to the high resolution grid and the difference between the true observations and this averaged data is taken to be the representativity error. Although providing promising estimates the method suffers from a number of limitations that make it impractical as a general method to calculate representativity error. The method requires high resolution observations that resolve all scales of variability; if not all these are resolved then the representativity error will be underestimated. The observations must also be interpolated to the model grid which may lead to an additional source of error.

As well as methods that give time independent estimates, there have also been attempts to estimate time dependent representativity error. This is important as forward model error is dependent on the true state. Time dependent estimates must be calculated within the assimilation scheme, which can cause problems due to the computational cost. Janjic and Cohn [2006] suggested a different approach where the forward model error is not calculated explicitly, but is accounted for in an augmented Kalman filter. The data assimilation scheme they developed accounts for the representativity error and allows the unrepresented scales to influence the scales resolved by the model. However, this method

is computationally costly and further approximations are needed due to the presence of unknown correlations between the resolved and unresolved scales.

More recent work by Li et al. [2009] has used the Desroziers diagnostic embedded in a local ensemble transform Kalman filter to give an estimate of the observation error variances. At each analysis step the Desroziers diagnostic is applied to a subset of observations to give a value for the observation error variance. This is iterated until the variance converges to what is assumed to be the correct static observation error matrix. This method has returned to the assumption of a diagonal matrix $\mathbf{R}$, and also assumes that each observation had the same associated error variance and that the true observation error variance is static. This work was extended in Miyoshi et al. [2013] to include a correlated $\mathbf{R}$ matrix. In this framework it is possible to average over a subset of observations as all observations have the same variance. However, as forward model error is time and state dependent averaging over observations may also be a poor assumption. In Chapter 7 we develop a method with similar ideas as in Li et al. [2009], but remove the assumptions of the diagonal $\mathbf{R}$, as in Miyoshi et al. [2013], and the need to average over a subset of observations.

In this thesis we consider two methods for calculating representativity error. One is a method developed by Daley [1993] and then used by Liu and Rabier [2002]. This method can be used for performing idealised experiments to calculate representativity error on a periodic 1D domain. We also develop our own method for calculating representativity error that makes use of the Desroziers et al. [2005] diagnostic. We now present in more detail the Daley [1993] method and the Desroziers et al. [2005] diagnostic.

3.3.1 The Daley [1993] method

We first present the method defined by Daley [1993] and Liu and Rabier [2002]. In this method it is assumed that the observations can be written as the mapping of a high resolution state into observation space, and that the model state $\mathbf{x}$ is a truncation of this high resolution state. This method also allows us to correctly specify the observation operator so our forward model errors consist only of errors of representativity. The assumptions made

by this method make it more suitable for calculating representativity error in idealised experiments on a 1D periodic domain.

We restrict our calculations to the 1D domain of length $l = 2a\pi$, where a is a constant that determines the length of the domain, and assume that the observation operator $\mathbf{H}$ is linear. It is assumed that the truth can be approximated by a high resolution state. The high resolution state $\mathbf{x}^t(r)$ at position r can be expressed as a Fourier series truncated at wave number K^t. At N^t points on the physical domain, $-a\pi \leq r \leq a\pi$, the function values $\mathbf{x}^t(r_j)$, $j = 1 \ldots N^t$, can be expressed in matrix form as

$$\mathbf{x}^t = \mathbf{F}^t \hat{\mathbf{x}}^t \tag{3.4}$$

where $\hat{\mathbf{x}}^\mathbf{t}$ is a vector of length $M^t = 2K^t + 1$ of spectral coefficients and $\mathbf{F}^t$ is a Fourier transform matrix of dimension $N^t \times M^t$. In this work a number of Fourier matrices are used to calculate forward model error. A Fourier matrix $\mathbf{F}$ of size $m \times n$ has elements

$$\mathbf{F}_{j,k} = \exp(\tfrac{2ikj\pi}{m}), \tag{3.5}$$

where $j = 1 \ldots m$ and $k = 1 \ldots n$.

The model representation of the actual state is a wave number limited filter of the high resolution state, $\hat{\mathbf{x}} = \mathbf{T}\hat{\mathbf{x}}^t$ where $\mathbf{T}$ is a truncation matrix that truncates the full spectral vector $\hat{\mathbf{x}}^t$ to the spectral vector $\hat{\mathbf{x}}$. The model representation of the actual state can be expressed as

$$\mathbf{x} = \mathbf{F}^m \hat{\mathbf{x}}, \tag{3.6}$$

where $\hat{\mathbf{x}}$ is a vector of length $M^m = 2K^m + 1$ of spectral coefficients and $\mathbf{F}^m$ is a Fourier transform matrix of dimension $N^m \times M^m$ with elements defined as in equation (3.5) but with no terms with wave number higher than K^m. The spectral coefficients for the model representation of the actual state are the Fourier spectral coefficients from $-K^m$ to K^m, $K^m < K^t$.

We define the observations by

$$y(r_o) = \int_{-a\pi}^{a\pi} x(r)w(r - r_o)dr. \tag{3.7}$$

Here the observations are defined as if they have been measured at point r_o by a remote sensing instrument. The choice of the weighting function $w(r)$ determines the type of observing instrument. Writing equation (3.7) in spectral space allows us to write the N^p equally spaced error free observations as

$$\mathbf{y}^t = \mathbf{F}^p \mathbf{W}^t \hat{\mathbf{x}}^t, \tag{3.8}$$

where $\mathbf{F}^p$ is a $N^p \times M^t$ Fourier transform matrix and $\mathbf{W}^t$ is a $M^t \times M^t$ diagonal matrix with elements $\hat{w}_k$, the spectral coefficients of the weighting function $w(r)$. $\mathbf{F}^p\mathbf{W}^t$ is an exact observation operator in spectral space. The measurement vector $\mathbf{y}$ is given by,

$$\mathbf{y} = \mathbf{F}^p \mathbf{W}^t \hat{\mathbf{x}}^t + \boldsymbol{\epsilon}^I, \tag{3.9}$$

where $\boldsymbol{\epsilon}^I$ is the instrument error.

The model representation of the observations is given by,

$$\mathbf{y}^m = \mathbf{F}^p_m \mathbf{W}^m \mathbf{T} \hat{\mathbf{x}}^t, \tag{3.10}$$

where $\mathbf{F}^p_m$ is the $N^p \times N^m$ Fourier matrix with elements defined as in equation (3.5). $\mathbf{W}^m$ is a $M^m \times M^m$ diagonal matrix with elements $\hat{w}_k$, the spectral coefficients of the weighting function $w(r)$. This method assumes that the low resolution model is a truncation of the high resolution model. This allows forward model error to be considered in the perfect model case.

To obtain an equation for forward model error we substitute the definitions of observations, equation (3.8), and model representation of the observation, equation (3.10), into equation (3.1) to give,

$$\boldsymbol{\epsilon}^H = \mathbf{F}^p \mathbf{W}^t \hat{\mathbf{x}}^t - \mathbf{F}^p_m \mathbf{W}^m \mathbf{T} \hat{\mathbf{x}}^t. \tag{3.11}$$

The expectation operation is applied to give the forward model error covariance matrix $E[\boldsymbol{\epsilon}^H \boldsymbol{\epsilon}^{\boldsymbol{H}*}] = \mathbf{R}^H$,

$$\begin{aligned} \mathbf{R}^H &= E[(\mathbf{F}^p\mathbf{W}^t\hat{\mathbf{x}}^t - \mathbf{F}^p_m\mathbf{W}^m\mathbf{T}\hat{\mathbf{x}}^t)(\mathbf{F}^p\mathbf{W}^t\hat{\mathbf{x}}^t - \mathbf{F}^p_m\mathbf{W}^m\mathbf{T}\hat{\mathbf{x}}^t)^*] \\ &= (\mathbf{F}^p\mathbf{W}^t - \mathbf{F}^p_m\mathbf{W}^m\mathbf{T})\hat{\mathbf{S}}(\mathbf{F}^p\mathbf{W}^t - \mathbf{F}^p_m\mathbf{W}^m\mathbf{T})^*, \end{aligned} \tag{3.12}$$

where $\hat{\mathbf{S}} = E[\hat{\mathbf{x}}^{\mathbf{t}}\hat{\mathbf{x}}^{\mathbf{t}*}]$ is the spectral covariance matrix for the high resolution state and $*$ denotes the complex conjugate transpose. The spectral covariance of the high resolution state, $\hat{\mathbf{S}}$, contains information on how different wave numbers are related. It can be calculated using

$$\hat{\mathbf{S}} = \mathbf{F}^{t*}\mathbf{S}\mathbf{F}^t, \tag{3.13}$$

where $\mathbf{F}^t$ is a Fourier transform matrix and $\mathbf{S} = E[\mathbf{x}^t\mathbf{x}^{t*}]$ is the covariance matrix of the high resolution state in physical space.

We now have an equation that can be used to calculate the forward model error covariance matrix. To use the method it is necessary to know the weighting matrices and the spectral covariance matrix. The spectral covariance matrix depends on the true state and the weighting matrix on the pseudo-observations. These differ for specific experiments so will be defined later when required.

This method is useful as it allows the forward model error to be calculated explicitly. It also has the advantage that the error in the observation operator can be removed, allowing the representativity error to be understood. However, the method assumes that the observation operator is linear and gives only a time independent estimate of forward model error. The other drawback to the method is that it makes the assumption that the low resolution model data is a truncation of the high resolution model data and that the data is given on a periodic domain. This assumption is not necessarily valid as the solution to a low resolution model may evolve differently from the solution of the high resolution model and the domain may not be periodic.

In Chapter 7 we develop a a more general method that can be used to calculate representativity error and forward model error. The method makes use of the Desroziers et al.

[2005] diagnostic which we now explain in greater detail.

3.3.2 The Desroziers diagnostic

The Desroziers et al. [2005] diagnostic, like the Hollingsworth-Lönnberg method, makes use of the innovation statistics. However, rather than just using the background innovation $\mathbf{d}^b = \mathbf{y} - \mathcal{H}(\mathbf{x}^b)$, the Desroziers diagnostic also uses the innovation of the analysis states $\mathbf{d}^a = \mathbf{y} - \mathcal{H}(\mathbf{x}^a)$. By making the tangent linear hypothesis on the observation operator and using equations (2.2) and (2.3) the background innovation can be written in terms of the background and observation errors,

$$\begin{aligned} \mathbf{d}^b = \mathbf{y} - \mathcal{H}(\mathbf{x}^b) &= \mathbf{y} - \mathcal{H}(\mathbf{x}^t) + \mathcal{H}(\mathbf{x}^t) - \mathcal{H}(\mathbf{x}^b), \\ &\approx \boldsymbol{\epsilon}^o + \mathbf{H}(\mathbf{x}^t - \mathbf{x}^b), \\ &\approx \boldsymbol{\epsilon}^o + \mathbf{H}\boldsymbol{\epsilon}^b, \end{aligned} \tag{3.14}$$

where $\mathbf{H}$ is the linearised version of $\mathcal{H}$. By using the analysis equation, equation (2.9), from the BLUE and again assuming the tangent linear hypothesis on the observation operator allows the innovations of the analysis to be written as,

$$\begin{aligned} \mathbf{d}^a &= \mathbf{y} - \mathcal{H}\mathbf{x}^a, \\ &= \mathbf{y} - \mathcal{H}(\mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathcal{H}\mathbf{x}^b)), \\ &= \mathbf{y} - \mathcal{H}(\mathbf{x}^b + \mathbf{K}\mathbf{d}^b), \\ &\approx \mathbf{d}^b - \mathbf{H}\mathbf{K}\mathbf{d}^b, \\ &= (\mathbf{I} - \mathbf{H}\mathbf{K})\mathbf{d}^b, \\ &= \mathbf{R}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathbf{T}} + \mathbf{R})^{-1}\mathbf{d}^b. \end{aligned} \tag{3.15}$$

Taking the cross product of the analysis and background innovations and assuming that the background and observation errors are uncorrelated results in

$$\mathbf{d}^a\mathbf{d}^{bT} = \mathbf{R}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d}^b\mathbf{d}^{bT}. \tag{3.16}$$

Then taking the expected value of this cross product gives

$$\begin{aligned} E[\mathbf{d}^a \mathbf{d}^{bT}] &= \mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1} E[\mathbf{d}^b \mathbf{d}^{bT}], \\ &= \mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1}(\mathbf{HBH}^T + \mathbf{R}), \\ &= \mathbf{R}. \end{aligned} \tag{3.17}$$

This is valid if the observation and background errors used in the gain matrix, $\mathbf{K}$, used to calculate the analysis, are the exact observation and background errors. However, Desroziers et al. [2005] has shown that a reasonable estimate of $\mathbf{R}$ can be obtained even if the $\mathbf{R}$ and $\mathbf{B}$ used in $\mathbf{K}$ are not correctly specified. It has also been shown that the method can be used as an iterative method to estimate $\mathbf{R}$ [Mènard et al., 2009, Desroziers et al., 2009]. The Desroziers diagnostic only provides information on the complete observation error covariance matrix. Therefore it is necessary to subtract the known instrument error to calculate the forward model error.

3.4 Summary

In this chapter we have introduced the ideas of forward model error and representativity error and given a mathematical description of forward model error. We have seen that currently forward model error is rarely treated explicitly in data assimilation schemes. Instead the techniques of covariance inflation and superobbing are used so the observation error covariance matrix can be assumed diagonal. As data assimilation methods evolve it will be important to have accurate estimates of the covariance matrices. We have considered a number of methods that have been developed to estimate the observation error covariance matrix $\mathbf{R}$, as well as those that give both time independent and dependent estimates of forward model errors. We have described in further detail the methods used in this thesis: A method proposed by Daley [1993] and then Liu and Rabier [2002], and the Desroziers diagnostic. We next use the Daley [1993] method to help us understand forward model error.

Chapter 4

Using the Daley [1993] Method

In the previous chapter we described a method introduced by Daley [1993] for calculating forward model and representativity error in idealised experiments on a 1d periodic domain. The method uses observations defined using weighting matrices, and requires the spectral covariance for the true state of the system. In this chapter we describe how these weighting matrices can be calculated for three specific cases. We also consider how in the general case the spectral covariance matrix can be calculated. We also consider the scheme in more detail to see if it is possible to obtain any theoretical results that explain the structure of representativity errors.

4.1 Defining the weighting matrices

To calculate the representativity error using the Daley [1993] method we require high resolution observations. We expect representativity error to depend on observation type. Different observation types are defined in equation (3.7), with the choice of weighting function determining the observation type. The weighting function can be represented using a weighting matrix. We choose the weighting matrices in (3.12) to correspond to different types of observing instruments. The elements of the weighting matrix are the spectral coefficients of the weighting function $w(r)$ which is used to define observations using equation (3.7). Pseudo-observations are created from high resolution data of the

true state using three weighting functions. The weighting functions used here are the same as those used in Liu and Rabier [2002]. Two of the weighting functions represent remotely-sensed observations. One follows a uniform curve where the weighting function is given by

$$w(r) = \begin{cases} 0 & \text{if } |r| > \frac{L_0}{2} \\ \frac{1}{L_0} & \text{if } |r| \leq \frac{L_0}{2} \end{cases} \tag{4.1}$$

and its spectral form is given by,

$$\hat{w}_k = \begin{cases} 1 & \text{if } k = 0 \\ \frac{\sin(\frac{kL_0}{2a})}{\frac{kL_0}{2a}} & \text{if } k \neq 0 \end{cases}. \tag{4.2}$$

The other weighting function is calculated using a Gaussian curve where the weighting function and its corresponding spectral form are given by,

$$w(r) = \frac{\exp(\frac{-4r^2}{L_o^2})}{\int_{-a\pi}^{a\pi} \exp(\frac{-4r^2}{L_o^2}) dr}, \tag{4.3}$$

$$\hat{w}_k = \exp(\frac{-k^2 L_o^2}{16a^2}). \tag{4.4}$$

The lengthscale parameter, L_o, in both the uniform and Gaussian weighting functions must be chosen to give the desired lengthscale for the observation.

We plot in Figure 4.1 both the uniform and Gaussian weighting functions at different lengthscales.

The different weighting functions give different ways to average the grid point data. The lengthscale determines how many grid points are averaged over.

We also consider in-situ measurements. For these direct observations the $w(r)$ term in equation (3.7) becomes a Dirac delta-function. In this case the diagonal elements $\hat{w}$ of the weighting matrix are all unity.

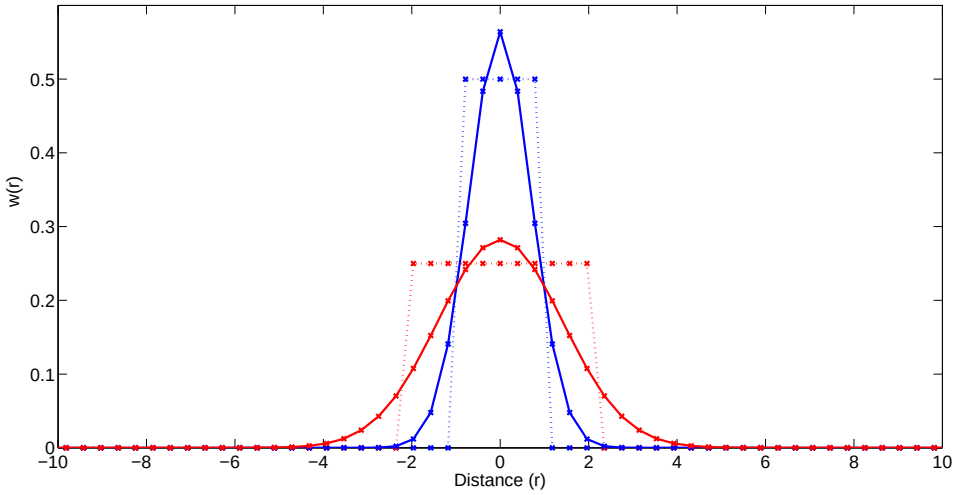


Figure 4.1: Weighting functions used to define pseudo-observations. Dotted lines, Uniform (top hat) function. Solid line, Gaussian function. Red: lengthscale $L_o = 4.0$ Blue: lengthscale $L_o = 2.0$

Now we have defined the weighting matrices we are left to determine the spectral covariance matrix $\hat{S}$.

4.2 Calculating the true correlation matrix

The spectral covariance of the true state, $\hat{S}$, contains information on how the state at one location is related to the state at another location.

From equation (3.13) we see that it is possible to calculate the spectral covariance using the covariance of the truth. This matrix S can be calculated from a number of samples, $\mathbf{x}_i$, of the truth using.

$$\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \tag{4.5}$$

where $\mathbf{x}_i$ is the i^{th} sample vector and $\bar{\mathbf{x}}$ is a vector of the mean of the samples.

It is also useful to consider $\mathbf{S} = \mathbf{DCD}$ where $\mathbf{D}$ is a diagonal matrix of standard deviations d_{ii} and $\mathbf{C}$ is a correlation matrix.

We can calculate the standard deviations using,

$$d_{ii} = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{ji} - \bar{x_j})^2} \tag{4.6}$$

The correlation matrix can be calculated from the covariance matrix using,

$$C_{ij} = \frac{S(i,j)}{\sqrt{S(i,i)S(j,j)}}. \tag{4.7}$$

As these matrices are calculated using a number of samples it is likely that they will contain sampling error. The Daley [1993] method requires the matrix to be isotropic and homogeneous, as this leads to a diagonal $\hat{\mathbf{S}}$ matrix, so it is likely that the sampling error will have to be compensated for. One possibility is to use covariance localisation, as described in Section 2.3.5.

Now we have shown how the weighting and spectral covariance matrices can be calculated we next show some theoretical results relating to the Daley [1993] method.

4.3 Theoretical results

We now show new results relating to the correlation structure and variance of representativity error. We show that the representativity error variance is independent of the number of available observations. We also show that the correlation structure of representativity error is dependent not on the number of observations, but the distance between them. We do this by considering the calculation of the elements of the representativity error covariance matrix.

Representativity error is calculated using equation (3.12). We summarise the elements of the matrices below,

- $\mathbf{F}^p$ with $j = 1 \dots p$ and $k = 1 \dots M$. Elements defined as in equation (3.5).
- $\mathbf{F}^p_m$ with $j = 1 \dots p$ and $k = 1 \dots M_m$. Elements defined as in equation (3.5).

- $\mathbf{W}^t$ with $j = 1 \dots M$ and $k = 1 \dots M$. Nonzero elements defined using a weighting function such as in equations (4.2) and (4.4) when $j = k$, 0 otherwise.
- $\mathbf{W}^m$ with $j = 1 \dots M_m$ and $k = 1 \dots M_m$. Elements $\hat{w}_j$ when $j = k$, 0 otherwise.
- $\hat{\mathbf{S}}$ with $j = 1 \dots M$ and $k = 1 \dots M$. Elements $\hat{s}_{jj}$ when $j = k$, 0 otherwise.
- $\mathbf{T}$ with $j = 1 \dots M_m$ and $k = 1 \dots M$. Elements 1 when $j = k$ and $k \leq M$, 0 otherwise.

For convenience we define $\mathbf{A} = \mathbf{F}^p\mathbf{W}^t$, $\mathbf{B} = \mathbf{F}^p_m\mathbf{W}^m\mathbf{T}$ and $\mathbf{C} = \mathbf{A} - \mathbf{B}$. We begin by calculating the elements of $\mathbf{A}$ and $\mathbf{B}$. As many of the elements are zero we find that $A_{j,k} = F^p_{j,k}W^t_{j,j}$ and $B_{j,k} = F^p_{mj,k}W^m_{j,j}$ when $k \leq M_m$, 0 otherwise.

Next we calculate elements of $\mathbf{C} = \mathbf{A} - \mathbf{B}$.

$$C_{j,k} = \begin{cases} \exp(\frac{2ikj\pi}{p})\hat{w}_j & M_m < k \leq M \\ 0 & \text{Otherwise} \end{cases} \tag{4.8}$$

Now we calculate $\mathbf{R}^H = \mathbf{C}\hat{\mathbf{S}}\mathbf{C}^*$. Elements of $\mathbf{E} = \mathbf{C}\hat{\mathbf{S}}$ are $E_{j,k} = C_{j,k}\hat{S}_{k,k}$. Finally we calculate $\mathbf{R}^H = \mathbf{E}\mathbf{C}^*$, where $*$ is the complex conjugate transpose and $\bar{\cdot}$ is the complex conjugate,

$$\begin{aligned} R^H_{j,k} &= \sum_{l=0}^{M} E_{j,l}C^*_{l,k}, \\ &= \sum_{l=1}^{M} C_{j,l}\hat{S}_{l,l}\bar{C}_{k,l}, \\ &= \sum_{l=1}^{M} \exp(\tfrac{2ijl\pi}{p})\hat{w}_l\hat{s}_l\hat{w}_l \exp(\tfrac{-2ikl\pi}{p}). \end{aligned} \tag{4.9}$$

We now show some theoretical results using (4.9).

We first show that the variance does not change when p changes, which is the case when $j = k$.

Theorem 4.3.1. *The variance of representativity error is independent of the number of observations*

Proof. We consider the variance of representativity error, that is the elements when $j = k$

$$\begin{aligned} \mathbf{R}^H_{j,j} &= \sum_{l=1}^{M} \exp(\tfrac{2ijl\pi}{p})\hat{w}_l\hat{s}_l\hat{w}_l \exp(\tfrac{-2ijl\pi}{p}), \\ &= \sum_{l=1}^{M} \hat{w}_l\hat{s}_l\hat{w}_l. \end{aligned} \tag{4.10}$$

This does not depend on p and hence we do not expect the variance to change when we use different numbers of observations to calculate representativity error. □

We now show that the correlation structure depends only on the distance between observations and not the number of observations.

Our model has N_m grid points separated by a spacing Δx and we have p observations. The distance between consecutive observations is $\Delta p = \frac{(N_m \Delta x)}{p}$.

Theorem 4.3.2. *The correlation structure of representativity error depends not on the number of observations but the distance between them.*

Proof. Suppose we have two observations separated by a distance d and assume that these are observation j and observation k. Then we have

$$\begin{aligned} d &= (j-k)\Delta p \\ &= \frac{(j-k)(N_m\Delta x)}{p}, \end{aligned} \tag{4.11}$$

and hence

$$\frac{(j-k)}{p} = \frac{d}{N_m \Delta x} \tag{4.12}$$

Substituting this into equation (4.9) we obtain

$$
\begin{aligned}
R^H_{j,k} &= \sum_{l=1}^{M} \hat{w}_l \hat{s}_l \hat{w}_l \exp(\tfrac{2il\pi(j-k)}{p}), \\
&= \sum_{l=1}^{M} \hat{w}_l \hat{s}_l \hat{w}_l \exp(\tfrac{2il\pi d}{N_m \Delta x}). \qquad (4.13)
\end{aligned}
$$

Hence the correlations depend only on the distance between the observations and not the number of observations. □

4.4 Summary

In this chapter we have described how we calculate the weighting matrices and the spectral covariance matrix that are required to calculate representativity error using the method defined by Daley [1993]. We have defined the weighting matrices, identity, uniform and Gaussian, that correspond to three different types of observations. We have also discussed how the covariance of the truth can be calculated statistically. We then presented some new theoretical results related to the Daley [1993] method. We showed that the variance of representativity error does not change when calculated with different numbers of observations. We also showed that the correlation structure of the representativity error depends only on the distance between observations and not the number of observations available. We now apply this method to calculate representativity and forward model error for the Kuramoto-Sivashinsky equation.

Chapter 5

Representativity Error for the Kuramoto-Sivashinsky Equation

In this chapter we use the Daley [1993] method to calculate forward model error. We hope to gain an understanding of the structure of forward model error and representativity error and how it changes under different circumstances. To help us understand representativity error we consider the Kuramoto-Sivashinsky (KS) equation, a non-linear partial differential equation (PDE). We start by introducing the KS equation and describing the numerical scheme we use to solve it. We then calculate the forward model error for the KS equation. We consider experiments where the error is a combination of errors in the observation operator and unresolved scales. However, as the Daley [1993] method allows us to specify a correct observation operator we are also able to consider the case where the forward model error consists only of representativity error. We analyse these results and use the theoretical results given in section 4.3 to help us understand the forward model errors and representativity errors.

5.1 The Kuramoto-Sivashinsky equation

The Kuramoto-Sivashinsky (KS) equation,

$$u_t = -uu_x - u_{xx} - u_{xxxx}, \tag{5.1}$$

a non-linear, non-dimensional PDE where t is time and x is a space variable, was proposed by Kuramoto [1978] and independently by Sivashinsky [1977] to model the non-linear evolution of instabilities in flame fronts. The equation produces complex behaviour due to the presence of the second and fourth order terms. The second order term has a destabilising effect as it injects energy into the system whereas the fourth order hyperviscous damping term has a stabilising effect. The non-linear term transfers energy from the low to high wavenumbers. The equation can be solved on both bounded and periodic domains and when this domain is sufficiently large the solutions exhibit multi-scale and chaotic behaviour [Gustafsson and Protas, 2010, Eguíluz et al., 1999]. This chaotic and multi-scale behaviour makes the KS equation a suitable low dimensional model that represents a complex fluid dynamic system. The KS equation has been used previously for the study of state estimation problems using both ensemble and variational methods [Protas, 2008, Jardak et al., 2000] and the multi-scale behaviour makes it particularly suitable model for the study of representativity error. There is no explicit solution to the KS equation, therefore it must be solved numerically. We now consider the numerical solution of the KS equation.

5.1.1 Numerical solution of the KS equation

Kassam and Trefethen [2005] have previously used the KS model to demonstrate how existing numerical methods can be modified to solve stiff non-linear PDEs. We now describe the method proposed in Kassam and Trefethen [2005] as we use it to solve the KS equation. We consider the solution on a periodic domain as it allows us to simplify the solution by solving in Fourier space.

We consider a PDE of the form,

$$u_t = \mathcal{L}u + \mathcal{N}(u,t), \tag{5.2}$$

where $\mathcal{L}$ is a linear operator and $\mathcal{N}$ is a non-linear operator. A system of ODEs can be obtained by discretising the spatial part of the PDEs,

$$u_t = \mathbf{L}u + \mathbf{N}(u,t). \tag{5.3}$$

In the form of (5.3) we can write the a semi-discreet version of the KS equation in Fourier space as

$$\hat{u}_t = (\mathbf{L}\hat{u})(k) + \mathbf{N}(\hat{u},t), \tag{5.4}$$

where

$$(\mathbf{L}\hat{u})(k) = (k^2 - k^4)\hat{u}(k), \tag{5.5}$$

and

$$\mathbf{N}(\hat{u},t) = \mathbf{N}(\hat{u}) = \frac{-ik}{2}(F((F^{-1}(\hat{u}))^2)). \tag{5.6}$$

In this form the equation can be solved using an exponential time differentiating Runge-Kutta 4 (ETDRK4) numerical scheme [Cox and Matthews, 2000]. The scheme is given in equations (5.7) to (5.10).

$$a_n = e^{\frac{\mathbf{L}h}{2}}u_n + \mathbf{L}^{-1}(e^{\frac{\mathbf{L}h}{2}} - I)\mathbf{N}(u_n, t_n), \tag{5.7}$$

$$b_n = e^{\frac{\mathbf{L}h}{2}}v_n + \mathbf{L}^{-1}(e^{\frac{\mathbf{L}h}{2}} - I)\mathbf{N}(a_n, t_n + \tfrac{h}{2}), \tag{5.8}$$

$$c_n = e^{\frac{\mathbf{L}h}{2}}a_n + \mathbf{L}^{-1}(e^{\frac{\mathbf{L}h}{2}} - I)(2\mathbf{N}(b_n, t_n + \tfrac{h}{2}) - \mathbf{N}(u_n, t_n)), \tag{5.9}$$

$$u_{n+1} = e^{\mathbf{L}h}u_n + h^{-2}\mathbf{L}^{-3}\{[-4 - \mathbf{L}h + e^{\mathbf{L}h}(4 - 3\mathbf{L}h + (\mathbf{L}h)^2)]\mathbf{N}(u_n, t_n) + 2[2 + \mathbf{L}h + e^{\mathbf{L}h}(-2 + \mathbf{L}h)](\mathbf{N}(a_n, t_n + \tfrac{h}{2}) + \mathbf{N}(b_n, t_n + \tfrac{h}{2})) + [-4 - 3\mathbf{L}h - (\mathbf{L}h)^2 + e^{\mathbf{L}h}(4 - \mathbf{L}h)]\mathbf{N}(c_n, t_n + h)\}. \quad (5.10)$$

Kassam and Trefethen [2005] were aware that this scheme suffers from numerical instability. This numerical instability is due to the terms in square brackets in equation (5.10). These coefficients are higher order equivalents of the function

$$g(z) = \frac{e^z - 1}{z}, \quad (5.11)$$

which for small values of z sufferers from cancellation error. To help reduce instability Kassam and Trefethen [2005] introduce a new way to approximate the terms in the form of (5.11), that makes use of complex analysis.

Their method evaluates (5.11) using a contour integral, the contour being in the complex plane, enclosing z and being well separated from 0. The contour Γ is chosen to be a circle of radius one, centred at z. Cauchy's integral formula [Nevanlinna, 1969],

$$f(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(t)}{t - z}dt. \quad (5.12)$$

is used to integrate over the contour. By first substituting equation (5.11) for $f(t)$ and then the choice of Γ, $t = z + e^{i\theta}$, into (5.12) we obtain,

$$\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_\Gamma \frac{e^t - 1}{t(t - z)}dt \\ &= \frac{1}{2\pi i}\int_0^{2\pi} \frac{(e^{z+e^{i\theta}} - 1)ie^{i\theta}}{(z + e^{i\theta})e^{i\theta}}d\theta \\ &= \frac{1}{2\pi}\int_0^{2\pi} \frac{(e^{z+e^{i\theta}} - 1)}{z + e^{i\theta}}d\theta \end{aligned} \quad (5.13)$$

This contour integral can now be approximated using the periodic trapezoid rule [Trefethen, 2000]

$$f(z) \approx I_N = \frac{2\pi}{N} \sum_{j=1}^{N} f(\theta_j), \tag{5.14}$$

where $\theta_j = \frac{j\pi}{N}$. The approximation of $f(z)$ is now

$$f(z) \approx \frac{1}{N} \sum_{j=1}^{N} \frac{(e^{z+e^{i\theta_j}} - 1)}{z + e^{i\theta_j}}. \tag{5.15}$$

This is just the mean of the function $f(z)$ evaluated at a number of different points around the circular contour. It is suggested that 32 points around the circle are sufficient to approximate this integral well, the number of points required is further reduced due to the $\pm i$ symmetry. To approximate $f(z)$ using (5.15) it is sufficient to consider 16 points in the upper half plane and take the real part of the result. This contour integral method approximates solutions to equations of the form in (5.11) well and reduces the numerical instability of the scheme.

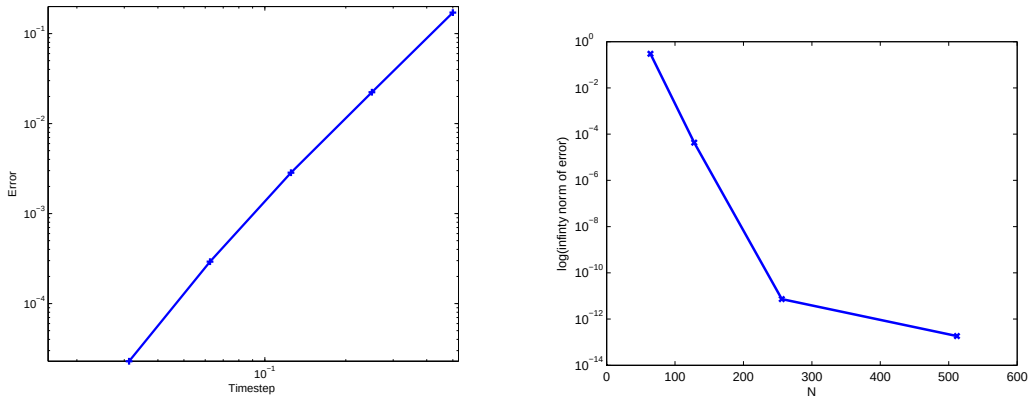
The ETDRK4 scheme is fourth order accurate in time and spectrally accurate in space. We wish to show that the scheme converges as expected when used to solve the KS equation and we use code provided by Kassam and Trefethen [2005] to do this.

5.1.1.1 Convergence of the ETDRK4 scheme

As there is no analytic solution to the equation we show convergence by taking a high resolution run of the code to be our truth. We then define the error to be the difference between this truth and a low resolution run.

To prove convergence in time we first create a truth run. We consider the solution to this equation on the periodic domain $0 \leq x \leq 32\pi$ from initial conditions $u = \cos(\frac{x}{16})(1 + \sin(\frac{x}{16}))$, and we use $N^m = 256$ spatial points and a time step $\Delta t = 0.015$. We then calculate the error from lower temporal resolution runs. For these lower resolution runs we fix the number of spatial points to $N^m = 256$ and vary the time step between $\Delta t = 0.015$ and $\Delta t = 0.5$. For convergence in space we use a truth run where the number of spatial

points is $N^m = 1024$ and the time step is $\Delta t = 0.03$. For the low spatial resolution runs we fix the time step to $\Delta t = 0.03$ and vary the number of spatial points between $N^m = 512$ and $N^m = 64$ to show the spectral convergence. We plot the L_∞ norm of the error against the time step and the log of the L_∞ norm of the error against the number of spatial points to show convergence. The time and space convergence plots are shown in figures 5.1(a) and 5.1(b) respectively.



(a) Fourth order convergence in time

(b) Spectral convergence in space

Figure 5.1: Spatial (a) and temporal (b) convergence of the ETDRK4 scheme

We see from Figure 5.1(a) that the ETDRK4 scheme is fourth order in time as expected. The first three points in Figure 5.1(b) show convergence as expected. We then see that the final point does not reduce the error as much as expected. To check that this was not due to a violation of the CFL condition, we consider the result when we use a smaller fixed time step. We find that the convergence is similar to that shown in Figure 5.1(b). From this we conclude that the smaller reduction in error for $N^m = 512$ is most likely due to computational rounding error.

5.1.1.2 Solution existence

While considering convergence it was found that a resolution of $N^m = 16$ and $N^m = 32$ spatial points was not sufficient to provide the multi-scale chaotic solution to the KS equation. To determine why this was the case we considered the solution to the equation at various points in time. In Figure 5.2 we show why, with a small number of spatial

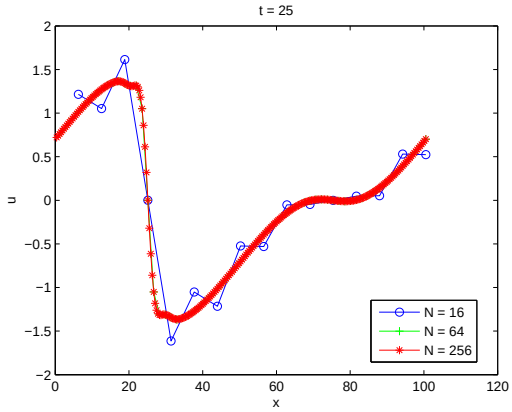


Figure 5.2: Different resolution solutions to the KS equations at $t = 25$. The $N^m = 16$ plot (blue) shows that this number of spatial points is not sufficient to capture the sharp gradients in the solution. Note that the $N^m = 64$ (green) solution lies directly under the $N^m = 256$ (red) solution

points, a full solution cannot be obtained.

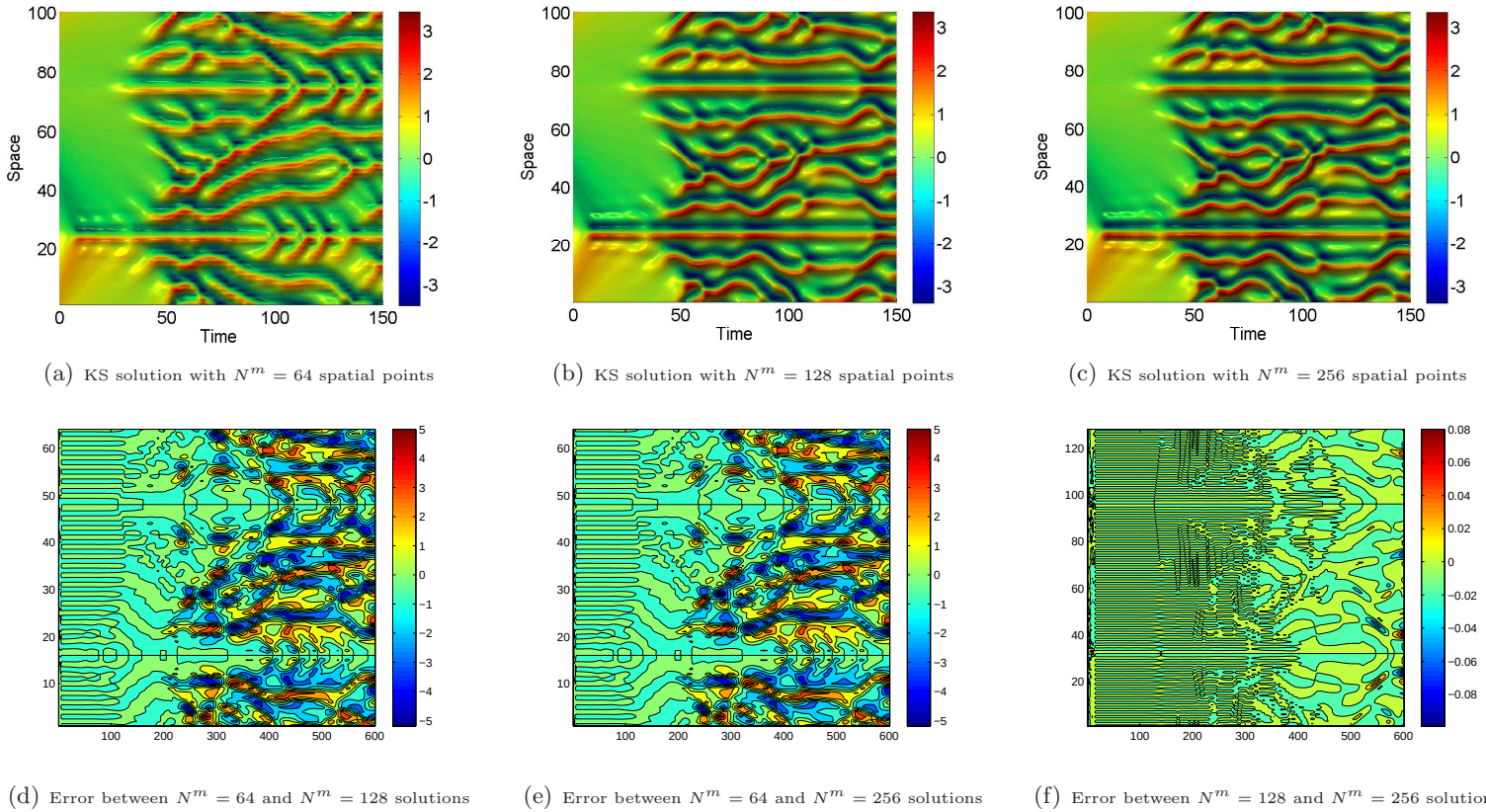
We see from Figure 5.2 that when a small number of spatial points ($N^m = 16$) is used there are not enough points to fully resolve the sharp gradients in the solution. Over time the point at the top and bottom of the gradient tend to plus and minus infinity to try account for the steep gradient. As these points tend to infinity the rest of the solution is affected, and eventually the whole solution blows up so a full solution can not be obtained.

5.2 Understanding the differences between solutions at different resolutions

5.2.1 Solutions at different resolutions

Before we calculate representativity errors we first compare different resolution runs of the model. We keep the time step fixed, $\Delta t = \frac{1}{4}$, and run our model for $N^m = 64, 128, 256$ spatial points.

The time evolutions of the solutions to the KS equation for $N^m = 64, 128, 256$ when $\Delta t = 0.25$ are shown in Figures 5.3(a), 5.3(b) and 5.3(c). We see that the solutions for $N^m = 128$ and $N^m = 256$ appear similar. We quantify this in Figure 5.3(f) where we plot



(a) KS solution with $N^m = 64$ spatial points

(b) KS solution with $N^m = 128$ spatial points

(c) KS solution with $N^m = 256$ spatial points

(d) Error between $N^m = 64$ and $N^m = 128$ solutions

(e) Error between $N^m = 64$ and $N^m = 256$ solutions

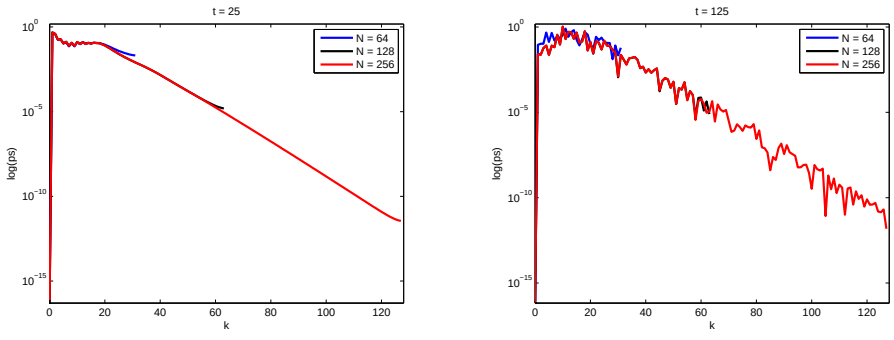
(f) Error between $N^m = 128$ and $N^m = 256$ solutions

Figure 5.3: Solutions to the Kuramoto-Sivashinsky equation at different resolution runs and the differences between them. Note the change in colourscales for plot d) and e) compared to f)

the error between these two solutions. We see that the errors are small with the errors only growing to ± 0.08 in the last 50 time steps. The solution for $N^m = 64$ is significantly different from the other two solutions as some of the smaller scale features are not captured. We again plot the error between the $N^m = 64$ and $N^m = 128$ solutions, Figure 5.3(d), and the error between the $N^m = 64$ and $N^m = 256$ solutions, Figure 5.3(e) to show how the solution with $N^m = 64$ differs from the higher resolution solutions. Both of these plots show very large errors, often larger than the value of the solution of the KS equation.

5.2.2 Power spectra

We wish to be able to understand the errors shown in Figures 5.3(d), 5.3(e) and 5.3(f). One way we can compare the solutions from different resolution runs is to consider the power spectra of the solutions. To obtain the power spectrum we calculate the square of the modulus of the Fourier transform coefficient and then scale by a factor of $\frac{2}{N^m}$. The power spectrum enables us to determine what portion of a signal's power falls at each wave number. In Figure 5.4 we plot the power spectra of the KS equation solutions with $N^m = 64$, $N^m = 128$ and $N^m = 256$ spatial points at times $t = 25$, where the solution is relatively smooth, and at $t = 125$ when the solution has evolved. In Figure 5.4(a) we see that the power in the highest wave numbers of the $N^m = 64$ solution is greater than the power in these wave numbers for the $N^m = 128$ and $N^m = 256$ solutions. We see the same



(a) Power spectra of the solution of the KS equation at $t = 25$

(b) Power spectra of the solution of the KS equation at $t = 125$

Figure 5.4: Power Spectra of the solutions of the KS equation. The power spectra of the solutions with $N^m = 64$, $N^m = 128$ and $N^m = 256$ spatial points are given by the blue, black and red lines respectively.

behaviour occurring with the power in the highest wave numbers of the $N^m = 128$ solution being greater than the power in these wave numbers for the $N^m = 256$ solution. This increase in power compensates for the power in higher wave numbers that the low resolution models cannot capture. This additional power relates to our error of representativity as it represents the small scales that cannot be resolved. We see similar behaviour in Figure 5.4(b) where power from higher wave numbers that cannot be captured is added to power in the highest wave numbers of a given resolution. As the solution is more chaotic the power spectrum is less simple to analyse. We would like to better understand how the distribution of power over different wave numbers can help us understand the structure of errors of representativity. We shall now consider these errors of representativity.

5.3 Understanding time independent representativity error

Now we have a numerical model we can use it to help us understand the structure of representativity error. We use the method described in Chapter 3 Section 3.3.1 to calculate both forward model and representativity error. We begin by describing the experimental design.

5.3.1 Experiment design

We first must define our truth. We solve the KS equation on the periodic domain $0 \leq x \leq 32\pi$ from initial conditions $u = \cos(\frac{x}{16})(1 + \sin(\frac{x}{16}))$. We use $N^t = 256$ spatial points and a time step $\Delta t = 0.015$. We calculate forward model error for a varying number of model points, $N^m = 32, 64$ and observations $p = 16, 32, 64$. It is possible to calculate representativity error for $N^m = 32$ as it is not necessary to run the forward model, hence it does not matter that the model is not stable at this resolution. Using the values on N^m and p the Fourier matrices can be defined as in equation (3.5). The truncation matrix is also determined by N^t and N^m. We use the weighting matrices defined in equations (4.2) and (4.4) with lengthscales $L_o = 2.0$ and $L_o = 4.0$. The spectral covariance matrix is calculated using equation (3.13) which requires the covariance matrix, $\mathbf{S}$ that can be calculated using

(4.5). For this we require a number of samples. We generate our samples by running our high resolution model for 80,000 time steps. After removing data from the burn-in period, the solutions of the equation before $t = 50$ where the chaotic nature of the solutions has not developed, we take the solution every 10^{th} time step to be a sample. We then use the MATLAB `cov`, `corrcoef` and `std` functions to calculate the covariance, correlation and standard deviation matrices from these samples. We plot five rows of the correlation matrix in Figure 5.5 to help us understand the structure of the matrix (note that this figure is plotted on a different scale to Figure 5.3). For each row the diagonal position is plotted in the centre of the graph and the size of the off-diagonal row elements are plotted symmetrically. We see that the covariance of the truth has a wave like structure, where the wave length in the correlation matrix is equal to the wavelength in the KS equation.

When analysing the calculated covariance matrix we find that it is not full rank. This is because the number of samples used to calculate the matrix is not large enough to provide all the information required. This problem can be dealt with by reducing the number of covariance parameters that need to be estimated. One way to do this is to assume that the covariances are homogeneous and isotropic. This results in a covariance matrix that is diagonal in spectral space, which is required for the matrix to be used in the Daley [1993] method. We see from Figure 5.5 that the correlation matrix is neither isotropic or homogeneous. At small separation distances the rows are similar and appear

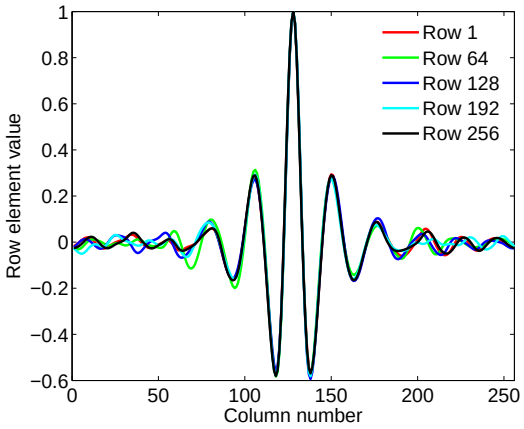


Figure 5.5: Five rows of the true correlation matrix

symmetric, at larger separation distances the rows appear less similar due to the reduced number of samples available to calculate the covariance. However, as the rows are similar and nearly symmetric it suggests that the matrix should be isotropic and homogeneous, and is prevented from being so by the sampling error. To make our matrix isotropic and homogeneous we take the first row of the calculated matrix, make it symmetric and then create a circulant matrix from this. The matrix is circulant as the correlations are homogeneous.

We now have a homogeneous and isotropic correlation matrix. However, the matrix does still contain sampling errors. One way to compensate for the sampling error and improve the correlation matrix is to use the technique of covariance localisation that was described in Chapter 2, section 2.3.5. As our localisation function we use the fifth order piecewise polynomial described in Gaspari and Cohn [1999] equation (4.10),

$$L(x,c) = \begin{cases} -\frac{1}{4}(\frac{|x|}{c})^5 + \frac{1}{2}(\frac{x}{c})^4 + \frac{5}{8}(\frac{|x|}{c})^3 - \frac{5}{3}(\frac{x}{c})^2 + 1 & \text{if } 0 \leq |x| \leq c, \\ \frac{1}{12}(\frac{|x|}{c})^5 - \frac{1}{2}(\frac{x}{c})^4 + \frac{5}{8}(\frac{|x|}{c})^3 + \frac{5}{3}(\frac{x}{c})^2 - 5(\frac{|x|}{c}) + 4 - \frac{2}{3}(\frac{c}{|x|}) & \text{if } c < |x| \leq 2c, \\ 0 & \text{if } 2c < |x|, \end{cases} \tag{5.16}$$

where c is the length scale. We choose this to be 16 as it allows the localizing function to capture the important features from the true correlation. This localisation function reduces the value of the correlations in Figure 5.5 where the rows do not have the same value (where the sampling error is more dominant). We plot this localisation function in Figure 5.6 along with the middle row of the correlation matrix calculated initially and the middle row of the localized correlation matrix. The effect of the localisation function on the correlation function is obvious. The localized correlation function is now zero where the localisation function is zero, but still resembles the true correlation function where the correlations are large in magnitude. It is clear that the localizing function has removed the sampling error from our correlation matrix estimate. We are now left with a good estimate of the correlation matrix. From this a full rank, isotropic and homogeneous covariance matrix can be calculated. When transformed in to spectral space this covariance matrix is real

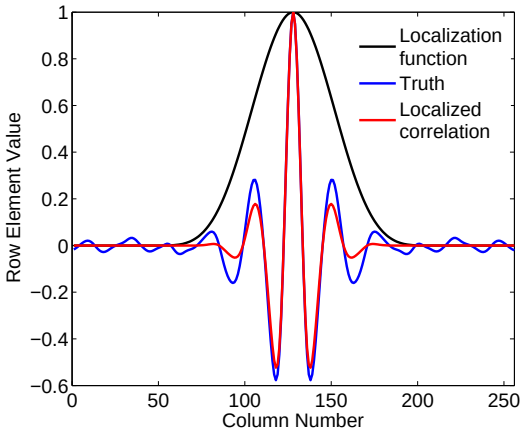


Figure 5.6: Comparison of the middle row of the covariance matrix (blue) and localized covariance matrix (red). The localizing function (black) is included to show the important length scale.

and diagonal. We can now use this covariance matrix to calculate the spectral covariance matrix. In turn this can be used to calculate the errors of representativity.

5.3.2 Numerical results

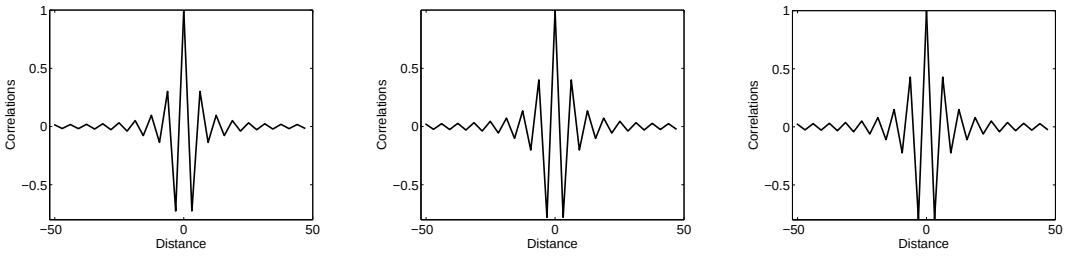
We now carry out a number of experiments to enable us to understand the structure of forward model error and representativity error. We calculate the representativity error using equation (3.12). We begin by considering the case where there is no error in the observation operator $\mathbf{H}$, and therefore the calculated error consists of only representativity error. We present the magnitude of the representativity error variances for the experiments carried out in Table 5.1.

5.3.2.1 Changing the observation type

First we consider the structure of representativity error when different observation types are used. Initially we consider direct observations. We assume that the model has 32 grid points, and that each of these grid points has an associated observation. The variance of the representativity error is given in Table 5.1 Experiment 1, and the correlations are plotted in Figure 5.7(a).

Experiment Number	Truncation	Number of Observations (p)	True Observation type (Length scale)	Assumed Observation Type (Length scale)	RE variance
1	32	32	Direct	Direct	2.81×10^{-1} (16.3%)
2	32	32	Uniform(2.0)	Uniform(2.0)	1.65×10^{-1} (9.6%)
3	32	32	Gaussian(2.0)	Gaussian(2.0)	1.35×10^{-1} (7.9%)
4	32	32	Gaussian(4.0)	Gaussian(4.0)	1.92×10^{-2} (1.1%)
5	32	16	Direct	Direct	2.81×10^{-1} (16.3%)
6	64	64	Direct	Direct	8.17×10^{-3} (0.5%)
7	32	32	Uniform(2.0)	Uniform(4.0)	2.36×10^{-1} (13.7%)
8	32	32	Gaussian(2.0)	Gaussian(4.0)	2.35×10^{-1} (13.7%)
9	32	32	Gaussian(2.0)	Uniform(2.0)	1.36×10^{-1} (7.9%)

Table 5.1: Representativity error (RE) variance for the KS equation. The values given in brackets with the representativity error variance are a comparison of the representativity error variance to the variance if the high resolution solution. Experiment details are given in the text.



(a) Direct Observations (b) Uniform Observations, $L_o = 2.0$ (c) Gaussian Observations, $L_o = 2.0$

Figure 5.7: Comparison of errors of representativity for observations calculated using different observation operators. The number of model grid points is $N^m = 32$ and every gridpoint is observed.

We first consider the correlation structure of the representativity error and we see that the errors are correlated. The variance of the representativity error is reasonably large, a little more than 16.0% of the variance of the high resolution state in the case of direct observations and approximately 10% of the values of the high resolution solution in the case of uniform observations. This suggests that in the case of direct observations the representativity error may be significant.

We now consider the representativity error when the observations are defined using a uniform weighting function with length scale $L_o = 2.0$. The variance is given in Table 5.1 Experiment 2, and the correlations are plotted in Figure 5.7(b). We see that the error is correlated, and the structure does not differ significantly from the structure when direct observations are used. However, we see that the variance has decreased, and that the representativity error is less significant for the error calculated with uniform observations. This is as expected as the uniform observations do not capture all the small scales that the direct observations can. We now consider the results where Gaussian observations, with $L_o = 2.0$ are used. We present the results in Table 5.1 Experiment 3, and the correlations are plotted in Figure 5.7(c). We expect the results to be similar to those where uniform weighted observations are used as the observations have the same length scale. We see that the variances are similar, with the variance for the Gaussian observations being slightly smaller. This is because the Gaussian weighting function averages over a slightly larger area, which means that the observations resolve fewer scales. We see again that the error

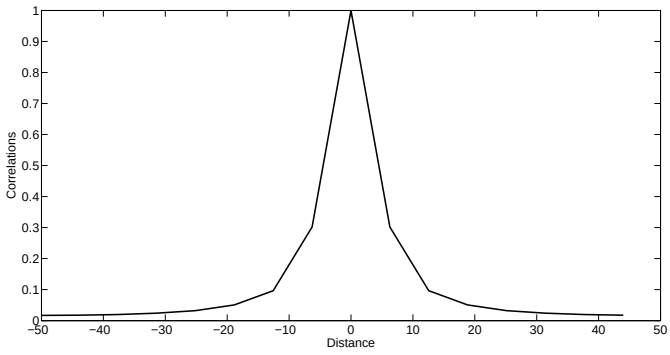


Figure 5.8: Representativity error correlation when the model resolution is $N^m = 32$ and half the gridpoints are observed.

is correlated. However, the structure of the correlations is similar to those found where direct and uniform observations were used.

We also consider the structure of representativity error when the observations are defined using a Gaussian, but the length scale is increased to $L_o = 4.0$. As these observations are averaged over a larger area of the domain we expect fewer scales to be resolved, and the representativity error to decrease. We give the results in Table 5.1 Experiment 4. We see that as expected the variance of the representativity error is much smaller and less significant.

5.3.2.2 Number of observations

We now consider what happens when we calculate the representativity error where fewer direct observations are available. Experiment 5 in Table 5.1 shows the error variance where only half the model grid points are directly observed, the correlations are plotted in Figure 5.8. We see that having fewer observations available does not alter the variance of the representativity error. This is expected as we have proved in Chapter 4 Section 4.3 that representativity error applies individually to each observation and is independent of other observations. From Figure 5.8 we see that the correlation no longer has a wave like structure, this is a result of the observation resolution. The observation resolution is such that only the tops of the waves in Figure 5.3 are sampled. By comparing Figures 5.7(a) and 5.8 we see that the correlation structure depends only on the distance between

observations and not the number of observations available. This supports the theoretical results given in Chapter 4 Section 4.3.

5.3.2.3 Number of model grid points

We now consider the effect of changing the model resolution. The higher the model resolution, the more scales we expect that the model can resolve. The more scales a model can resolve, the smaller the representativity error should be. We calculate the representativity error using direct observations. However, we set our model to have 64 grid points, double the resolution of the previous experiments. We present the result in Table 5.1 Experiment 6. We see that the variance of the representativity error has reduced significantly, and is now only 0.5% of the true variance. This supports the conclusion that models with higher resolution have smaller representativity errors.

5.3.2.4 Forward model error

So far we have calculated forward model error that consisted only of representativity error. We now calculate forward model error when there are contributions from both representativity error and error in the observation operator. Due to this error we expect the forward model error to increase. In Experiments 7 and 8 we introduce error in the observation operator by assuming the model observations have a different length scale to the high resolution observations. In both these cases, uniform and Gaussian observations, the forward model error is much larger than when the observation error is exact. We also consider the case when the observations are modelled with the wrong weighting function, but with the same length scale. In Experiment 9 the true observations are created using a Gaussian weighting function, but the model observations are calculated with a uniform weighting function. In this case we find that there is no great increase of the variance of the forward model error. The introduction of error in the observation operator increases the forward model error. However, it appears that less error is introduced when the length scale of the observations is correctly modelled.

5.4 Summary

In this chapter we have used the Kuramoto-Sivashinsky equation and the method proposed by Daley [1993] to help us understand the structure of representativity error. We started by describing the KS equation, which has a solution that exhibits chaotic and multi-scale behaviour. We then described the ETDRK4 method, a numerical method that we used to solve the KS equation. We demonstrated convergence of the scheme and showed that a certain spatial resolution is required for the numerical scheme to be stable. We considered solutions to the KS equation at different resolutions, and the power spectra of these solutions. We then calculated representativity error for the KS equation. We found that representativity error is correlated. We showed that representativity error reduced as the length scale of the observation increased. This is because observations with larger length scales resolve fewer scales, and therefore the difference between the resolved scales in the observation and model is reduced, hence reducing the representativity error. We also showed that representativity error variance does not depend on the number of observations available. The correlations are also independent of the number of observations, and only dependent on the distance between observations. This supported the theoretical results given in Chapter 4. We also showed that representativity error was reduced when model resolution was increased. This is because a model at higher resolution resolves more of the scales that are resolved by the observations. We then calculated forward model error for the KS equations. We introduced error into the observation operators by either assuming the wrong observation lengthscale or the incorrect observation type. We found that there was a large contribution to the forward model error when the wrong observation lengthscale was assumed. However when the observations were assumed uniform, where they were in fact Gaussian, the forward model error was not much greater than when the observation operator was correct. We have calculated representativity error for a simple system. We now consider if the conclusions found in this chapter hold in the context of NWP.

Chapter 6

Representativity Error for Temperature and Humidity Using the Met Office High Resolution Model

In this chapter we use the Daley [1993] method described in Section 3.3.1 to calculate forward model error for data from the Met Office UKV model. The observation operator is always correctly specified so the forward model errors consist only of errors of representativity. We wish to to show that our conclusions about representativity error hold in the context of NWP. To verify this we carry out a number of experiments similar to those in the previous chapter. We also investigate the significance of representativity error for both temperature and specific humidity over the UK. Previous work has shown that observation error statistics are correlated for certain observation types [Stewart et al., 2009, 2012b, Bormann et al., 2002, Bormann and Bauer, 2010, Bormann et al., 2010]. We consider whether these significant correlations in the observation error covariance matrix could be attributed to representativity error. We note that the work in this chapter has been submitted for publication and is available as a preprint [Waller , née Pocock].

6.1 The model and data

We begin by introducing the model and available data. The calculation of representativity error by the method of Daley [1993] assumes that the actual state can be taken from a high resolution model. We use data from the Met Office UK variational (UKV) model as a proxy for our actual state. The UKV model is a variable resolution model that covers the UK. The model has fixed regular grid on the interior with 1.5km square grid boxes. The regular grid is surrounded by a variable resolution grid where grid boxes smoothly increase in size to 4km. For this study we consider two sets of data, previously used in Pavelin et al. [2009]. The data covers sub-domains, each of 450km $\times$ 450km (300×300 grid points with 1.5km grid boxes), of the UKV model. The lateral boundary conditions for the 1.5km models are taken from a 4km resolution regional model which is nested in the 12km model that covers the North Atlantic and Europe (NAE). The boundary conditions blend into the 1.5 km model field over a transition zone of 10km [Pavelin et al., 2009] and we therefore exclude the data in this region from our study.

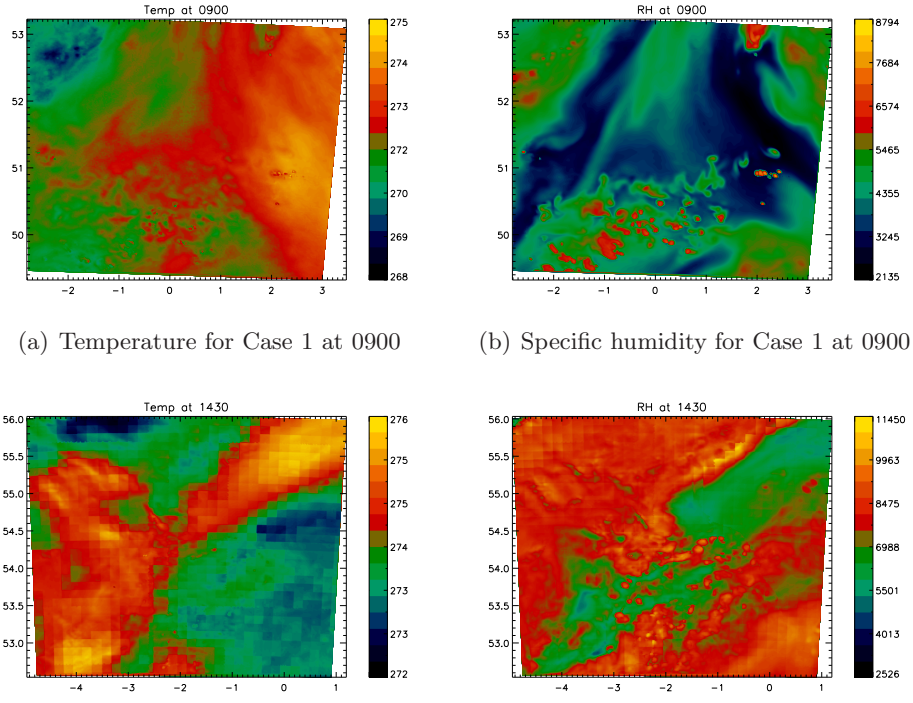
Since we are considering representativity error it is also necessary to ensure that the model spectra have fully adjusted to the higher spatial resolution. This is not fully understood for this suite of models. However, qualitative measures of the distance it takes convection to spin up due to features advecting in from the boundaries are given in Lean et al. [2008], Tang et al. [2012] and Kendon et al. [2012]. We remove further data from the boundary so that approximately 30km are removed in total. We expect the 1.5km model to be spun up from the 4km boundary conditions by this distance, although this is would not be guaranteed for a rapidly changing synoptic situation.

In this work we calculate representativity error using the assumption that the model state is a truncation of high resolution data. For the majority of our experiments we truncate the data so the model grid spacing is equivalent to the grid spacing that is used in the Met Office NAE model. The Met Office NAE model has a grid spacing of 12km (in mid-latitudes) and covers Europe and the North Atlantic.

6.1.1 The data available

We use temperature and humidity data over the UK from two cases. The first case consists of data from 7^{th} August 2007 at times 0830UTC, 0900UTC and 0930UTC on an area over the southern UK that covers -3.04^oW to 3.71^oE and 49.18^oN to 53.36^oN. In this case there are partly clear skies with convection occurring over the south east [Eden, 2007]. The second set of data is from 6^{th} September 2008 at 1400UTC, 1430UTC and 1500UTC and covers -5.00^oW to 1.20^oE and 52.5^oN to 56.00^oN. In this case a deep depression is tracking slowly east-northeast across England [Eden, 2008]. The data is available on a 300×300 square of a latitude and longitude grid at each of 43 pressure levels. We plot the temperature and humidity data for the 749hPa pressure level for the first case at time 0900UTC and the second case at time 1430UTC in Figure 6.1.

To calculate representativity error using the Daley [1993] method we require the spectral covariance matrix for the truth. In Chapter 4 we showed that it is possible to calculate



(a) Temperature for Case 1 at 0900

(b) Specific humidity for Case 1 at 0900

(c) Temperature for Case 2 at 1430

(d) Specific humidity for Case 2 at 1430

Figure 6.1: Temperature (a & c) and humidity fields (b & d) for Case 1 (a & b) at time 0900 and Case 2 (c & d) at time 1430

the spectral covariance using the covariance of the high resolution data. In equation (4.5) we showed that it is possible to calculate the covariance of the high resolution data using a number of samples. We now describe how we create samples from the available data.

6.1.2 Creating samples from the data

There are some limitations to the data. Data near the boundary is contaminated by the boundary conditions taken from the coarser model. We remove this data by reducing the grid to a 256×256 mesh centred on the main grid. We need to sample the data to calculate the covariance matrices for the actual state. The data we have is available on a 2D gridded domain; however, the Daley [1993] method provides an equation to calculate representativity error on a 1D domain. To convert our data to 1D we take the individual rows of the data from the 749hPa pressure level. We use this level as it is outside the lower boundary layer and therefore not affected by complex processes such as small scale turbulence, but should still include the small scale features that are relevant when calculating representativity error. We consider temperature and the natural logarithm of specific humidity data for each of the two synoptic cases. For each synoptic case we have 256 samples at three different times and, therefore, we have 768 samples to calculate the covariance matrices. A covariance calculated with this number of samples is dominated by sampling error and hence this is not a sufficient number of samples to calculate an accurate representation of the required covariances. One way to overcome this would be to take data from more times. However, this will reduce the time dependence of the calculated representativity error. A further problem is that the samples are not periodic, but the Daley [1993] method assumes a periodic domain with a circulant $\mathbf{S}$. To overcome this and to increase the number of samples we detrend and process the data.

6.1.3 Data processing

To create surrogate samples from each of the available samples the data must be detrended. Detrending gives data on a homogeneous field; this is required by our chosen method for

calculating representativity error. Data is detrended by removing a best fit line using an appropriate polynomial of order no greater than 3 [Bendat and Piersol, 2011]. It is justifiable to detrend the data as only trends with large length scales are removed. All scales that contribute to the representativity error still remain. We detrend the 256 latitude samples at each available time. Different orders of polynomial were considered for detrending and the lowest order polynomial that resulted in homogeneous data was chosen. A linear trend was removed from the temperature data, and a cubic trend from the log of the specific humidity data. Removing polynomials of higher order had little effect on the representativity error results. This detrended data is now used to create new samples from each existing sample.

The method of Fourier randomization is used to generate surrogate samples from the same statistical disribution [Theiler et al., 1992, Small and Tse, 2002]. Fourier randomization consists of perturbing the phase of a set of data to create a new sample with a different phase, but where each wave number retains the same power. As the power spectrum of the sample is unchanged the linear covariances are preserved. Therefore any choice of phase shift should result in data with the same covariance. As the covariance is preserved we do not expect the choice of phase shift to affect the results when representativity error is calculated. Here we calculate circulant samples, which corresponds to shifting the phase of the data. This also gives the data the required periodicity. A circulant sample is created by shifting each element of the sample one position and taking the final element and making it the first entry in the sample. Each element can be shifted to each position, which means a sample with n elements can be used to create n circulant samples. Therefore creating surrogate samples increases the number of available samples we have for calculating the covariance of the high resolution data. We have available 256 samples at three different times. Creating circulant samples gives us 65536 samples at each time, and a total of 196608 samples to estimate each of the covariance matrices, which is a sufficient number of samples.

6.2 Experiments

We use equation (4.5) and the circulant samples calculated from the UKV model data to calculate the covariance matrices for the temperature and humidity fields at the 749hPa pressure level for both cases. We give the variances in Table 6.1 and plot a row of each of the true state correlation matrices in Figure 6.2.

	Temperature (K^2)	log(Specific Humidity) (kg^2/kg^2)
Case 1	0.6638	0.0812
Case 2	0.1934	0.0178

Table 6.1: Variances for the true state at the 749hPa pressure level

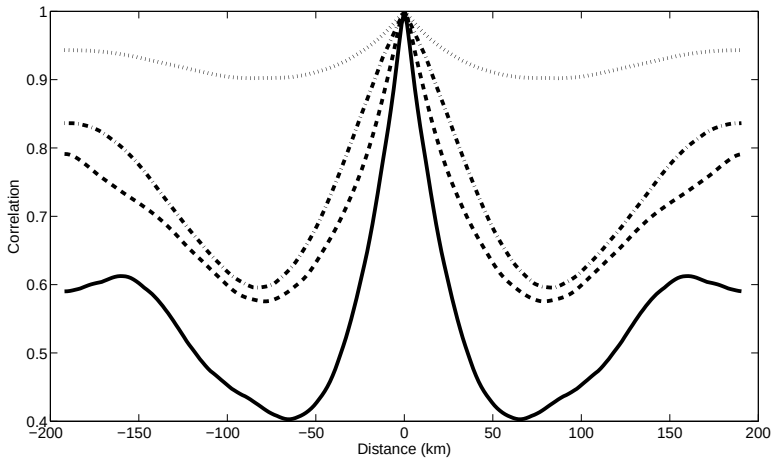


Figure 6.2: Correlation structure for the true temperature and specific humidity fields at the 749hPa pressure level. Temperature: Case 1 dotted line, Case 2 dot dash line. Specific Humidity: Case 1 dashed line, Case 2 solid line

From Table 6.1 we see that the variances for Case 2 are smaller than those for Case 1. When considering the correlations plotted in Figure 6.2, we see that the temperature fields have larger correlations than humidity. For Case 1 the temperature correlations are very high; this is expected as, after detrending, this field is fairly constant. We also note that the correlations for Case 2 are smaller than the correlations for Case 1. This is due to the synoptic situation. In Case 1 the field is more homogenous with small scale features over the domain. However, in Case 2 the features are large scale and less homogeneous.

For the estimates of representativity error to be exact we require the correct covariances of the actual state. As our truth we are using data from the UKV model, and therefore our estimates of the covariances will only be as accurate as the spectra of the UKV model. As the UKV does not resolve all the scales in the truth it is likely that the estimates of representativity error given by the Daley [1993] method will be an underestimate. However, as the UKV model gives a reasonable estimate of the truth and we are measuring the loss of information between the low and high resolution models we can still expect to understand more about the behaviour and structure of representativity error.

Again we use equations (4.2) and (4.4) to define high resolution pseudo observations. We use the uniform curve with a width of approximately 5km and a Gaussian curve with a width of approximately 20km. We also consider in-situ measurements where the diagonal elements $\hat{w}$ of the weighting matrix are all unity.

Now we have the appropriate weighting matrices and the covariance matrices for the high resolution data at the 749hPa pressure level. This allows us to calculate representativity errors for temperature and log specific humidity. In the next section we present the results of our experiments.

6.3 Results

We now carry out a number of experiments to enable us to understand the nature of representativity error. As in Chapter 5 we present initially only the magnitude of the representativity error variance. The results for experiments carried out with data from Case 1 are given in Table 6.2, and for Case 2 in Table 6.3.

6.3.1 Temperature and humidity representativity errors

We first consider how the errors of representativity differ between the temperature and log humidity fields. We consider the representativity error for the case where the model has 32 points. This relates to a grid spacing of 12km equivalent to the grid spacing that

Experiment Number	Truncation	Number of Observations (p)	Observation type	Temperature RE variance ($(\mathrm{K})^2$)	Humidity RE variance ($(\mathrm{kg/kg})^2$)
1.1	32	32	Direct	4.81×10^{-3} (0.7%)	1.51×10^{-3} (1.9%)
1.2	32	32	Uniform	2.71×10^{-3} (0.4%)	1.08×10^{-3} (1.3%)
1.3	32	32	Gaussian	8.99×10^{-4} (0.1%)	3.80×10^{-4} (0.5%)
1.4	32	16	Direct	4.81×10^{-3} (0.7%)	1.51×10^{-3} (1.9%)
1.5	32	16	Uniform	2.71×10^{-3} (0.4%)	1.08×10^{-3} (1.3%)
1.6	32	16	Gaussian	8.99×10^{-4} (0.1%)	3.80×10^{-4} (0.5%)
1.7	64	64	Direct	2.13×10^{-3} (0.3%)	4.04×10^{-4} (0.5%)
1.8	64	64	Uniform	5.40×10^{-4} (0.1%)	1.71×10^{-4} (0.2%)
1.9	64	64	Gaussian	2.76×10^{-5} (0.0%)	1.07×10^{-5} (0.0%)

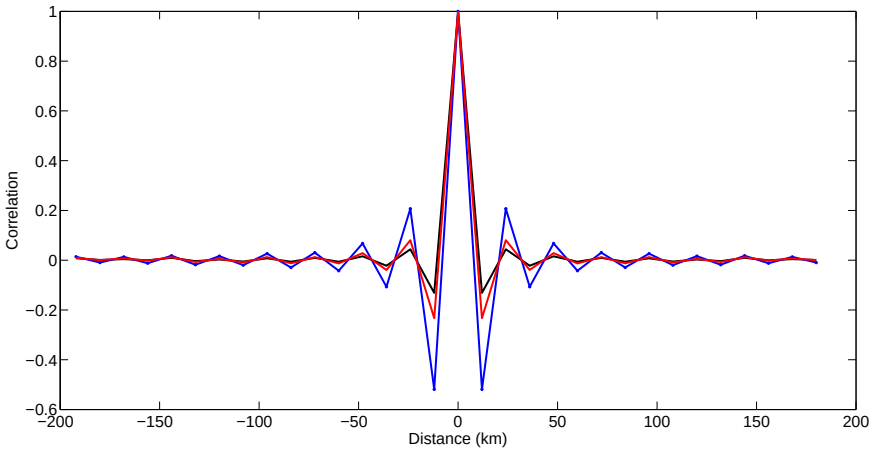
Table 6.2: Representativity error (RE) variances for Case 1 at the 749hPa pressure level. The values given in brackets are a comparison of the representativity error variance to the high resolution data variance.

Experiment Number	Truncation	Number of Observations (p)	Observation type	Temperature RE variance ($(\mathrm{K})^2$)	Humidity RE variance ($(\mathrm{kg/kg})^2$)
2.1	32	32	Direct	2.21×10^{-3} (1.1%)	7.14×10^{-4} (4.0%)
2.2	32	32	Uniform	1.30×10^{-3} (0.7%)	4.84×10^{-4} (2.7%)
2.3	32	32	Gaussian	3.96×10^{-4} (0.2%)	1.60×10^{-4} (0.9%)
2.4	32	16	Direct	2.21×10^{-3} (1.1%)	7.14×10^{-4} (4.0%)
2.5	32	16	Uniform	1.30×10^{-3} (0.7%)	4.84×10^{-4} (2.7%)
2.6	32	16	Gaussian	3.96×10^{-4} (0.2%)	1.60×10^{-4} (0.9%)
2.7	64	64	Direct	1.12×10^{-3} (0.6%)	2.50×10^{-4} (1.4%)
2.8	64	64	Uniform	3.14×10^{-4} (0.2%)	8.81×10^{-5} (0.5%)
2.9	64	64	Gaussian	1.86×10^{-5} (0.0%)	5.36×10^{-6} (0.0%)

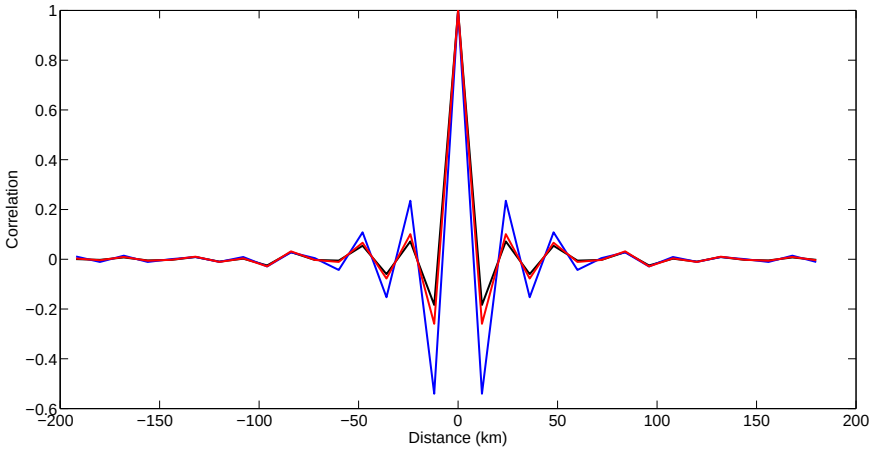
Table 6.3: Representativity error (RE) variances for Case 2 at the 749hPa pressure level. The values given in brackets are a comparison of the representativity error variance to the high resolution data variance.

is used in the Met Office NAE model. This is a truncation of a factor of eight from the high resolution model, which has 256 points. We start by assuming that we have direct observations. The values of the representativity error variance are given in Table 6.2 Experiment 1.1. We plot in Figures 6.3(a) and 6.3(b) (black lines) the middle row of the representativity error correlation matrices for temperature and log specific humidity from Case 1 respectively.

When we compare the variance of representativity error against the variance of the ac-



(a) Temperature (K)



(b) log(Specific humidity) (kg/kg)

Figure 6.3: Representativity error correlations between observation centre points for Case 1 with truncation to 32 points (12km resolution) with every model grid point observed using direct (black line), uniform-weighted (red line) and Gaussian-weighted (blue line) observations

tual states we see that representativity error is more significant for humidity than it is for temperature. We find that the representativity error variance for temperature is 0.7% of the high resolution temperature variance, whereas the humidity representativity error variance is 1.9% of the high resolution humidity field variance. When comparing the variance from this experiment to the same experiment carried out with Case 2 data (Table 6.3, Experiment 2.1), we see that the representativity error variances are smaller for Case 2. This is expected as there is less variance in the true fields in Case 2. These experiments show, however, that the representativity error is more significant in this case. The representavity error for temperature is 1.1% of the high resolution temperature variance and humidity representativity error is 4.0% of the high resolution variance. For Case 1 from Figures 6.3(a) and 6.3(b) (black line) we see that the correlation structure is similar for both temperature and specific humidity. The correlations rapidly decrease in magnitude as the observation separation distance increases. The correlations for specific humidity are slightly larger, and decay less rapidly than the correlations for temperature.

6.3.2 Changing the observation type

We now consider what happens where the observations are defined with a uniform weighting matrix. This uniform weighting acts on the temperature and log of specific humidity fields. The variance of the representativity error is given in Table 6.2 Experiment 1.2. We see again, as expected, that the representativity error is more significant for humidity than it is for temperature. We see that the assumption of uniformly weighted observations has decreased the representativity error for both fields when compared to Experiment 1.1. This supports the results seen in Chapter 5 that the representativity error decreases as the observation length scale increases. We see from Figures 6.3(a) and 6.3(b) (red line) that the correlations are larger than those when direct observations are used. This is because two consecutive observations have some overlap in physical space. We see from Table 6.3 that Experiment 2.2 supports these results as the representativity error variance is smaller than those seen in Experiment 2.1.

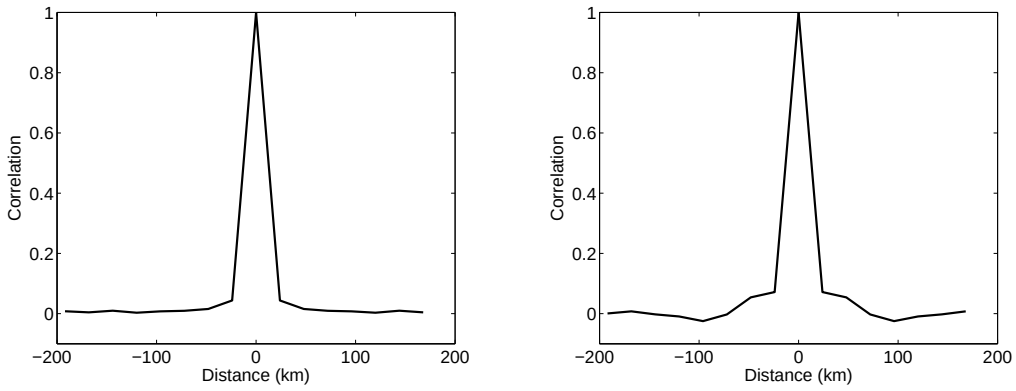
We now consider what happens where the observations are defined using a Gaussian-

weighting matrix. The results are given in Table 6.2 Experiment 1.3 and Table 6.3 Experiment 2.3. We plot the middle row of the representativity error correlation matrices for temperature and log specific humidity from Case 1 in Figures 6.3(a) and 6.3(b) (blue lines). We find that the error variance is smaller than when either direct or uniform observations are assumed. In this case Gaussian weighted observations are defined using a larger length scale and hence capture fewer small scale features than the direct and uniform observations. Therefore the representativity error is smaller as the model captures a larger proportion of the scales captured by the observations. From the figures we see that the correlations for the representativity error calculated with these Gaussian-weighted observations are larger than the representativity error correlations present when direct observations are used. This is due to the overlapping of the weighting functions in physical space of nearby observations. The overlapping weighting functions result in adjacent observations sharing information about particular points in state space. Hence the greater the overlap in weighting function the larger the correlations seen in the representativity error.

By comparing the experiments with different weighting functions we see that the larger the weighting function lengthscale used to define the observation the lower the representativity error variance. Observations defined using weighting functions with larger length-scales average over the smaller spatial spatial scales. Therefore the difference between a larger length-scale observation and the model representation of the observation is smaller than a small lengthscale observation and the model representation of the observation. Hence observations defined using weighting functions with larger lengthscales result in smaller representativity error variance.

6.3.3 Number of observations

We now consider what happens when we calculate the representativity error where fewer direct observations are available. As shown in the theoretical results in Chapter 4 and the numerical results using the KS equation in Chapter 5 we expect that the variance of the representativity error should not change. Experiments 1.4 in Table 6.2 and 2.4 in



(a) Temperature (K)

(b) log(Specific humidity) (kg/kg)

Figure 6.4: Representativity error correlations between observation centre points for Case 1 with truncation to 32 points (12km resolution) with every other model grid point observed using direct observations

Table 6.3 show the error variance where only half the model grid points have associated direct observations. By comparing with Experiment 1.1 we see, as expected, that having fewer observations available does not alter the variance of the representativity error. Experiments 1.5 and 2.5 with uniform observations and Experiments 1.6 and 2.6 with Gaussian observations also support this conclusion. We now consider how the structure of the representativity error correlations. We plot in Figures 6.4(a) and 6.4(a) the correlation structures of the temperature and specific humidity representativity errors. By comparing these figures with the direct observation results (black line) shown in Figures 6.3(a) and 6.3(a) we see that the representativity error correlation structure is not dependent on the number of observations, only the distance between them. This supports the numerical results seen in the previous chapter and the theoretical results presented in Chapter 4 Section 4.3.

6.3.4 Number of model grid points

We now consider the results when the model has a larger number of grid points, $N^m = 64$. This is a smaller truncation, so the model should be able to resolve more small scale features, and hence we expect the errors of representativity to decrease. We give the results for experiments with direct observations available every grid point in Table 6.2

Experiment 1.7 and Table 6.3 Experiment 2.7. In these experiments as well as increasing the number of model grid points, we have also increased the number of observations. However, as representativity error variance is not affected by the number of observations we assume that any differences in the representativity error variance can be attributed to the change in model resolution. From the results of Experiments 1.7 and 2.7 we see that the representativity error variances have decreased. For direct observations the representativity error has been approximately halved. Experiments 1.8 and 2.8 with uniform observations and Experiments 1.9 and 2.9 with Gaussian observations produce results that also support this conclusion.

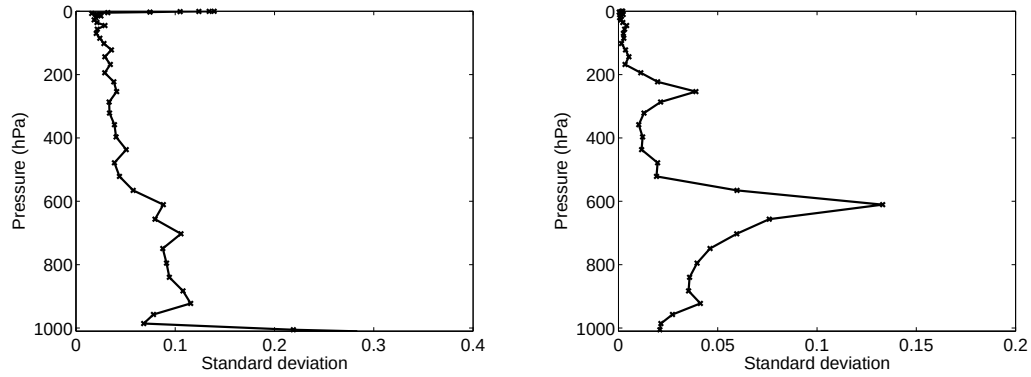
6.3.5 Representativity errors at different model levels

So far we have only considered the representativity error at the 749hPa model level height. We now calculate a representativity error for each pressure level of the model. This will allow us to consider the variation of representativity error with height. From this we can determine if one realisation of representativity error would be suitable at every pressure level, or if it is more appropriate to use the correct representativity error for each level.

Before calculating the representativity error for each model level, we must first calculate the covariance matrices for the high resolution data for temperature and specific humidity for each pressure level. We use the same data, but at the correct pressure level, and the same preprocessing techniques as described in section 6.1.

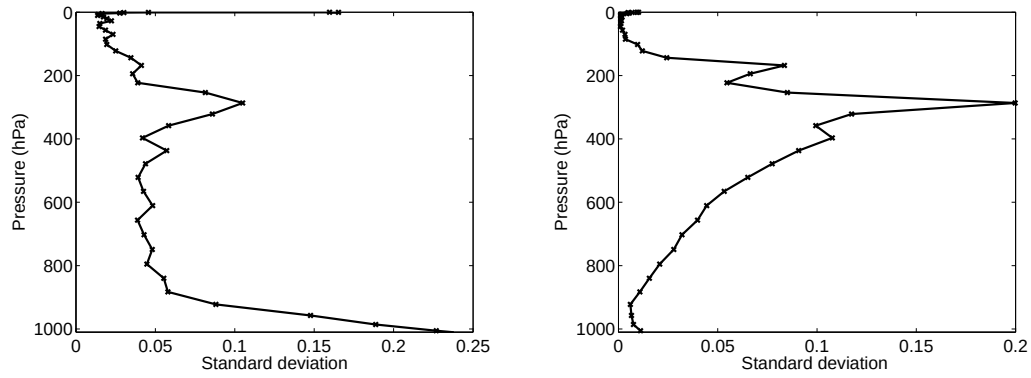
We consider the case where we have truncated to 32 grid points and have 32 direct observations available. We plot the standard deviation of representativity error for Case 1 in Figure 6.5(a) (temperature) and 6.5(b) (specific humidity) and for Case 2 in Figure 6.6(a) (temperature) and 6.6(b) (specific humidity).

From the figures we see that representativity error for temperature is more constant with height than specific humidity. The exception to this is in the boundary layer, where the temperature representativity error is large. For specific humidity in Case 1 we see a large



(a) Temperature (K) (b) log(Specific humidity) (kg/kg)

Figure 6.5: Change in representativity error standard deviation with model level height. Case 1 with 32 direct observations (every model grid point observed)



(a) Temperature (K) (b) log(Specific humidity) (kg/kg)

Figure 6.6: Change in representativity error standard deviation with model level height. Case 2 with 32 direct observations (every model grid point observed)

increase in the representativity error standard deviation between 749hPa and 610hPa. For Case 2 the largest peak in representativity error is seen at 300hPa. These levels are where cloud is seen and hence it is at these levels where the small scale humidity features exist, which results in the larger representativity error variances. Finally we consider how the correlation structure varies with height. We find that for both temperature and specific humidity at different pressure levels the correlation structures of the representativity errors are qualitatively similar to those for the 749hPa level, as seen in Figure 6.3. The difference in variance and minimal difference in correlation structure can be attributed to the different scales in the true state which are represented in the matrix, $\hat{\mathbf{S}}$, used to calculate the representativity error (Chapter 4, equation 4.9). The results support our conclusions that

representativity error is strongly case dependent.

6.4 Summary

In this chapter we use a method defined in Daley [1993] and Liu and Rabier [2002] to calculate representativity error. Previously the method has been used to investigate representativity error for a simple system. We adopt a new approach by applying the method to NWP data. We investigated the significance of representativity error for temperature and specific humidity. We showed that significant covariances in the observation error covariance matrix [Stewart et al., 2009, Stewart, 2010] could be attributed to representativity error. To calculate the representativity error using the Daley [1993] method it was necessary to have an estimate of the covariance of the high resolution data. This covariance was calculated using data from the Met Office UKV model. The accuracy of the representativity error estimates depends on the accuracy of these covariances and as the UKV model cannot represent all the scales in the truth it is possible that the representativity error is underestimated. Experiments using data from the Met Office UKV model showed that representativity error was more significant for humidity than temperature. We calculated representativity error using data from two different cases and showed that representativity error is sensitive to the synoptic situation, which supports claims by Janjic and Cohn [2006].

We showed that conclusions made in the previous chapter also hold when calculating representativity error for NWP data. We showed that representativity error decreases as observation lengthscale increases and model resolution increases. We also showed that the representativity error variance does not change when calculated with different numbers of observations. Also the representativity error correlations are dependent only on the distance between observations, and not the number of observations available. Finally we considered how the representativity error standard deviation varied at different pressure levels. We found that representativity does vary at different pressure levels and this means that assumptions such as those in Dee and Da Silva [1999] where errors at different model

levels are fixed, may not be suitable when representativity error is taken into account in data assimilation systems.

So far we have considered a method that provides a time independent estimate of representativity error. We have shown that representativity error is state and time dependent. This provides motive to develop a method to calculate time dependent estimate of the error. We next introduce and investigate a new method for estimating the representativity error.

Chapter 7

Calculating Observation Error Covariances Using the Ensemble Transform Kalman Filter

So far we have considered only time independent estimates of forward model and representativity error. In the previous chapter we showed that representativity error is case dependent, and therefore using a time independent estimate for forward model error for all cases may be inappropriate. Work by Li et al. [2009] used the Desroziers diagnostic embedded in a local ensemble transform Kalman filter to give a estimate of the observation error covariance matrix $\mathbf{R}$ under the assumption that $\mathbf{R}$ is diagonal and that the true observation error covariance matrix is static. At each analysis step the Desroziers diagnostic is applied to a subset of observations to give a value for the observation error variance. This work was extended in Miyoshi et al. [2013] to include a correlated matrix $\mathbf{R}$. In the framework described by Li et al. [2009] and Miyoshi et al. [2013] it is possible to average over a subset of observations as all observations have the same variance. However as forward model error is time and state dependent averaging over observations may be a poor assumption. In this chapter we introduce a new method, similar to that seen in Miyoshi et al. [2013], that combines an ensemble filter with the Desroziers diagnostic. Rather than averaging over

a set of observations at a given time, our method uses statistics from observations over a short period of time to produce a slowly time varying estimate of the observation error covariance matrix $\mathbf{R}$. Subtracting the known instrument error from this estimate gives a time varying estimate for forward model error. After presenting the new method we carry out a number of experiments. We first show that it is possible to estimate the observation error covariance using the Desroziers diagnostic. We then show how we can estimate a slowly varying time estimate of $\mathbf{R}$. Finally we estimate $\mathbf{R}$ and substitute this estimated error covariance matrix back in to the assimilation scheme. We show that this will improve both the estimate of $\mathbf{R}$ and the analysis. We note that the method and work described in this chapter were developed before the work of Miyoshi et al. [2013] appeared.

7.1 Including observation error estimation in the ensemble transform Kalman filter

In Chapter 2 we introduced ensemble Kalman filtering and described in detail the ensemble transform Kalman filter (Chapter 2 Table 2.3). We use this as the base of our method introduced here; however any deterministic ensemble Kalman filter should be suitable. The idea is to estimate the observation error covariance matrix within ETKF. We use the ETKF to provide the samples of the background and analysis innovations to be used in the Desroziers diagnostic that was introduced in Chapter 3 Section 3.3.2. The filter is split into two stages, a static $\mathbf{R}$ stage and a 'varying estimate' stage. In the initialisation stage an initial set of samples for use with the Desroziers diagnostics are calculated. In the second stage at each assimilation step the samples for use with the Desroziers diagnostic are updated and a new estimate of $\mathbf{R}$ is calculated. This estimate of the observation error covariance matrix is substituted back into the assimilation scheme. This improves further estimates of $\mathbf{R}$ calculated within the scheme, and improves the analysis. We now describe the method we have developed.

The method is presented in Table 7.1. We start by describing the initialisation phase. We begin with an initial ensemble, at $t = 0$, that has an associated initial covariance matrix.

We also assume an initial estimate of the observation error covariance matrix $\mathbf{R}_0$; it is possible that this could just consist of the instrument error. The first step, Step 1 of Table 7.1, is to use the full non-linear model to forecast each ensemble member. Then the ensemble mean and covariances are calculated using equations (2.13) and (2.16). Using the ensemble mean the background innovations $\mathbf{d}_n^b = \mathbf{y}_n - \mathcal{H}\bar{\mathbf{x}}_n^f$ at time t_n are calculated in Step 4 of Table 7.1. We then carry out the update steps. We update the ensemble mean using equation (2.20), Step 5 of Table 7.1, and the ensemble perturbations using equation (2.22), Step 6 of Table 7.1. The analysis mean is then used to calculate the analysis innovations, Step 7 of Table 7.1. These steps, an application of the standard ETKF, are repeated for a number of assimilation steps N^s. This number of assimilation steps is dependent on the number of samples required to calculate an accurate estimate of the observation error covariance using the Desroziers diagnostic, and further investigation is required to determine this number. Once these assimilation steps are completed a new time averaged estimate of $\mathbf{R}$ is calculated using the Desroziers diagnostic from equation (3.17), $\mathbf{R} = E[\mathbf{d}^a \mathbf{d}^{bT}]$. We do this using the samples of the innovations collected using,

$$\mathbf{R}_{N^s+1} = \frac{1}{N^s - 1} \sum_{i=1}^{i=N^s} \mathbf{d}_i^a \mathbf{d}_i^{bT}. \tag{7.1}$$

Now we have a set of samples we can begin to include and update the estimate of $\mathbf{R}$. We continue running the ETKF using the updated $\mathbf{R}_n$ in place of our initial guess for $\mathbf{R}$. After the forecast and analysis stages we calculate a new estimate for the observation error covariance matrix $\mathbf{R}$ by removing the oldest samples for $\mathbf{d}^b$ and $\mathbf{d}^a$ and replacing them with those calculated in the current assimilation step, that is,

$$\mathbf{R}_{n+1} = \frac{1}{N^s - 1} \sum_{i=n-N^s+1}^{i=n} \mathbf{d}_i^a \mathbf{d}_i^{bT}. \tag{7.2}$$

This means that at every assimilation step $\mathbf{R}$ is updated using the latest information, with the oldest information being discarded. Although this does not give a completely time dependent estimate of $\mathbf{R}$ it should give a slowly time varying estimate that should take into account the most recent information relating to the true state. We summarize the

The Determanistic Ensemble Kalman Filter Algorithm with $\mathbf{R}$ estimation

Initilisation

1. Determine the initial ensemble $\mathbf{x}_0^i$ for $i = 1 \dots N$.
2. Set $\mathbf{R}_0$ to an initial guess $\mathbf{R}^{int}$.

Itterations

1. forecast each ensemble member,
$$\mathbf{x}_n^f, i = \mathbf{M}_n \mathbf{x}_{n-1}^{a,i}.$$
2. Determine the ensemble mean,
$$\bar{\mathbf{x}}_n^f = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_n^{f,i}$$
3. Determine the ensemble perturbation matrix
$$\mathbf{X'}_n^f = (\mathbf{x}_n^f, 1 - \bar{\mathbf{x}}_n^f, \dots, \mathbf{x}_n^f, N - \bar{\mathbf{x}}_n^f)$$
4. Calculate the background innovations,
$$\mathbf{d}_n^f = \mathbf{y}_n - \mathcal{H}(\bar{\mathbf{x}}_n^f)$$
5. Update the ensemble mean,
$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^f + \mathbf{K}_n \mathbf{d}_n^f,$$
where $\mathbf{K}_n = \mathbf{X'}_n^f (\mathbf{Y'}_n^f)^T (\mathbf{Y'}_n^f (\mathbf{Y'}_n^f)^T + \mathbf{R}_n)^{-1}$
6. Update the ensemble perturbations,
$$\mathbf{X'}_n^a = \mathbf{X'}_n^f (\mathbf{I} - \mathbf{Y'}_n^{f\,T} \mathbf{S}^{-1} \mathbf{Y'}_n^f)^{-\frac{1}{2}}$$
7. Calculate the analysis innovations,
$$\mathbf{d}_n^a = \mathbf{y}_n - \mathcal{H}(\bar{\mathbf{x}}_n^a)$$
8. If $n > N^s$ update $\mathbf{R}$ using
$$\mathbf{R}_{n+1} = \frac{1}{N^s - 1} \sum_{a=n-N^s}^{a=n} \mathbf{d}_n^a \mathbf{d}_n^{b\,T}.$$
Otherwise keep $\mathbf{R}_{n+1} = \mathbf{R}^{int}$

Table 7.1: An algorithm for the EnKF with $\mathbf{R}$ estimation

algorithm in Table 7.1.

There is a number of limitations with the method, one of which is that a large number of samples may be required to calculate the observation error covariance matrix. The method does allow any number of samples to be generated before the $\mathbf{R}$ estimation begins. However using a larger number of samples means the estimate of the observation error

covariance matrix is an average over a large period of time. Therefore the larger the number of samples the less time dependent the estimate becomes. Even if the required number of samples is small the time dependence may be reduced if the observations are sparse. Observations that are only available at large time intervals may represent very different synoptic situations and therefore an average over these innovations will be less related to the current synoptic situation. This suggests that this method for estimating $\mathbf{R}$ would be most suited to a situation where observations are available frequently.

We now describe the experiments we use to test this algorithm.

7.2 Experiment design

To analyse the method we run a series of twin experiments. We use the Kuramoto-Sivashinsky equation, as described in Chapter 5 Section 5.1, as our model. We use as our truth the solution to the KS equation on the periodic domain $0 \leq x \leq 32\pi$ from initial conditions $u = \cos(\frac{x}{16})(1 + \sin(\frac{x}{16}))$ until time $T = 10000$, using $N = 256$ spatial points and a time step of $\Delta t = 0.25$. To minimise model error and representativity error, we run our model at the same spatial and temporal resolution as the truth. We use a slightly perturbed initial condition, created by adding an error from the distribution $\mathcal{N}(0, 0.1)$ to the true initial condition. From this initial condition the $N = 1000$ ensemble members are created by adding errors from the initial background error distribution, which is chosen to also be $\mathcal{N}(0, 0.1)$. Hence the background error covariance matrix is $\mathbf{B} = 0.1\mathbf{I}$. We choose the large number of ensemble members to minimise the risk of ensemble collapse and to help obtain an accurate background error covariance matrix as we wish to avoid using covariance inflation and localisation. We require the background error covariance matrix to be as accurate as possible so the Desroziers diagnostic produces the best estimate of $\mathbf{R}$. We choose to use direct observations with added instrument error, which are calculated by adding error from $\mathcal{N}(0, 0.1)$ to the values of the truth. As we have removed the source of forward model error, as the model has the same resolution as the truth, we must artificially add this correlated error to our observations. As the correlation function for our artificial

representativity error we use the SOAR function where ρ is the correlation between two points separated by distance r,

$$\rho_t(r) = \left\{ \cos(br) + \frac{\sin(br)}{L_t b} \right\} e^{\frac{-r}{L_t}}, \tag{7.3}$$

where we set constants $b = 3.8$ and $L_t = 15$ and $0 \leq r \leq a\pi$. We use this SOAR function to determine a circulant covariance matrix. Each row of the matrix contains the SOAR function shifted by one element. To calculate our true representativity error covariance matrix that we aim to estimate, we multiply the circulant matrix by the representativity error variance which is chosen to be 0.1. The true observation error covariance matrix is obtained by adding the instrument and reperesentativity error covariance matrices. The result is plotted in Figure 7.1. Having a specified observation error covariance matrix allows

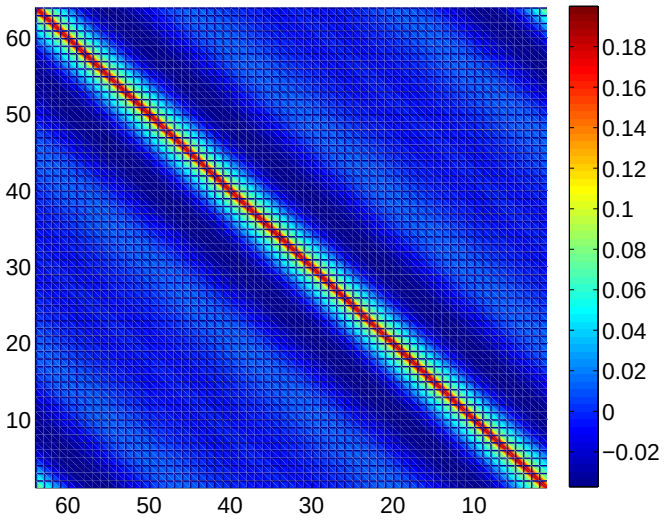


Figure 7.1: True observation error covariance matrix

us to determine how well the method is working as we know the observation error covariance matrix that we are trying to approximate. 64 equally spaced observations, drawn from the distribution $\mathcal{N}(0, \mathbf{R}^t)$, are available at each assimilation step and the frequency varies between experiments, with the chosen frequencies being observations available every 40 and 100 time steps, that is every 10 and 25 time units respectively.

As well as changing the frequency of the observations we also consider different true $\mathbf{R}$

matrices. For each different $\mathbf{R}$ and observation frequency we run four different types of experiment.

- Type A: Filter run without estimating the observation error covariance matrix $\mathbf{R}$. The $\mathbf{R}$ used in the assimilation is the true observation error covariance matrix $\mathbf{R}^{int} = \mathbf{R}^{t}$. The estimated observation error covariance matrix $\mathbf{R}^{est}$ is calculated after the final assimilation using all the available background and analysis innovations.

- Type B: Filter run without estimating the observation error covariance matrix $\mathbf{R}$. The $\mathbf{R}$ used in the assimilation consists of the only the instrument error covariance matrix $\mathbf{R}^{int} = \mathbf{R}^{I}$. The estimated observation error covariance matrix $\mathbf{R}^{est}$ is calculated after the final assimilation using all the available background and analysis innovations.

- Type C: Filter run without estimating the observation error covariance matrix $\mathbf{R}$. The $\mathbf{R}$ used in the assimilation consists of the the instrument error covariance matrix inflated by a scaler factor α so $\mathbf{R}^{int} = \alpha\mathbf{R}^{I}$. The inflation factor α is chosen so the variance of the observation error covariance matrix used in the assimilation is equal to the true observation error variance. The estimated observation error covariance matrix $\mathbf{R}^{est}$ is calculated after the final assimilation using all the available background and analysis innovations.

- Type D: Filter run with observation error covariance matrix $\mathbf{R}$ estimation. The initial $\mathbf{R}$ used in the assimilation consists of the only the instrument error covariance matrix $\mathbf{R}^{int} = \mathbf{R}^{I}$. The number of samples used to calculate $\mathbf{R}^{est}$ is $N^{s} = 250$. The observation error covariance matrix is first estimated after 250 assimilation steps, with this estimate being included in the next assimilation step. The observation error covariance matrix is then updated every assimilation step using step 8 of the algorithm given in Table 7.1.

Running this set of experiments allows us to determine how well the filter is working. It also allows us to see if inflating the variance accounts for unknown forward model error. We also consider if the matrix $\mathbf{R}$, and hence forward model error, can be estimated within

an assimilation scheme. We now present the results of our experiments.

7.3 Results

We present the results from all our experiments in Table 7.2. We give details of the matrix used as the true observation error covariance matrix $\mathbf{R}^t$. We also give the type of experiment (A-D) used and the frequency of the observations. We also give two time averaged analysis RMSEs. The two RMSEs are calculated using different error realisations for the background and instrument errors as well as the perturbations for the initial conditions and the initial creation of the ensemble members. The RMSE allows us to compare the performance of the filter for each experiment. We also provide figures to aid our analysis of the experiments. We plot rank histograms to give information about the ensemble spread as consequently this may affect the analysis and the estimation of the observation error covariance matrix. We also plot a row of the true and estimated observation error covariance matrix. To give an idea of the accuracy of the estimated covariance we also include, in the figure caption, the RMSE of one row of the estimated observation error covariance matrix. The RMSE of the row of the estimated observation error covariance matrix is calculated using equation (2.26). The truth is one row of the true observation error covariance matrix and as our estimated covariance we use the average calculated covariance structure. The average covariance structure is calculated by averaging the permuted rows of the estimated observation error matrix.

We begin by describing the experiments where the true matrix $\mathbf{R}$ is static and the observations are available frequently.

7.3.1 Results with a static R and frequent observations

We begin by setting the true matrix $\mathbf{R}^t$, to the matrix shown in Figure 7.1. We run the the model until a final time of $T = 10000$, a total of 40000 time steps, and observations are available every 40 time steps. This allows us to run the filter for 1000 assimilation steps.

Experiment Number	True R	Experiment Type	Obs Freq (time steps)	Background, Instrument and representativity error variance	Time Av analysis RMSE Error realisation 1	Time Av analysis RMSE Error realisation 2
1	SOAR + RI	A	40	0.1	0.246	0.250
2	SOAR + RI	B	40	0.1	0.275	0.276
3	SOAR + RI	C	40	0.1	0.273	0.276
4	SOAR + RI	D	40	0.1	0.251	0.260
5	SOAR + RI	A	100	0.1	0.353	0.350
6	SOAR + RI	B	100	0.1	0.380	0.381
7	SOAR + RI	C	100	0.1	0.375	0.373
8	SOAR + RI	D	100	0.1	0.357	0.372
9	SOAR + RI/RI	A	40	0.1	0.270	0.269
10	SOAR + RI/RI	B	40	0.1	0.284	0.284
11	SOAR + RI/RI	C	40	0.1	0.281	0.284
12	SOAR + RI/RI	D	40	0.1	0.279	0.281
13	Time dependent (L_o 3.7 to 4.0)	D	40	0.1	0.255	0.248
14	Time dependent (L_o 4.0 to 3.7)	D	40	0.1	0.252	0.249
15	Time dependent (L_o 3.7 to 4.0)	D	40	0.01	0.060	0.058
16	Time dependent (L_o 3.7 to 4.0)	D	40	1.0	0.704	0.704

Table 7.2: Details of experiments executed to investigate the performance of the ETKF with observation error covariance estimation

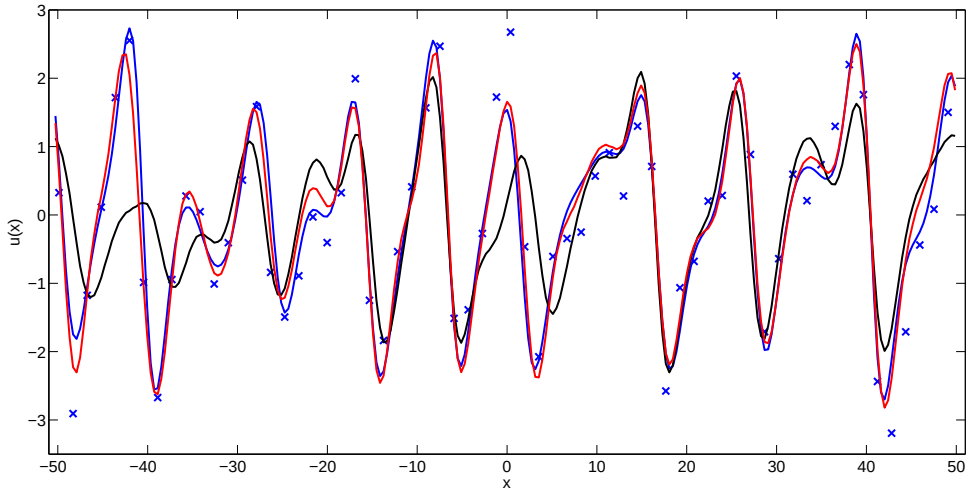


Figure 7.2: Experiment 1: Truth (blue), observations (crosses), forecast (black) and analysis (red) for the final time $T = 10000$

Experiment Type A

We begin by running the filter without estimating $\mathbf{R}$. At each assimilation time we calculate the background and analysis innovations. At the end of the assimilation window we use the 1000 background and analysis innovation statistics to calculate the matrix $\mathbf{R}$. This allows us to check that the filter is working correctly and that the observation error covariance matrix can be calculated using the Desrozier diagnostic.

We first run the filter using the correct $\mathbf{R}$. We calculate the innovation statistics at each assimilation step. However we only use them to estimate $\mathbf{R}$ after the final assimilation has been completed. We run the assimilation using the correct $\mathbf{R}$ throughout as this allows us to show that the assimilation and Desroziers diagnostic perform well. As we are using the correct $\mathbf{R}$ this shows the best performance we can expect from the filter. We show the final time truth, observations, ensemble mean forecast and analysis state in Figure 7.2. The rank histogram is plotted in Figure 7.3 and the time averaged analysis RMSE for Experiment 1 is given in Table 7.2. We expect the RMSE to be the lowest of all the experiments as $\mathbf{R}$ is specified correctly and the observations are frequent. We see that the assimilation is working well with the analysis being a better approximation of the truth than the background, particularly in the first half of the domain. We see that the rank

histogram is fairly flat indicating that the ensemble spread is sufficient to capture the observations and the filter is not divergent. As well as considering the performance of the filter we also consider how well the Desrozier diagnostic performs in this case. As we are using the correct $\mathbf{R}$ and assuming we have a good estimate for $\mathbf{P}^f$ we expect the Desroziers diagnostic to give a good estimate of the matrix $\mathbf{R}$. We use the innovations from each of the 1000 assimilation steps to calculate the observation error covariance matrix. As the true observation error covariance matrix is isotropic and homogeneous we plot the correlation function obtained by averaging the rows of the covariance matrix, which removes some of the sampling error. It may be possible to reduce this sampling error by increasing the number of samples. However where we introduce a time varying $\mathbf{R}$, increasing the number of samples reduces the time dependence of the matrix $\mathbf{R}$ we are calculating, as we will use samples that span a larger amount of time. It may also be possible to overcome the sampling error with localisation; however we do not consider this here. The middle row of the true correlation matrix (blue), and the average of the calculated covariance matrix (red) are plotted in Figure 7.4. We see that the Desroziers diagnostic is working well, providing a good estimate of the matrix $\mathbf{R}$.

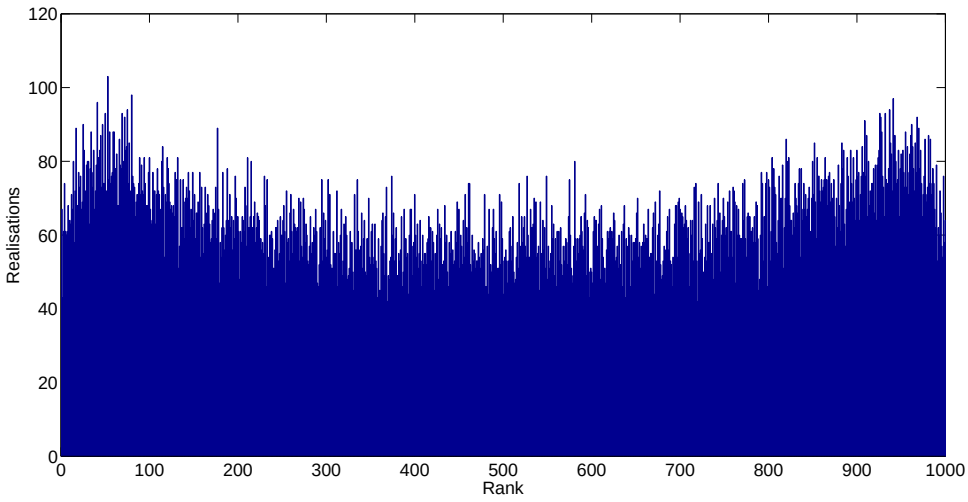


Figure 7.3: Rank Histogram for Experiment 1 (experiment type A with frequent observations)

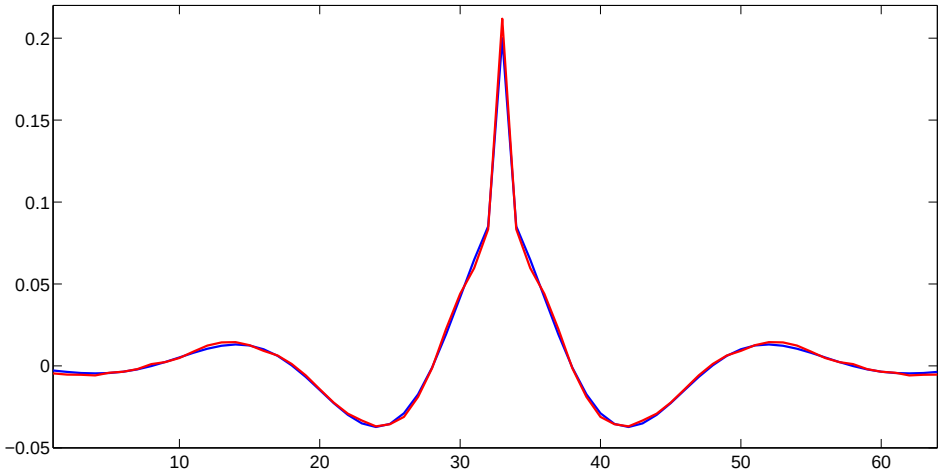
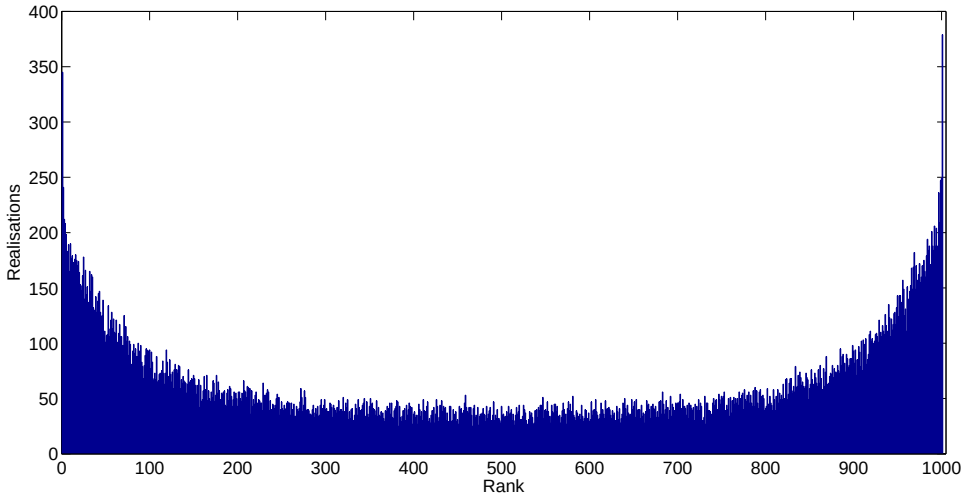


Figure 7.4: Rows of the true (blue) and estimated (red) covariance matrices for Experiment 1 (experiment type A with frequent observations). Observation error covariance RMSE 0.002

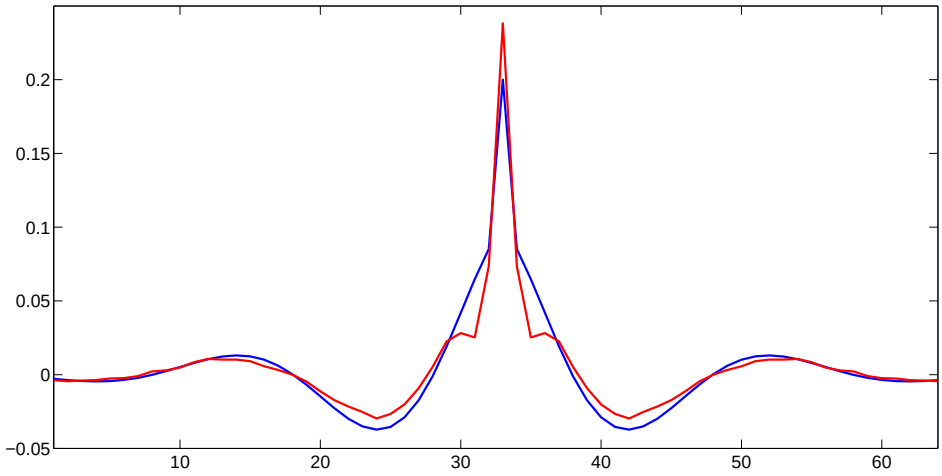
Experiment Type B

We next consider the case where it is assumed that the observation error covariance matrix used in the assimilation consists only of the instrument error covariance, that is $\mathbf{R} = \mathbf{R}^I$. All 1000 analysis and background innovations are used as samples for the Desroziers diagnostic. The analysis RMSE is given in Table 7.2 Experiment 2, while the true and calculated matrices are plotted in Figure 7.5(b). The rank histogram is plotted in Figure 7.5(a).

We see that the rank histogram has a definite U shape suggesting that there is not enough variability in the ensemble members. This suggests that the variance in the matrix $\mathbf{P}^f$ is not as large as it should be, which may affect the estimation of $\mathbf{R}$. As expected the assimilation does not perform as well as in Experiment 1 and the time averaged RMSE is larger. From Figure 7.5(b) we see that even where it is assumed that $\mathbf{R} = \mathbf{R}^I$ the Desroziers diagnostic still gives a reasonable estimate of the true covariance with approximately correct length scales. This suggests that even if forward model error is initially unknown it should be possible to use the Desroziers diagnostic to produce an estimate of forward model error.



(a) Rank Histogram for Experiment 2

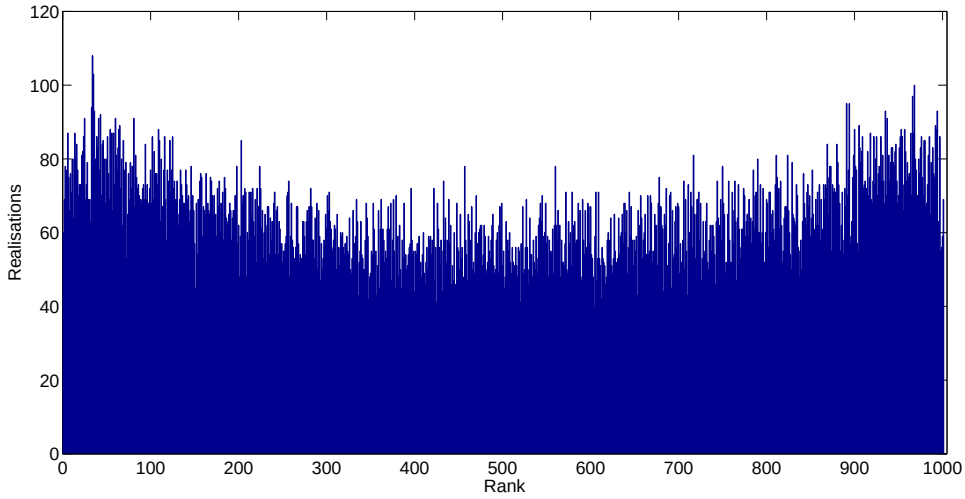


(b) Rows of the true (blue) and estimated (red) covariance matrices for Experiment 2. Observation error covariance RMSE 0.010

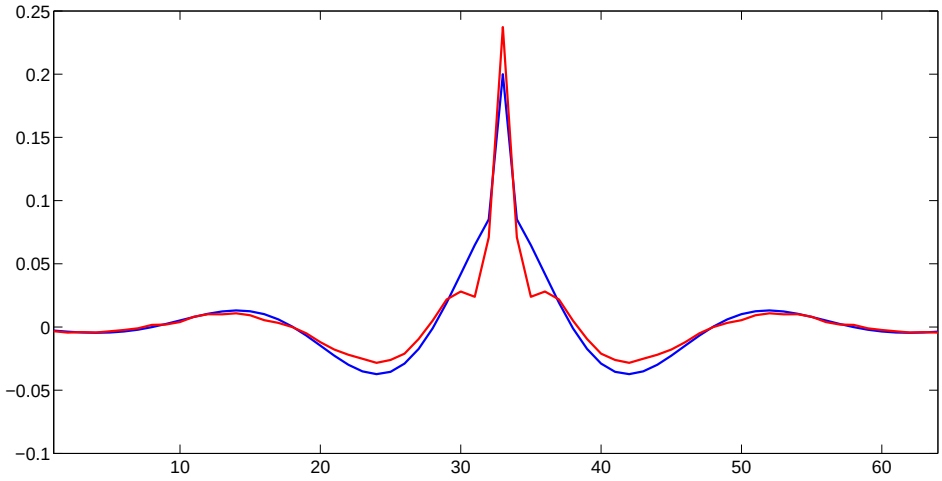
Figure 7.5: Diagnostics for Experiment 2 (experiment type B with frequent observations)

Experiment Type C

In Experiment 3 we use variance inflation to increase the variance of the instrument error covariance. We choose the inflation factor to be 2, so that $\mathbf{R} = 2\mathbf{R}^I$, which gives a diagonal covariance matrix with the same variance as the true matrix $\mathbf{R}$. We see from the rank histogram in Figure 7.6(a) by comparison with Figure 7.5(a) that inflating the assumed observation error variance has increased the ensemble spread.



(a) Rank Histogram for Experiment 3



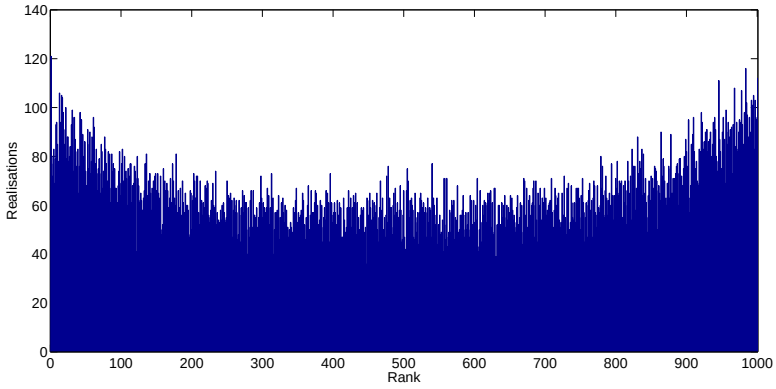
(b) Rows of the true (blue) and estimated (red) covariance matrices for Experiment 3. Observation error covariance RMSE 0.010

Figure 7.6: Diagnostics for Experiment 3 (experiment type C with frequent observations)

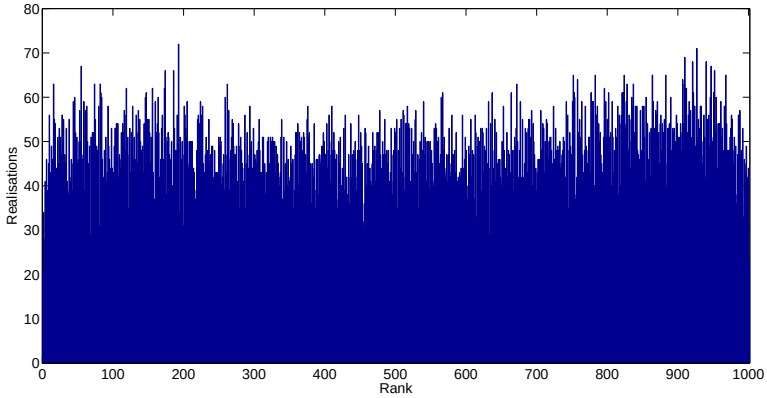
Considering the RMSE shows that the assimilation is comparable to Experiment 2. Again all 1000 background and analysis innovations are used to calculate our estimate of $\mathbf{R}$. We see from Figures 7.6(b) that using the Desroziers diagnostic has provided a reasonable estimate of $\mathbf{R}$ and it is similar to the estimate from Experiment 2.

Experiment Type D

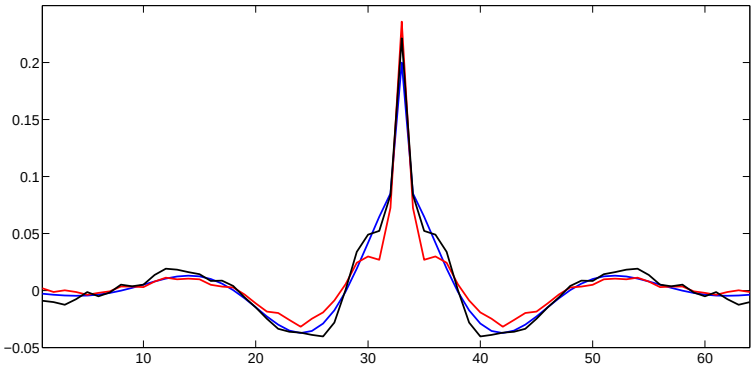
We now consider what happens where we estimate $\mathbf{R}$ within the assimilation scheme as described in Table 7.1 with $N^s = 250$. We find that with $N^s = 250$ samples, for the estimated matrix to be full rank it is necessary to remove the sampling error. To remove the sampling error we make the matrix isotropic and homogeneous by taking the mean of the shifted rows of estimated matrix. This averaged row is then used to reconstruct a circulant matrix. This makes the assumption that all the observations have the same correlations. Although this may be the case in this experiment, it is not necessarily true in a NWP scenario. As previously mentioned it may be possible to overcome sampling error by increasing the number of samples; however this reduces the time dependence of the estimation. We choose to make the homogeneous assumption to increase the time dependence of our estimated $\mathbf{R}$. We verify that the method proposed is able to improve the analysis by including improved estimates of $\mathbf{R}$ in the assimilation scheme. We begin by assuming that the observation error covariance matrix consists of only the instrument error. We plot in Figure 7.7(a) the rank histogram for Experiment 4. We see that that the rank histogram is U shaped, but the shape is not as severe as in Experiment 2, which suggests that there is a lack of variability in the ensembles. As the first 250 time steps are equivalent to Experiment 2, it is possible that it is these initial assimilation steps that contribute to the shape of the rank histogram, and that once the estimated $\mathbf{R}$ is used in the assimilation the ensemble spread is increased. To see if this is the case we plot the rank histogram, Figure 7.7(b), for the last 750 assimilation steps. We see that the rank histogram for the section of the assimilation that uses the estimated $\mathbf{R}$ is flat. This suggests that using the correct $\mathbf{R}$ has helped increase the spread of the ensemble. This shows that overall the assimilation scheme is performing better than the cases where the observation error covariance matrix was assumed diagonal. The RMSE is now closer to the RMSE obtained where the correct $\mathbf{R}$ was used. In Figure 7.7(c) we plot the true covariance (blue) as well as the first estimate of the covariance calculated using the first 250 background and analysis innovations (red) and the last estimate of the covariance calculated using the last 250 background and analysis innovations (black). We note that



(a) Rank Histogram for Experiment 4



(b) Rank Histogram for the last 750 assimilation steps using estimated $\mathbf{R}$ in Experiment 4



(c) Rows of the true (blue) and estimated covariance matrices for Experiment 4. Covariance calculated using the first 250 background and analysis innovations (red), observation error covariance RMSE 0.010. Covariance calculated using the last 250 background and analysis innovations (black), observation error covariance RMSE 0.006.

Figure 7.7: Diagnostics for Experiment 4 (experiment type D with frequent observations)

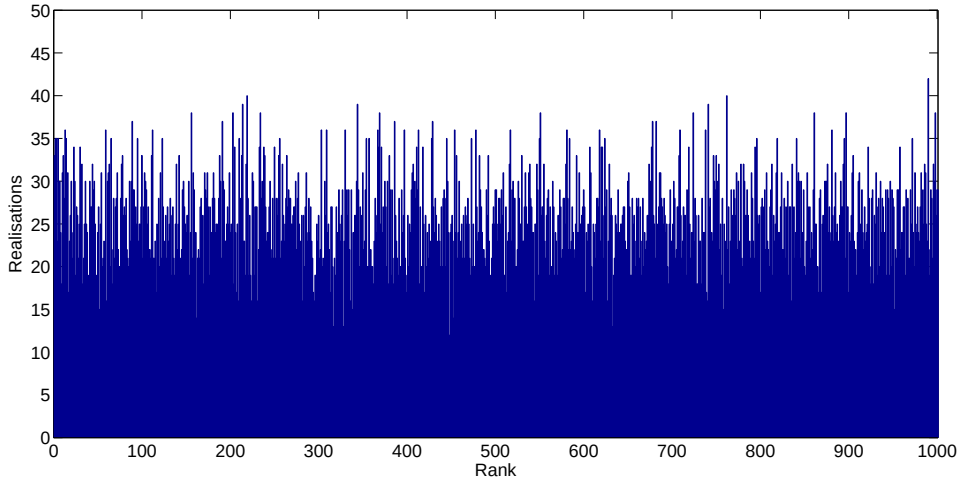
as $\mathbf{R}$ is estimated using fewer samples than the previous experiments we expect the results to be more noisy due to the increased sampling error. We see that the first estimate of the covariance structure is similar to that calculated in Experiment 2. This is because all the innovations used as samples were calculated assuming that $\mathbf{R} = \mathbf{R}^I$. We see that the last estimate of the covariance structure is closer to the true covariance structure. This suggests that iterating the estimation of $\mathbf{R}$ within the ETKF improves the estimation of a static $\mathbf{R}$. It also suggests that it should be possible to gain a time dependent estimate of forward model error. So far we have considered the case where observations are available every 40 time steps. However it is possible that the ETKF with $\mathbf{R}$ estimation is sensitive to the time interval between observations. To test this we now consider the case where observations are only available every 100 time steps.

7.3.2 Static R, infrequent observations

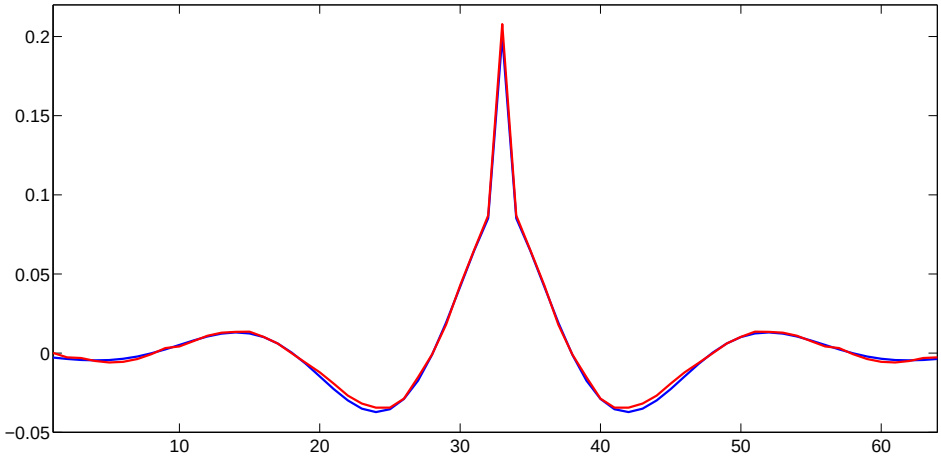
Again we start by only estimating $\mathbf{R}$ after the final assimilation has finished. Observations are available only every 100 time steps so we have 400 assimilation steps, and hence 400 background and analysis innovations that can be used to estimate $\mathbf{R}$. We consider the cases where $\mathbf{R}$ is correctly specified, set to be the instrument error, and an inflated instrument error. The RMSE for these experiments are given in Table 7.2 Experiments 5, 6 and 7.

Experiment Type A

We begin by using the correct observation error covariance matrix in the assimilation. As we expect, the lowest RMSE is for Experiment 5. This is because we are using the correct observation error covariance matrix $\mathbf{R}^t$ in the assimilation scheme. The RMSE is larger than in Experiment 1 because there is a larger time between observations so there is more time for the model to move away from the observations and therefore the assimilation has a larger correction to make. However the flat rank histogram seen in Figure 7.8(a) suggests that there is enough variance in the ensemble. The estimation of the covariance structure,



(a) Rank Histogram for Experiment 5



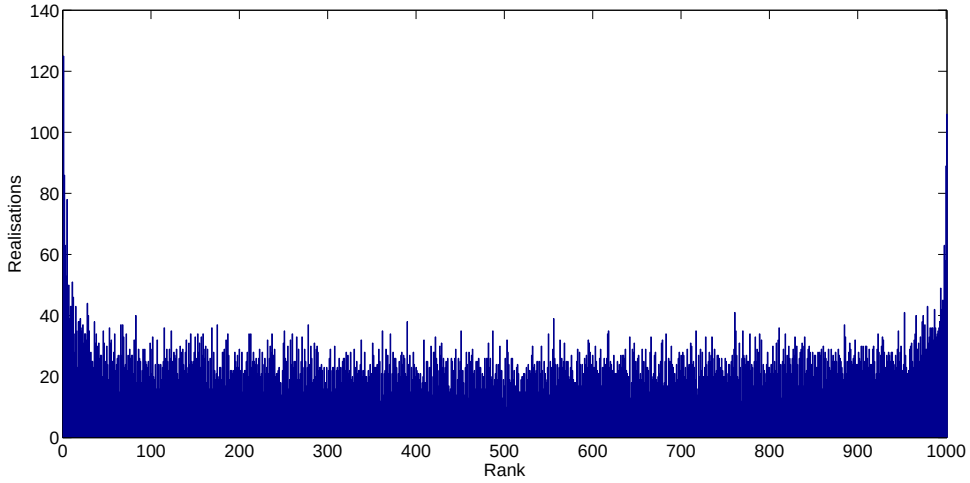
(b) Rows of the true (blue) and estimated (red) covariance matrices for Experiment 5. Observation error covariance RMSE 0.002

Figure 7.8: Diagnostics for Experiment 5 (experiment type A with infrequent observations)

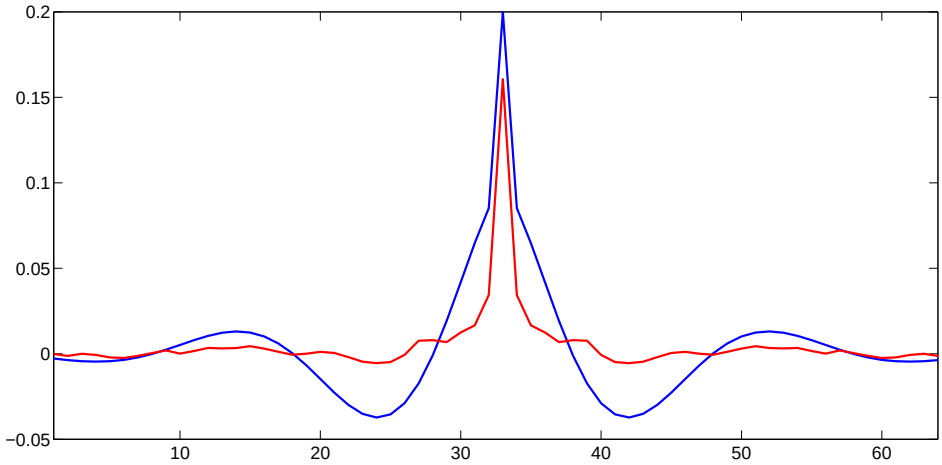
plotted in Figure 7.8(b), is again accurate.

Experiment Type B

Where $\mathbf{R} = \mathbf{R}^I$ we see from Figure 7.9(a) that the ensemble is under determined. From Table 7.2 we see that the RMSE larger than for Experiment 5. From Figure 7.9(b) it is obvious that the estimate of the covariance function is not as good as Experiment 2, and this suggests that the larger spacing between observations does affect how well the



(a) Rank Histogram for Experiment 6



(b) Rows of the true (blue) and estimated (red) covariance matrices for Experiment 6. Observation error covariance RMSE 0.020

Figure 7.9: Diagnostics for Experiment 6 (experiment type B with infrequent observations)

Desrozier diagnostic estimates $\mathbf{R}$ if the $\mathbf{R}$ used for calculating the innovations is incorrect. We also see that in this case the estimate of the variance is an underestimate of the true variance, whereas where the observations are more frequent (Experiment 2) the estimated variance is an overestimate. As well as being linked to the frequency of the observations, it is likely that the estimation of $\mathbf{R}$ is also related to the matrix $\mathbf{P}^f$. With fewer observations the ensemble has more time to spread; this in turn may lead to a matrix $\mathbf{P}^f$ with larger variances, whereas more frequent observations lead to a $\mathbf{P}^f$ with smaller variances. The

Desroziers diagnostic uses the matrix $\mathbf{P}^f$ in the calculation of the observation covariance matrix. It is possible that with less frequent observations the large variances in $\mathbf{P}^f$ are dominant in the calculation of the observation error covariance matrix, and hence the observation error variances are underestimated.

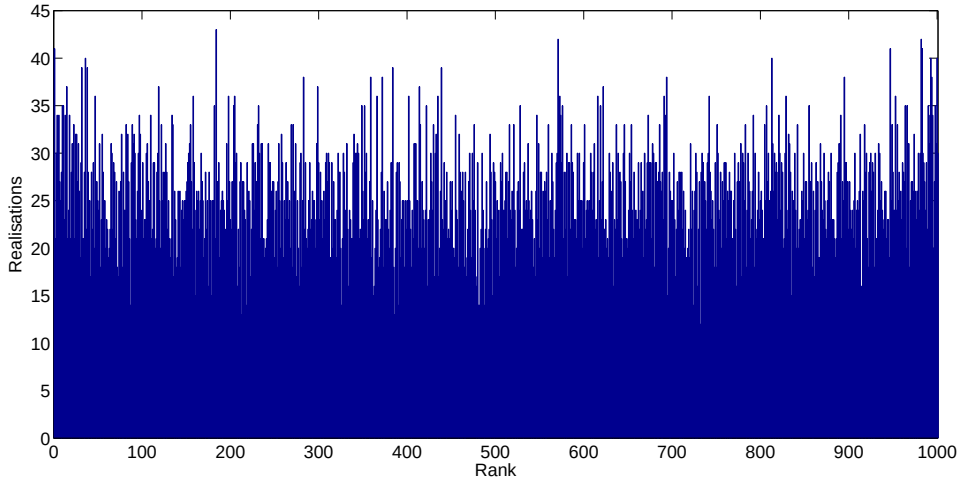
Experiment Type C

Assuming an inflated error variance, $\mathbf{R} = 2\mathbf{R}^I$, again helps keep the ensemble spread (Figure 7.10(a)) and the RMSE (Experiment 7, Table 7.2) is also slightly lower than Experiment 6. The variance estimation of $\mathbf{R}$ is improved from Experiment 6. It is likely that the improvement is because the assumed observation error variance is equal to the true error variance. However we show in Figure 7.10(b) that the covariance structure is not well estimated.

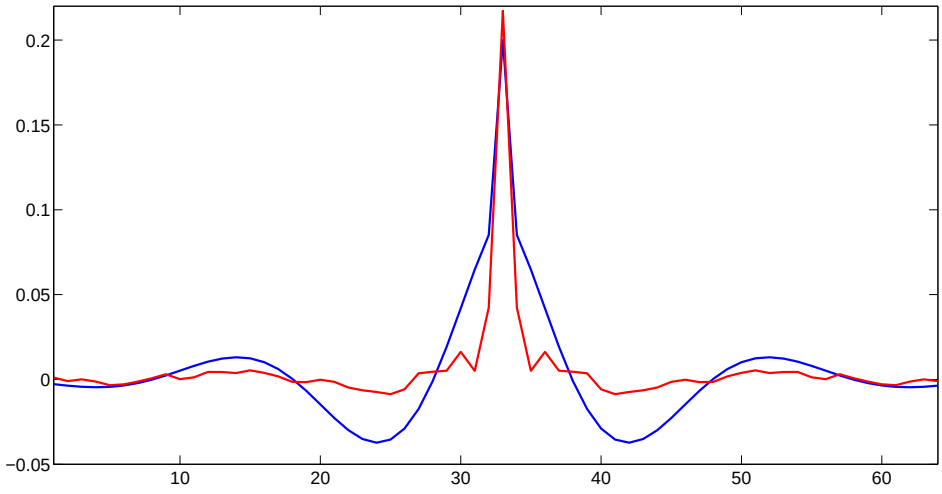
Experiment Type D

We now estimate $\mathbf{R}$ within the scheme and then reuse the estimated $\mathbf{R}$ at the next assimilation step. As our initial error covariance we choose the instrument error covariance, $\mathbf{R} = \mathbf{R}^I$. We see from Table 7.2 that the RMSE is lower than both the cases where $\mathbf{R}$ was assumed diagonal, and is close to the RMSE where the correct $\mathbf{R}$ was used. The rank histogram plotted in Figure 7.11(a) again suggests that the ensemble is slightly under determined. Again this may be due to the initial 250 assimilation steps where $\mathbf{R}$ is assumed diagonal. As with Experiment 4 we plot in Figure 7.11(b) the true covariance structure as well as the first and last estimated correlation structures. The first estimate is the equivalent of calculating $\mathbf{R}$ after the first 250 assimilation steps of Experiment 6. It is for this reason that we see an under estimate of the correlation structure. We see that the last estimate of the correlation variance is closer to the truth. The structure is also improved, but does not closely follow the truth. This is partly because some of the background and analysis innovations were calculated using the diagonal $\mathbf{R}$. As we are considering a static observation error matrix we expect the estimated $\mathbf{R}$ to improve with every assimilation

step. If the assimilation is run for longer period of time we would expect the estimated $\mathbf{R}$ to converge to the truth. However if the observations are not frequent enough this method may not be suitable for producing a time varying estimate of $\mathbf{R}$. We now return to the the case of more frequent observations, but consider a different true $\mathbf{R}$.

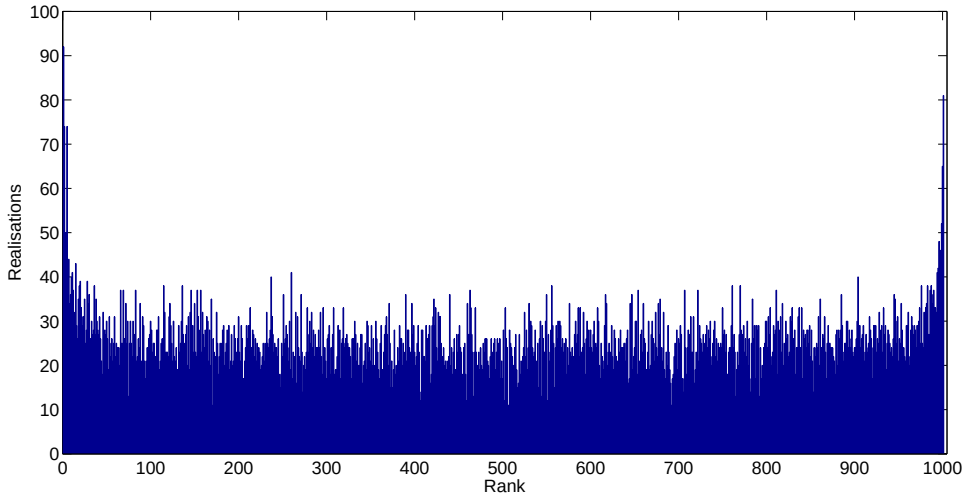


(a) Rank Histogram for Experiment 7

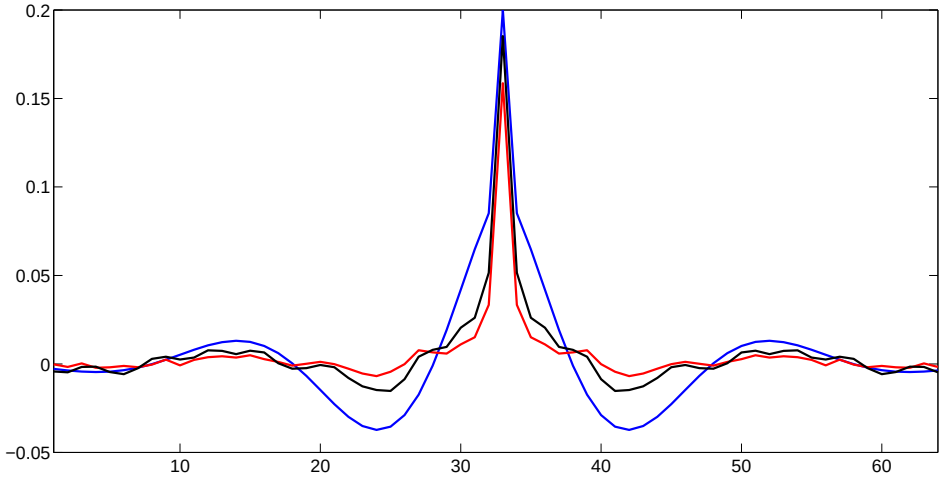


(b) Rows of the true (blue) and estimated (red) covariance matrices for Experiment 7. Observation error covariance RMSE 0.020

Figure 7.10: Diagnostics for Experiment 7 (experiment type C with infrequent observations)



(a) Rank Histogram for Experiment 8



(b) Rows of the true (blue) and estimated covariance matrices for Experiment 8. Covariance calculated using the first 250 background and analysis innovations (red), observation error covariance RMSE 0.021. Covariance calculated using the last 250 background and analysis innovations (black), observation error covariance RMSE 0.015.

Figure 7.11: Diagnostics for Experiment 8 (experiment type D with infrequent observations)

7.3.3 Two different observation error covariance matrices with frequent observations

We now consider the case where for the first 500 time steps the true $\mathbf{R}$ is defined using the SOAR function as in the previous experiments, Figure 7.1. For time steps 501 to 1000

$\mathbf{R}^t$ is defined to be $\mathbf{R}^t = 2\mathbf{R}^I$. We show that using our method allows us to obtain good estimates of the both the matrix $\mathbf{R}$ defined by the SOAR function and the diagonal matrix $\mathbf{R}$.

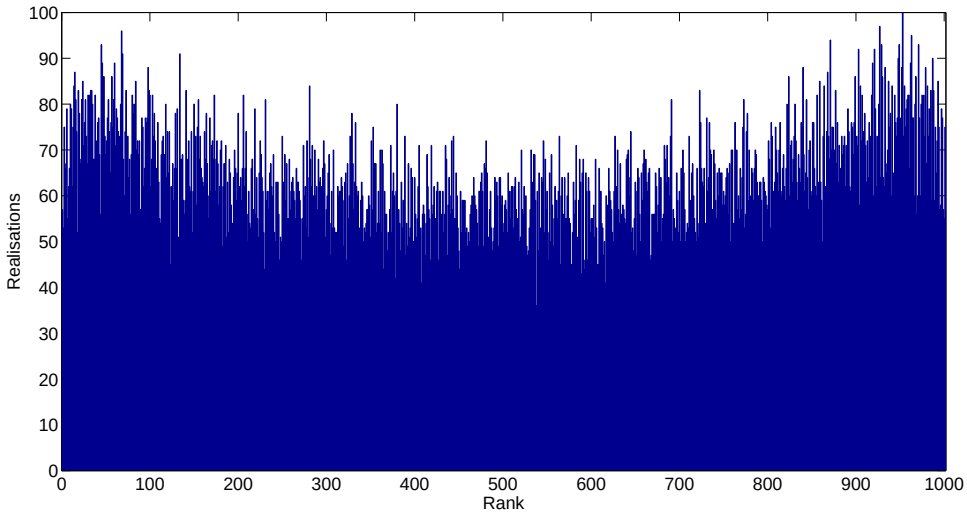
Experiments Type A, B and C

We run Experiments 9 (Type A), 10 (Type B) and 11 (Type C) using the described $\mathbf{R}$ and observations every 40 time steps. The RMSEs are given in Table 7.2 Experiments 9, 10 and 11 and the rank histograms and estimates of the observation error covariance matrix are plotted in Figures 7.12, 7.13, and 7.14 respectively. The estimate of the covariance matrix defined by the SOAR is calculated by using the first 500 background and analysis innovations. The background and analysis innovations from assimilation steps 501 to 1000 are used to estimate the true diagonal $\mathbf{R}$.

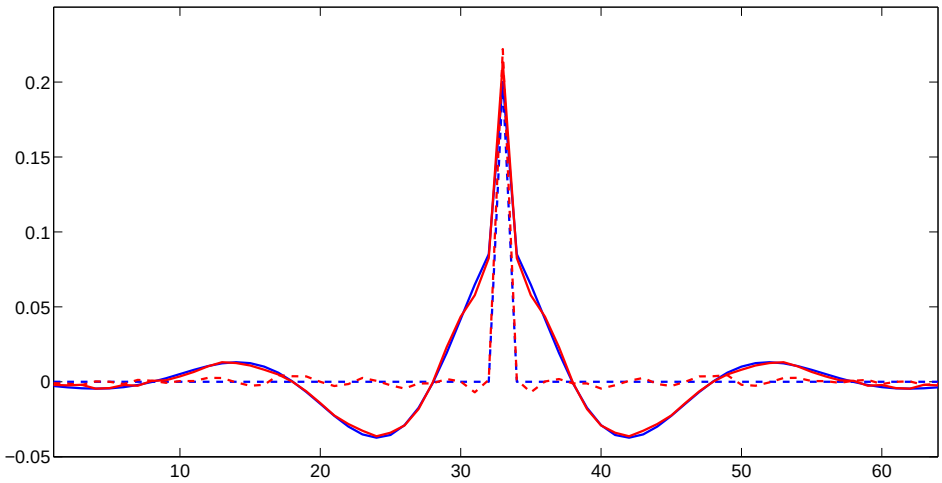
We see that using the correct $\mathbf{R}$ in the assimilation gives the lowest RMSE; however the filter also performs as well using an $\mathbf{R}$ with an inflated variance. Using $\mathbf{R}^{int} = \mathbf{R}^I$ results in a poorer performance of the filter both in terms of the RMSE and the rank histogram which suggests that the ensemble is under determined. However all three experiments result in good estimates of both covariance structures. We now consider if the observation error covariance matrix can be estimated within the assimilation scheme.

Experiment Type D

We show in Experiment 12 the case where $\mathbf{R}$ is estimated within the scheme, we initially assume $\mathbf{R} = \mathbf{R}^I$. By considering the rank histogram (Figure 7.15(a)) of the steps where $\mathbf{R}$ has been estimated we see that there is enough spread in the ensemble, and the RMSE suggests that the analysis had been slightly improved. We plot the estimates of the correlation structure of $\mathbf{R}$ every fifty time steps, as well as the two true correlation structures, in Figure 7.15(b). We see that the estimate of the correlation function at time $t_n = 300$ is already a good approximation of the true correlation function. This approximation is gradually improved over time. We see that after $t_n = 500$ the approximated correlation

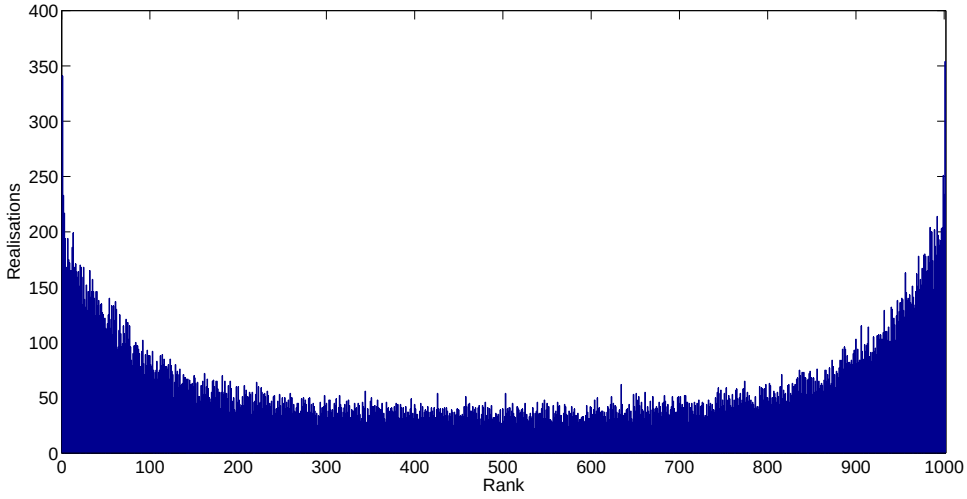


(a) Rank Histogram for Experiment 9

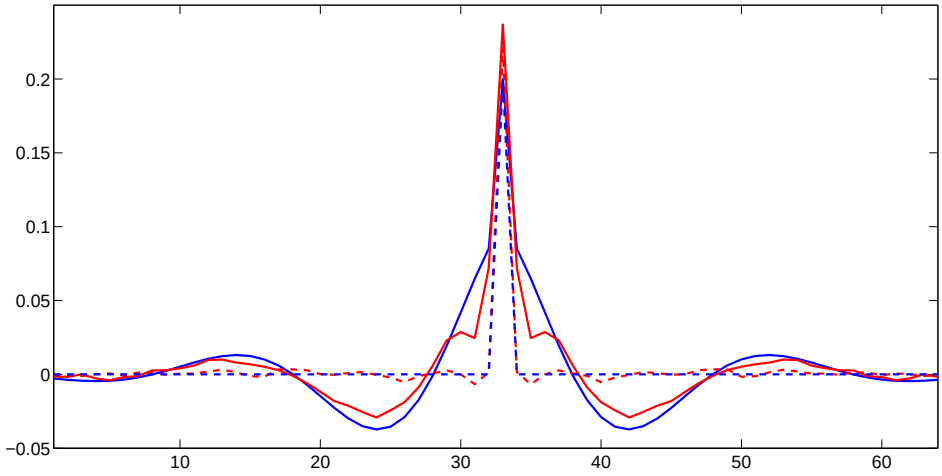


(b) Rows of the true (blue) and estimated (red) covariance matrices for the SOAR (solid lines) and diagonal (dashed lines) covariance functions for Experiment 9. Observation error covariance RMSE using first 500 innovations 0.010. Observation error covariance RMSE using innovations from assimilation time 501 to 1000 0.003

Figure 7.12: Diagnostics for Experiment 9 (experiment type A with two different observation error covariance matrices with frequent observations)

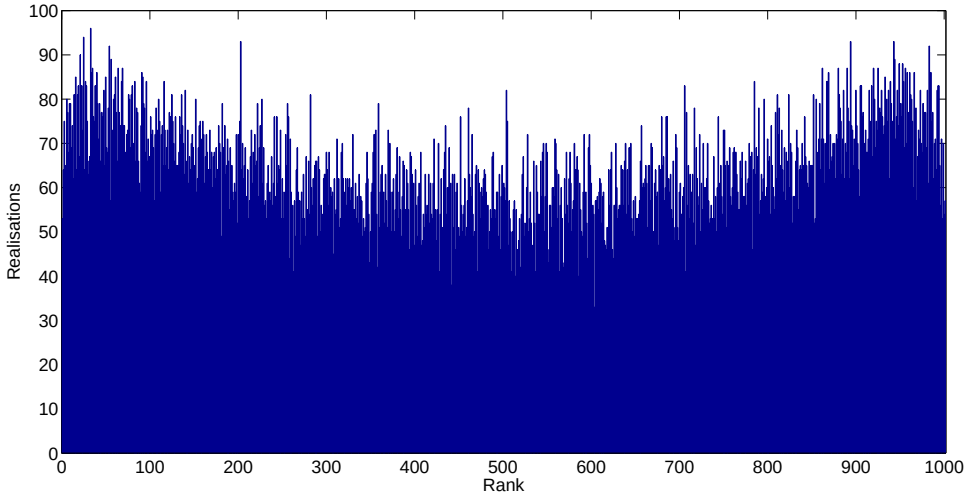


(a) Rank Histogram for Experiment 10

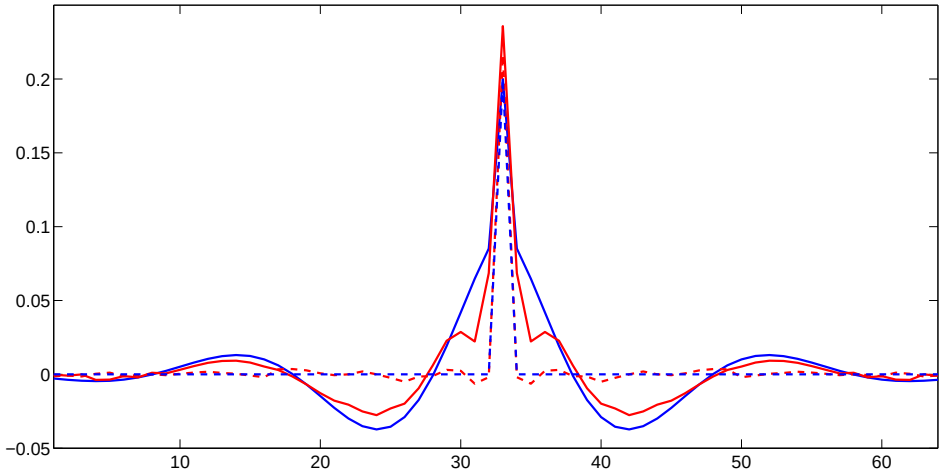


(b) Rows of the true (blue) and estimated (red) covariance matrices for the SOAR (solid lines) and diagonal (dashed lines) covariance functions for Experiment 10. Observation error covariance RMSE using first 500 innovations 0.010. Observation error covariance RMSE using innovations from assimilation time 501 to 1000 0.003

Figure 7.13: Diagnostics for Experiment 10 (experiment type B with two different observation error covariance matrices with frequent observations)

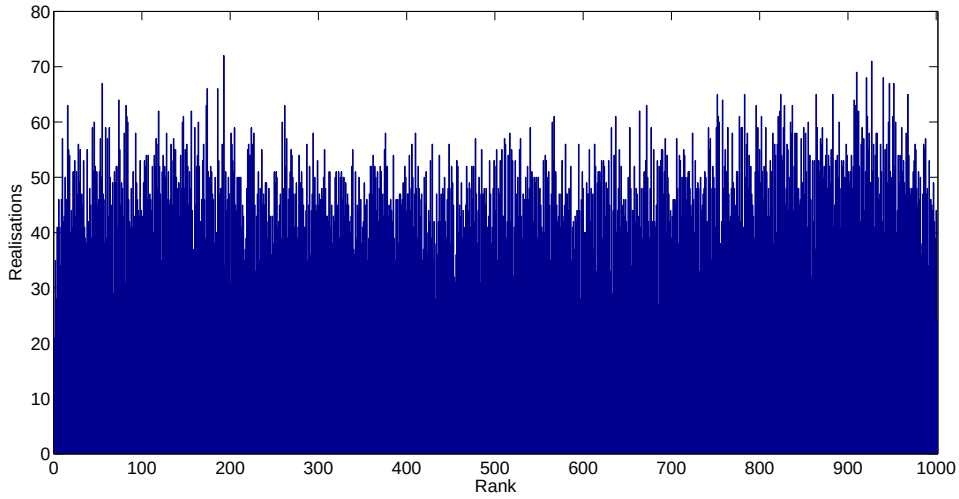


(a) Rank Histogram for Experiment 11

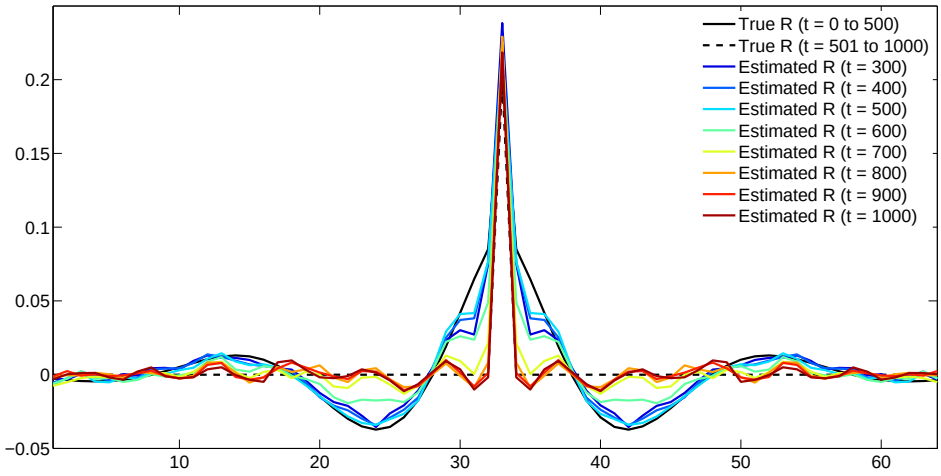


(b) Rows of the true (blue) and estimated (red) covariance matrices for the SOAR (solid lines) and diagonal (dashed lines) covariance functions for Experiment 11. Observation error covariance RMSE using first 500 innovations 0.011. Observation error covariance RMSE using innovations from assimilation time 501 to 1000 0.003

Figure 7.14: Diagnostics for Experiment 11 (experiment type C with two different observation error covariance matrices with frequent observations)



(a) Rank Histogram for Experiment 12



(b) Rows of the true (blue) and estimated covariance matrices for Experiment 12. Observation error covariance RMSE using innovations from assimilation time 1 to 250 0.010. Observation error covariance RMSE using innovations from assimilation time 251 to 500 0.006. Observation error covariance RMSE using innovations from assimilation time 501 to 750 0.006. Observation error covariance RMSE using innovations from assimilation times 750 to 1000 0.005.

Figure 7.15: Diagnostics for Experiment 12 (experiment type D with two different observation error covariance matrices with frequent observations)

function moves away from the SOAR function and by $t_n = 700$ the approximation is a good estimate of the new true covariance. Although the method can estimate the true covariance in this case once the true covariance structure is $\mathbf{R} = \mathbf{R}^{\mathbf{I}}$ it takes approximately 200 assimilation steps for the estimated $\mathbf{R}$ to be a good approximation of the truth. This

is no surprise as it is the current and previous 249 innovations that are being used for the estimation of $\mathbf{R}$. Until $t_n = 750$ some of the innovations will have been calculated where $\mathbf{R} = \mathbf{R}^I + SOAR$ so time is needed before the knowledge of these innovations is forgotten. The aim of developing this method was to approximate a slowly time varying $\mathbf{R}$ in that could be used to calculate a time dependent estimate of forward model error. In the case we have just presented the error is not slowly time varying, instead the method is presented with a discrete jump from one covariance to another, and this is the cause of the delay in the estimation of $\mathbf{R}$. We show that if the true $\mathbf{R}$ is slowly varying then innovations used in the $\mathbf{R}$ estimation will not be so different as in the case just presented, therefore a good approximation of $\mathbf{R}$ can be obtained. We now go on to consider this case.

7.3.4 Time dependent R

For this experiment we set the true $\mathbf{R}$ to be time dependent. We choose the correlation to be the SOAR function as described in equation (7.3) with $L_t = 15$. To create time dependence we vary the length scale with time by increasing b. At the initial time we set $b = 3.7$ and at each assimilation step b is increased by 3.0×10^{-4}, until the final assimilation time where $b = 4.0$. We only consider the case where $\mathbf{R}$ is estimated and used within the assimilation. We show that it is possible to use the ETKF and Desroziers diagnostic to estimate a time varying observation error covariance matrix. To show how well the filter is performing we give the analysis RMSE in Table 7.2 Experiment 13 and plot the rank histogram in Figure 7.16.

We see that the RMSE is low and the histogram is flat suggesting that the assimilation is working well and the ensemble spread is maintained. We now show how well the Desroziers diagnostic estimates the true observation error covariance matrix. We plot the estimates at every fifty time units in Figure 7.17. We see that the first estimate of $\mathbf{R}$ captures the true correlation structure well. Considering the estimates at each of the times plotted we see that the true correlation structure is well approximated. The ETKF with $\mathbf{R}$ estimation gives good estimates of a slowly time varying observation error covariance matrix. We now consider what happens when the covariance length scale decreases with time.

Again we chose the correlation to be the SOAR function as described in equation (7.3) with $L_t = 15$. To create time dependence we vary the length scale with time by decreasing b. At the initial time we set $b = 4.0$ at each assimilation step b is decreased by 3.0×10^{-4}, until the final assimilation time where $b = 3.7$. To show how well the filter is performing we give the analysis RMSE in Table 7.2 Experiment 14, plot the rank histogram in Figure 7.18 and we plot the estimates of the covariance matrix at every fifty time units in Figure 7.19. Again we see that the method produces reasonable estimates of the time varying observation error covariance matrix.

So far we have always considered how the method performs where the instrument, representativity and initial background error variances have been set to 0.1. We now consider how well the method performs where these errors are larger and smaller. We return to using the time varying observation error covariance matrix used in Experiment 13, but with the instrument, representativity and initial background error variances set to 0.01. We give the analysis RMSE in Table 7.2 Experiment 15, plot the rank histogram in Figure 7.20 and we plot the estimates of the covariance matrix at every fifty time units in Figure 7.21. We see that the RMSE is considerably lower than any of the previous experiments and this is a result of the increased accuracy of the observations. We see from the figure that the method works well where the error variances are small.

We now consider how the method performs where the instrument, representativity and initial background error variance are set to 1. We give the analysis RMSE in Table 7.2 Experiment 16, plot the rank histogram in Figure 7.22 and we plot the estimates of the covariance matrix at every fifty time units in Figure 7.23. The RMSE is large, and this is the result of the inaccurate observations in the assimilation. We also see that the rank histogram is U shaped, suggesting a lack of variability in the ensemble. Despite both the lack of variability and large RMSE we see that the method is still able to provide good approximations to the observation error covariance matrix. Comparing Experiments 13, 15 and 16, it appears that the size of the instrument and representativity error variance do not affect the accuracy of the approximation of the observation error covariance matrix.

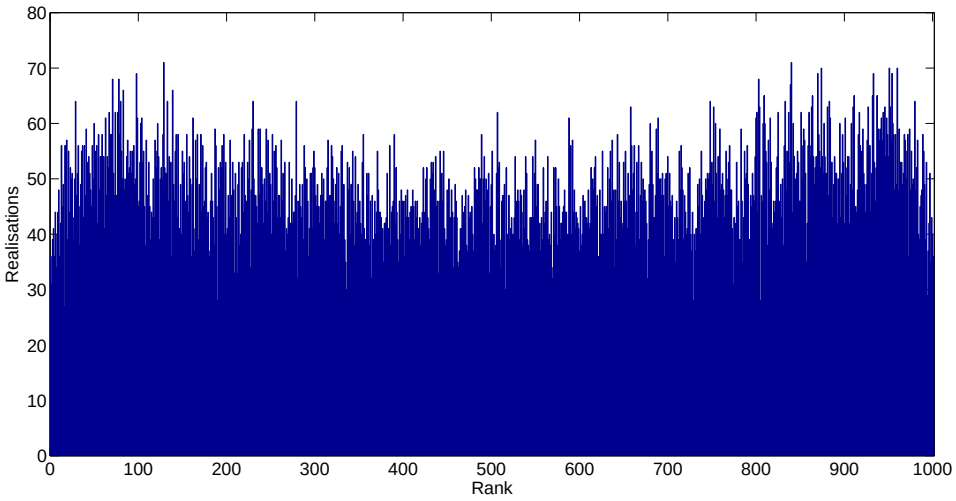


Figure 7.16: Rank Histogram for Experiment 13 (experiment type D with a time dependent $\mathbf{R}$, where L_o varies from 3.7 to 4.0, frequent observations and initial background, instrument and representativity error variances set to 0.1)

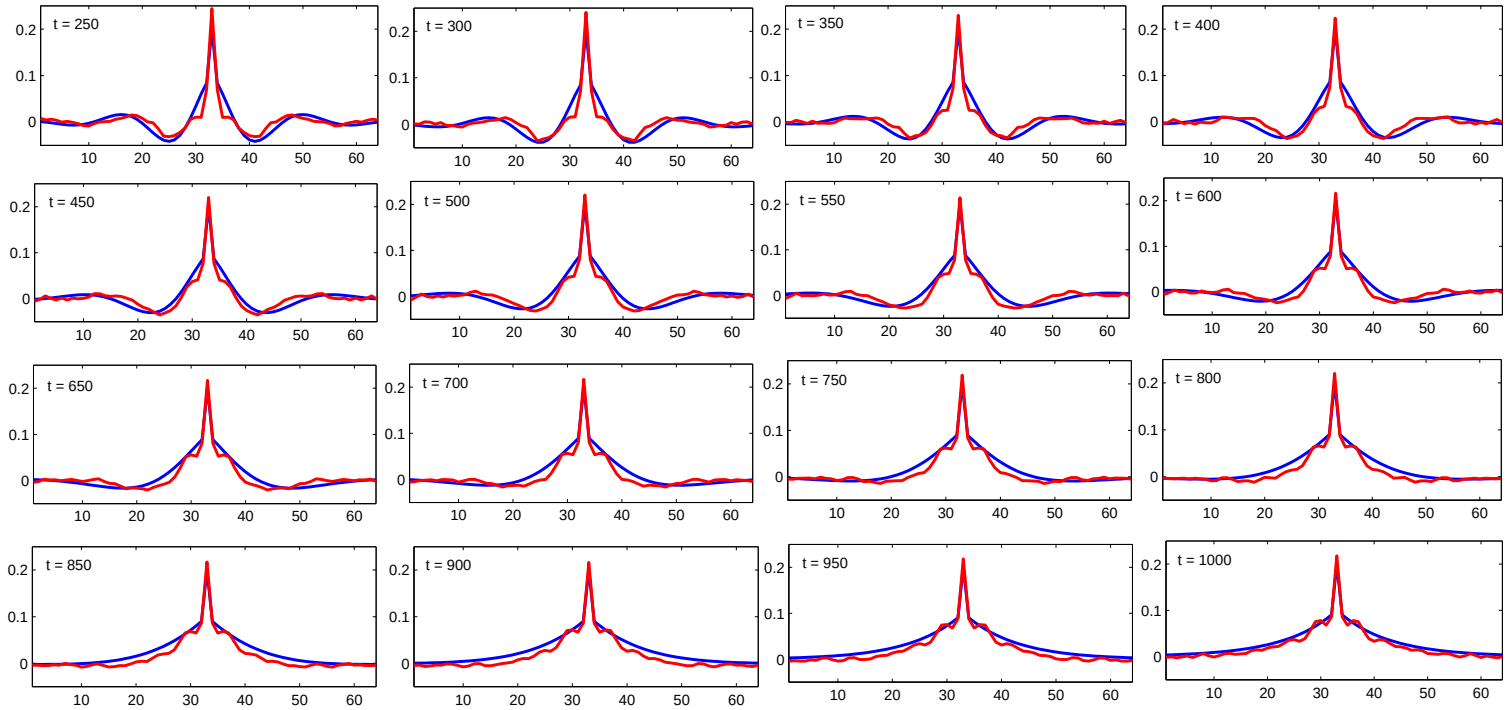


Figure 7.17: Rows of the true (blue) and estimated (red) covariance matrices for Experiment 13 (experiment type D with a time dependent $\mathbf{R}$, where L_o varies from 3.7 to 4.0, frequent observations and initial background, instrument and representativity error variances set to 0.1). Observation error covariance RMSE for final covariance estimate 0.008.

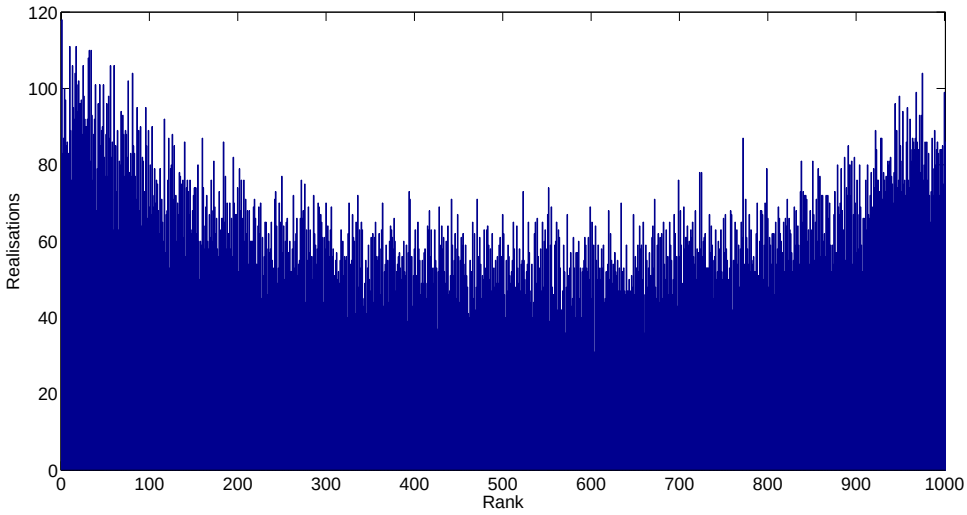


Figure 7.18: Rank Histogram for Experiment 14 (experiment type D with a time dependent **R**, where L_o varies from 4.0 to 3.7, frequent observations and initial background, instrument and representativity error variances set to 0.1)

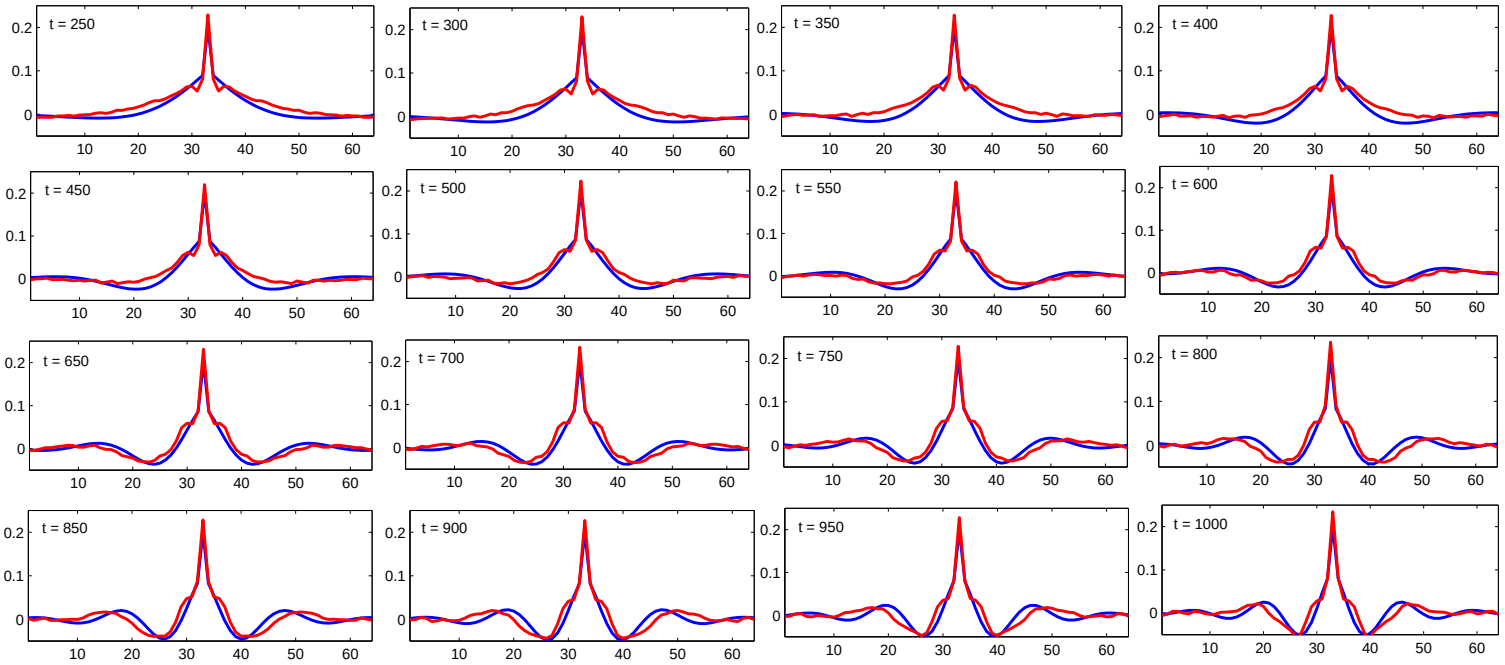


Figure 7.19: Rows of the true (blue) and estimated (red) covariance matrices for Experiment 14 (experiment type D with a time dependent $\mathbf{R}$, where L_o varies from 4.0 to 3.7, frequent observations and initial background, instrument and representativity error variances set to 0.1). Observation error covariance RMSE for final covariance estimate 0.014.

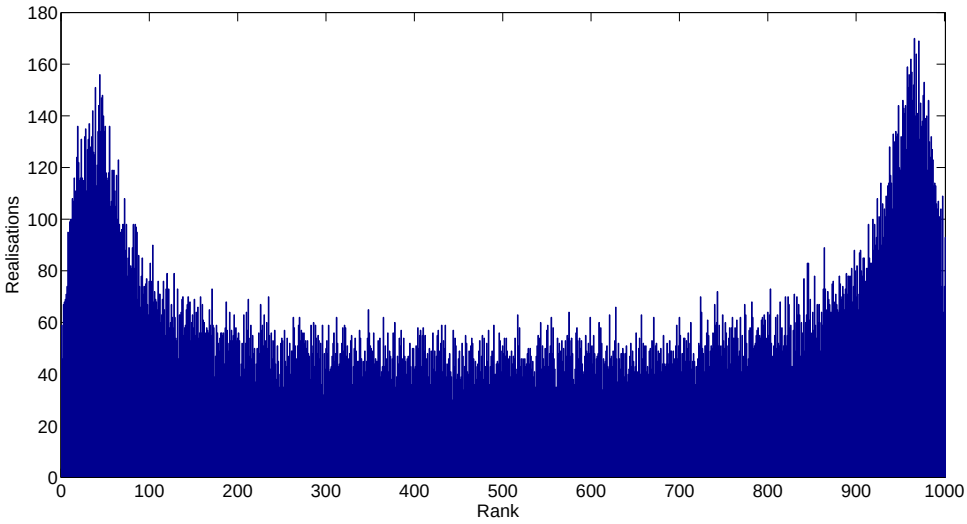


Figure 7.20: Rank Histogram for Experiment 15 (experiment type D with a time dependent **R**, where L_o varies from 3.7 to 4.0, frequent observations and initial background, instrument and representativity error variances set to 0.01)

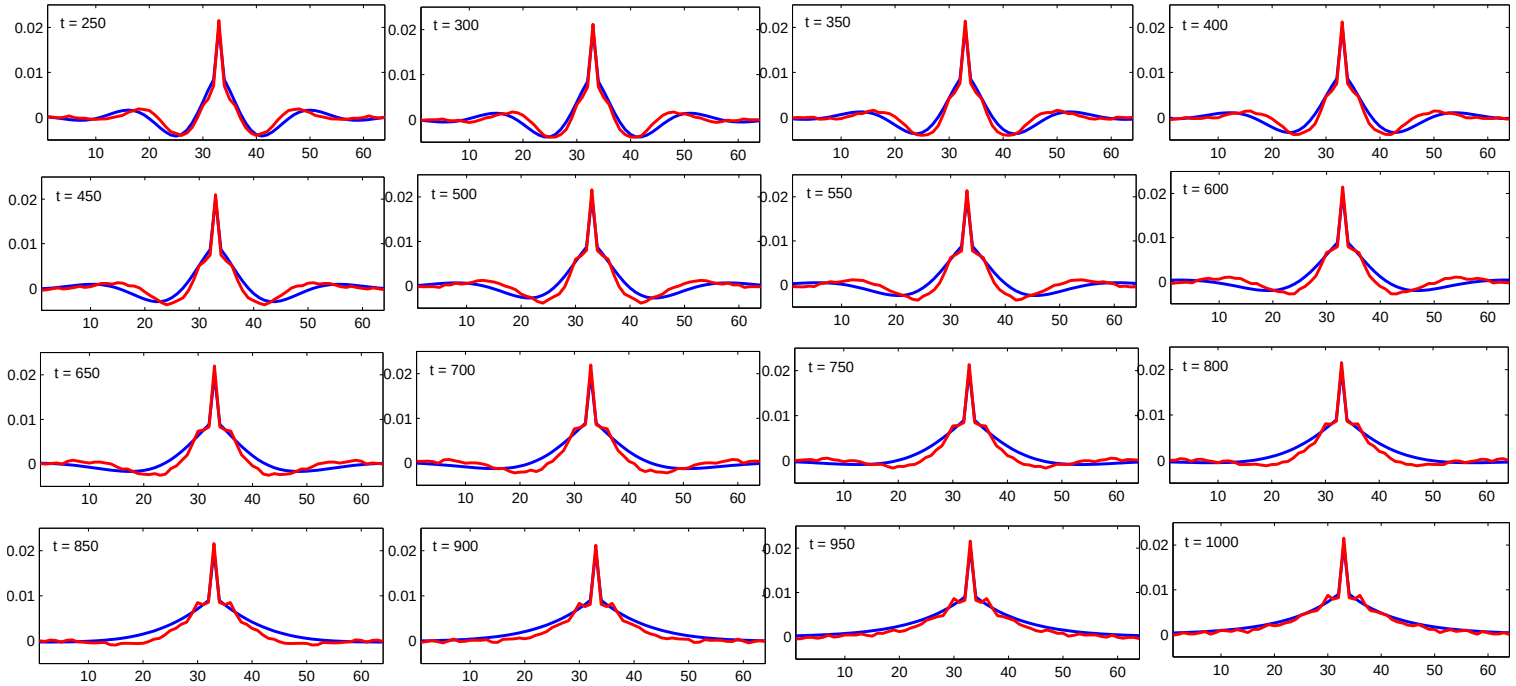


Figure 7.21: Rows of the true (blue) and estimated (red) covariance matrices for Experiment 15 (experiment type D with a time dependent $\mathbf{R}$, where L_o varies from 3.7 to 4.0, frequent observations and initial background, instrument and representativity error variances set to 0.01). Observation error covariance RMSE for final covariance estimate 0.001.

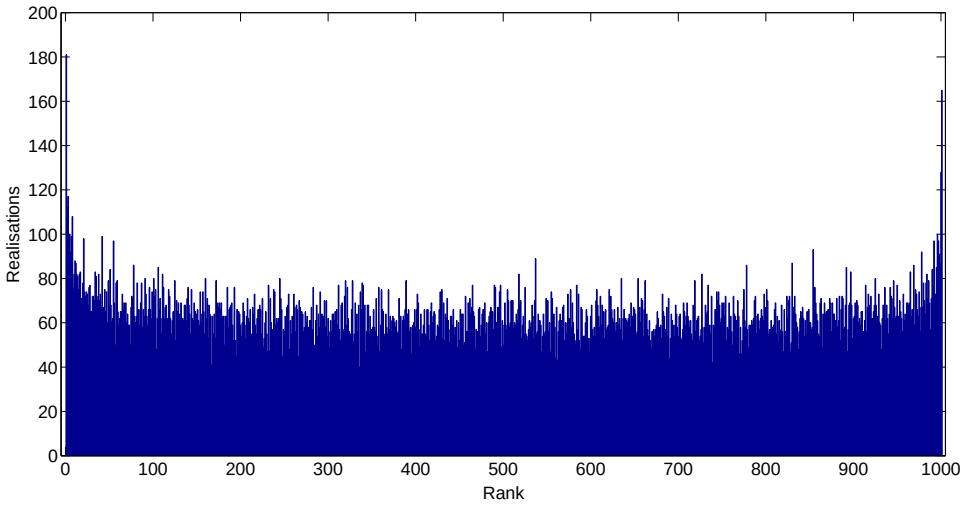


Figure 7.22: Rank Histogram for Experiment 16 (experiment type D with a time dependent $\mathbf{R}$, where L_o varies from 3.7 to 4.0, frequent observations and initial background, instrument and representativity error variances set to 1.0)

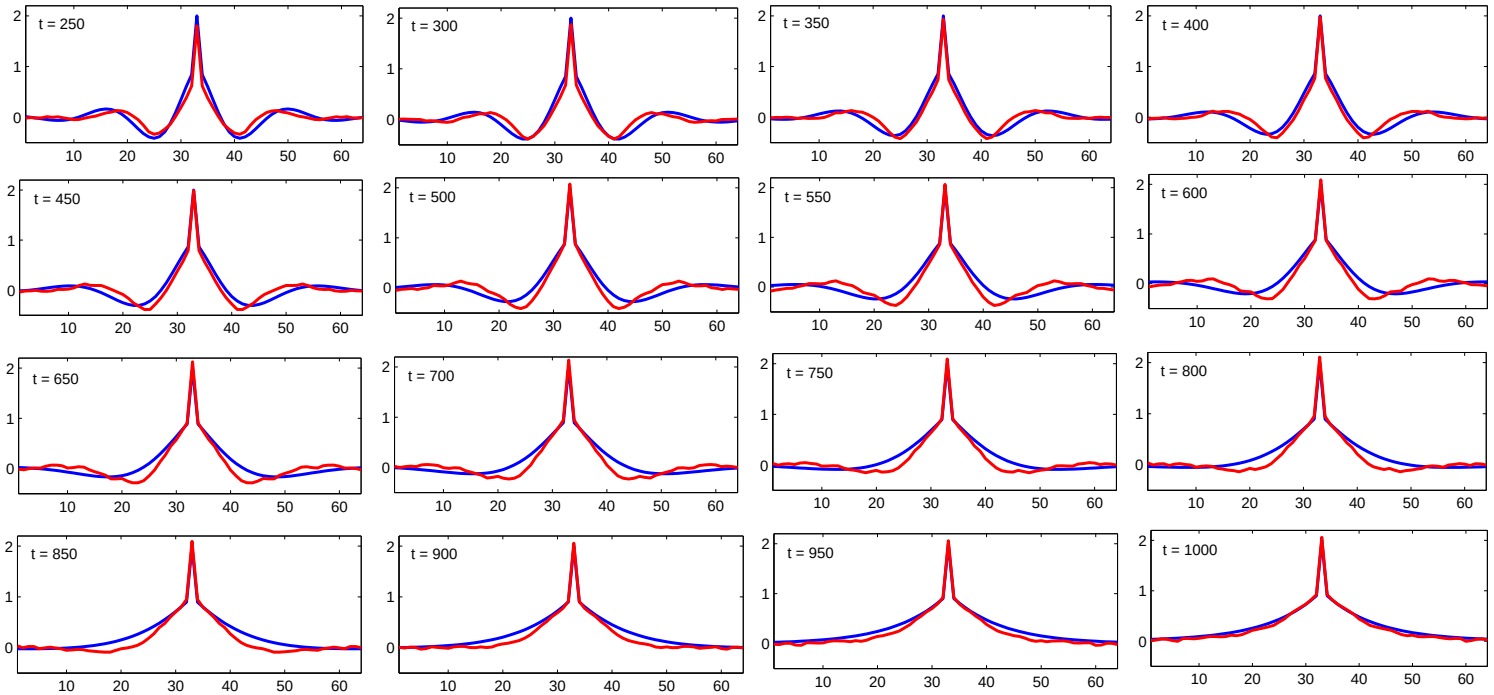


Figure 7.23: Rows of the true (blue) and estimated (red) covariance matrices for Experiment 16 (experiment type D with a time dependent $\mathbf{R}$, where L_o varies from 3.7 to 4.0, frequent observations and initial background, instrument and representativity error variances set to 1.0). Observation error covariance RMSE for final covariance estimate 0.044.

7.4 Summary

In this chapter we have introduced an ensemble transform Kalman filter with observation error covariance matrix estimation. This is an ETKF where analysis and background innovations are calculated at each time step and the most recent set of these innovations is used to estimate the matrix $\mathbf{R}$ using the Desroziers diagnostic. This estimate of $\mathbf{R}$ is then used in the next assimilation step. The method has been developed to allow a slowly time varying estimate of the observation error covariance matrix, and hence forward model error, to be calculated. We showed it is possible to obtain a good estimate of $\mathbf{R}$ using the Desroziers diagnostic; the best result is obtained where the correct matrix is used in the assimilation. However, even if the $\mathbf{R}$ used in the assimilation is diagonal it is still possible to obtain a reasonable estimate of the true correlation structure. We then showed that estimating $\mathbf{R}$ within the ETKF worked well, with good estimates obtained, the ensemble spread maintained and the analysis RMSE reduced compared to the case where the matrix $\mathbf{R}$ is always assumed diagonal. We also showed that the method does not work as well where the observations are less frequent. However the method still produces a reasonable estimate of $\mathbf{R}$, maintains the ensemble variance and the analysis RMSE is lower than where a diagonal $\mathbf{R}$ is used. We also showed that the method worked well where the true matrix $\mathbf{R}$ is defined by the SOAR function for the first 500 time units and then diagonal for the second half of the assimilation. We showed that in this situation the method worked well initially; however due to the discrete change in the matrix $\mathbf{R}$ the method took approximately 250 assimilation steps to approximate well the true $\mathbf{R}$ defined for the second half of the assimilation. Finally we considered a case where $\mathbf{R}$ varied slowly with time. It is not known how quickly representativity error will vary, but given the case dependence seen in the previous chapter it is likely that representativity error will vary at the same rate as the synoptic situation changes. We showed that the method worked well where the true $\mathbf{R}$ was defined to slowly vary with time. The analysis RMSE was low and the rank histogram suggested that the ensemble spread was maintained. The estimates of the correlation structure were good, suggesting that the method is capable of estimating a slowly time varying observation error covariance matrix. We also showed

that the ability for the method to approximate the correlation structure was not sensitive to the background errors or the true magnitude of the observation error variance. This suggests that the method would be suitable to give a time dependent estimate of forward model or representativity error. We note that the effectiveness of the method will depend on how rapidly the synoptic situation and hence representativity error is changing and how often observations are available. The representativity error will also be dependent on the NWP model that is being used. For models designed to capture rapidly developing situations, where representativity error is likely to change rapidly, assimilation cycling and observation frequency within the assimilation is expected to be more frequent and hence more data is available for estimating the representativity error.

Chapter 8

Conclusions

Data assimilation is an important technique that combines observations with a model prediction to find the best estimate of the true state of a dynamical system [Kalnay, 2002]. It is used to provide a complete set of initial conditions as input into a numerical model. The accuracy of the initial conditions is important as any error will be propagated by the numerical model. Both the observations and model prediction contain errors and their statistics are included in the assimilation in the observation and background error covariance matrices. The errors associated with the observations are the instrument and forward model errors. The instrument error is determined for specific instruments under a set of test conditions by the instrument manufacturer or from in-orbit calibration data. Forward model errors consist of the errors due to a misspecified observation operator and errors of representativity. Errors of representativity are the result of the small scale observation information being incorrectly represented in the model [Daley, 1993]. Currently little is known about representativity errors and they are not correctly included in data assimilation schemes. Previous work in the context of atmospheric data assimilation [Stewart et al., 2009, 2012b, Bormann et al., 2002, Bormann and Bauer, 2010, Bormann et al., 2010] has shown that the observation error covariance matrix may be correlated, but it is not known if these correlations are in part caused by the representativity error. In this thesis we have use existing methods and developed our own schemes to investigate forward model error and representativity error. A better understanding of these errors would allow them

to be incorporated into the observation error statistics to provide more accurate observation error covariance matrix and also would allow us to make better use of the available observations. In turn this could improve the analysis, which would provide better initial conditions for forecasting. We now summarise the work in this thesis and highlight the main conclusions. We then present ideas for further work.

8.1 Summary

In Chapter 1 we set out the main aims that we proposed to answer throughout the thesis. In particular we wished to:

- Understand what representativity error is and how it can be calculated and included in the data assimilation scheme.
- Understand the structure of representativity error and see if it may be a cause of correlations in the observation error covariance matrix.
- Understand if representativity error is significant.
- To see if the inclusion of representativity error in the data assimilation scheme can improve the analysis.
- To see if it is possible to calculate a time dependent estimate of forward model error.

In Chapter 2 we introduced the concepts of data assimilation. The notation for dynamical systems and data assimilation used throughout this thesis was also introduced. A brief overview of some different types of sequential and variational data assimilation was given, with the best linear unbiased estimate and ensemble transform Kalman filter described in greater detail. Two methods used to overcome problems with ensemble filtering were introduced and diagnostics for data assimilation were also considered.

In Chapter 3 the ideas of forward model error and representativity error were introduced and a mathematical description of forward model error was given. The current treatment of forward model error was discussed. The methods that have been previously developed

to estimate $\mathbf{R}$, as well as those that give both time independent and dependent estimates of forward model errors were also reviewed. A method proposed by Daley [1993] and the Desroziers diagnostic were discussed in greater detail.

The use of the Daley [1993] method was described in Chapter 4. The method, suitable for calculating representativity errors in idealised situations, requires pseudo-observations and other matrices and the details of these were derived in this chapter. Some new theoretical results related to this method were then presented.

In Chapter 5 we introduced the Kuramoto-Sivashinsky equation and used it along with the method proposed by Daley [1993] to help us understand the structure of representativity error. The ETDRK4 numerical method used to solve the KS equation was introduced. We considered solutions to the KS equation at different resolutions, and the power spectra of these solutions. We then calculated representativity error for the KS equation. We considered the effect of altering the number and type of observations as well as altering the model resolution.

In Chapter 6 we investigated the significance of representativity error for temperature and specific humidity fields. We also showed that significant correlations in the observation error covariance matrix [Stewart et al., 2009, Stewart, 2010] could be attributed to representativity error. This was achieved by calculating representativity error for temperature and humidity data from the Met Office UKV model for two different cases. The sensitivity of representativity error to the synoptic state was also investigated.

Finally in Chapter 7 we introduced a new method that combines an ensemble transform Kalman filter with observation error covariance matrix estimation. The method combined an ETKF with the Desroziers diagnostic, which is used to estimate the observation error covariance matrix $\mathbf{R}$. This method was introduced as it could be used to calculate a time dependent estimate of representativity error. We used this method to assimilate high resolution observations with the model solutions of the KS equation. We carried out experiments to show how well the method worked. We showed that this new method could be used to estimate both static and time varying observation error covariance matrices.

8.2 Conclusions

In this thesis we have used existing methods and developed our own schemes to investigate forward model error and representativity error. We have shown that:

- Representativity error is correlated.

  In Chapter 5 and 6 the structure of representativity error was investigated. We calculated representativity error for both idealised and real data using a method first developed by Daley [1993] and then by Liu and Rabier [2002]. All the results showed that representativity error is correlated. This suggests that the correlations found in Stewart [2010] and Weston [2011] are likely to be caused, in part, by representativity error.

- The representativity error variance is independent of the number of available observations and the correlation structure of representativity error is dependent not on the number of observations, but the distance between them.

  In Chapter 4 we presented theoretical results relating to the Daley [1993] method. We proved that when using the Daley [1993] method the variance of representativity error does not change when calculated with different numbers of observations. We also showed that the correlation structure of the representativity error depends only on the distance between observations and not the number of observations available. The numerical results when using the KS equation, Chapter 5, and the Met Office data, Chapter 6 supported the theoretical results. Although the theoretical results are specific to the Daley [1993] method, in general we expect that the representativity error should not be dependent on the number of available observations, only on the distance between them.

- Representativity error can be reduced by increasing the model resolution.

  We considered the effect that changing the model resolution had on the representativity error. Results when using the KS equation, Chapter 5, and the Met Office data, Chapter 6 showed that representativity error decreases as model resolution in-

creases. This is because a model at higher resolution resolves more of the scales that are resolved by the observations.

- Representativity error is lower for observations with larger lengthscales

  We considered the effect that changing the observation type and lengthscale had on the representativity error. Results when using the KS equation, Chapter 5, and the Met Office data, Chapter 6 showed that variance of the representativity error reduced as the lengthscale of the observation increased. This is because observations with larger lengthscales average over the smallest scales and therefore the difference between the resolved scales in the observation and model is reduced, hence reducing the representativity error. We also showed that for observations with larger lengthscales the correlations were more significant. This increase in significance is related to the fact that neighbouring observations overlap, hence they share information about the state.

- Representativity error is case dependent, more significant for humidity than temperature and varies throughout the atmosphere.

  In Chapter 6 we used data from the high resolution Met Office model to calculate representativity error for temperature and specific humidity for two different synoptic situations. We found that the representativity error was sensitive to the synoptic situation, which supports claims by Janjic and Cohn [2006] that representativity error is time and state dependent.

  The experiments showed also that representativity error was more significant for humidity than temperature.

  We considered how the representativity error standard deviation varied at different pressure levels. We found that representativity error does vary at different pressure levels and this means that assumptions such as those in Dee and Da Silva [1999] where errors at different model levels are fixed may not be suitable when representativity error is taken into account in assimilation schemes. We found that the highest representativity errors for specific humidity occurred at the pressure levels where

cloud was present.

- It is possible to estimate time varying observation error covariance matrices, and including these within the assimilation scheme can improve the analysis.

  In Chapter 7 we showed it was possible to calculate time dependant representativity error and improve the analyisis by accounting for representativity error in the assimilation scheme. We developed a new method that provides an online estimation of the observation error covariance matrix. We showed it was possible to obtain a good estimate of $\mathbf{R}$ using the Desroziers diagnostic. We then showed that estimating $\mathbf{R}$ within the ETKF worked well, with good estimates obtained, the ensemble spread maintained and the RMSE reduced compared to the case where the matrix $\mathbf{R}$ was always assumed diagonal. We also showed that the method does not work as well when the observations are less frequent. However the method still produces a reasonable estimate of $\mathbf{R}$ and maintains the ensemble variance. Finally we showed that the method worked well where the true matrix $\mathbf{R}$ was defined to slowly vary with time. The estimates of the correlation structure were good, suggesting that the method is capable of estimating a slowly time varying observation error covariance matrix. We conclude that the method would be suitable to give a time dependent estimate of forward model or representativity error.

8.3 Future work

In this thesis we have attempted to understand the structure of representativity error and include it in an assimilation scheme. In Chapter 7 we introduced a method that combined an ETKF with the Desroziers diagnostic that can be used to give a time dependent estimate of representativity error. Under the assumptions we made, the method worked well and provided a good estimate of a time dependent observation error covariance matrix. When this error covariance matrix was included in the assimilation scheme the analysis was improved. However the method should be tested with the assumptions removed.

Although not a requirement of the method, all our experiments have assumed that the

observation error covariance matrix is homogeneous. This has allowed the sampling error to be reduced. In a true NWP situation the observation error covariance matrix is unlikely to be homogeneous, so rather than make this assumption it will be necessary to determine a different way to deal with the sampling error. One way to do this would be to increase the number of samples, as samples are taken over time increasing the number of samples will reduce the time dependency of the $\mathbf{R}$ estimate. Work needs to be done to understand how many samples are required for there to be a balance between a good time dependent estimate of $\mathbf{R}$ and a reduction in sampling error. As well as considering increasing the number of samples, it may also be beneficial to consider covariance localisation to reduce the sampling error.

For the Desroziers diagnostic to give an accurate estimate of $\mathbf{R}$ it is suggested that the $\mathbf{R}$ and $\mathbf{P}^f$ used in the analysis should be correctly specified. To ensure that we have an accurate $\mathbf{P}^f$ we have used $N = 1000$ ensemble members. Using this large number of ensembles is computationally costly and will not be practical in larger systems. In Chapter 2 we discussed the methods of covariance inflation and localisation that can be used to increase the effective ensemble size. Further work is required to test the method with a reduced ensemble size and the inclusion of inflation and localisation to see what effect these methods would have on the estimates of $\mathbf{B}$, and subsequently the estimate of $\mathbf{R}$.

We have also only considered calculating a time dependent $\mathbf{R}$ for direct observations. It would be interesting to compare how this method works for calculating $\mathbf{R}$ for a different observation type. This could be done by creating pseudo-observations using the weighting matrices defined in Chapter 4. After considering different observation types it would be interesting to see if the method is suitable for estimating $\mathbf{R}$ that is associated with more than one observation type.

All these further research ideas, could be carried out using the same technique used in Chapter 7 where the model used is at the same resolution as the truth and a pseudo-representativity or forward model error is added to the observations. This allows the method to be examined without the interaction of any other model errors. Once the method is better understood it should be used in twin experiments where the model is run

at a lower resolution than the 'truth', and hence real representativity error is present. The method should be tested in this case to calculate forward model errors and include them in the scheme to improve the analysis.

In this thesis errors of representativity have been investigated. It has been shown that these errors are correlated and a method has been developed that allows a time dependent estimate of these errors to be calculated. With further development of these methods it is possible that representativity errors could be correctly included in the assimilation in a context of numerical weather prediction.

Bibliography

J. L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129:2884–2903, 2001.

J. L. Anderson and S. L. Anderson. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 126:2741–2758, 1999.

J. S. Bendat and A. G. Piersol. *Random Data: Analysis and Measurement Procedures*. John Wiley and Son, fourth edition, 2011.

C. Bishop, B. Etherton, and S. Majumdar. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129:420–436, 2001.

W. Bolstad. *Introduction to Bayesian Statistics*. John Wiley and Son, second edition, 2007.

N. Bormann and P. Bauer. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Quarterly Journal of the Royal Meteorological Society*, 136: 1036–1050, 2010.

N. Bormann, S. Saariene, G. Kelly, and J. Thepaut. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Monthly Weather Review*, 131:706–718, 2002.

N. Bormann, A. Collard, and P. Bauer. Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II:

Application to AIRS and IASI data. *Quarterly Journal of the Royal Meteorological Society*, 136:1051–1063, 2010.

M. Buehner and M. Charron. Spectral and spatial localization of background-error correlations for data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 133: 615–630, 2007.

M. Buehner, P. Houtekamer, C. Charette, H. Mitchell, and B. He. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Monthly Weather Review*, 138:1567–1586, 2010.

G. Burgers, P. van Leeuwen, and G. Evensen. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126:1719–1724, 1998.

A. M. Clayton, A. C. Lorenc, and D. M. Barker. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quarterly Journal of the Royal Meteorological Society*, 2012. Early View. doi: 10.1002/qj.2054.

S. E. Cohn. Introduction to estimation theory. *Journal of the Meteorological Society of Japan*, 75:275–288, 1997.

S. M. Cox and P. C. Matthews. Exponential time differencing for stiff systems. *Journal of Computational Physics*, 176:430–455, 2000.

R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, 1991.

R. Daley. Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophysicae*, 11:634–647, 1993.

M. Dando, A. Thorpe, and J. Eyre. The optimal density of atmospheric sounder observations in the met office NWP system. *Quarterly Journal of the Royal Meteorological Society*, 133:1933–1943, 2007.

D. P. Dee and A. M. Da Silva. Maximum-likelihood estimation of forecast and observation

error covariance parameters. Part I: Methodology. *Monthly Weather Review*, 127:1822–1843, 1999.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131:3385–3396, 2005.

G. Desroziers, L. Berre, and B. Chapnik. Objective validation of data assimilation systems: diagnosing sub-optimality. *In Proceedings of Workshop on diagnostics of data assimilation system performance, 15-17 June 2009*, 2009.

P. Eden. Weather log: August 2007. *Weather*, 62:i–iv, 2007.

P. Eden. Weather log: September 2008. *Weather*, 63:i–iv, 2008.

V. M. Eguíluz, P. Alstrøm, E. Hernández-García, and O. Piro. Average patterns of spatiotemporal chaos: A boundary effect. *Phys. Rev. E*, 59:2822–2825, 1999.

G. Evensen. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical research*, 99: 10143–10162, 1994.

G. Evensen. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

R. Furrer and T. Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98:227–255, 2007.

G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125:723–757, 1999.

P. Gauthier, M. Tanguay, S. Laroche, and S. Pellerin. Extension of 3D-Var to 4D-Var: Implementation of 4D-Var at the meteorological service of Canada. *Monthly Weather Review*, 135:2339–2354, 2007.

A. Gelb, editor. *Applied Optimal Estimation*, chapter 6. The M.I.T. Press, 1974.

J. Gustafsson and B. Protas. Regularization of the backward-in-time Kuramoto-Sivashinsky equation. *Journal of Computational and Applied Mathematics*, 234:398–406, 2010.

T. M. Hamill. Interpretation of rank histograms for varifying ensemble forecasts. *Monthly Weather Review*, 129:550–560, 2000.

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

S. B. Healy and A. A. White. Use of discrete Fourier transforms in the 1D-Var retrieval problem. *Quarterly Journalof the Royal Meteorological Society*, 131:63–72, 2005.

F. Hilton, A. Collard, V. Guidard, R. Randriamampianina, and M. Schwaerz. Assimilation of IASI radiances at european NWP centres. *In Proceedings of Workshop on the assimilation of IASI data in NWP, ECMWF, Reading, UK, 6-8 May 2009*, pages 39–48, 2009.

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

P. L. Houtekamer and H. L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126:796–811, 1998.

P. L. Houtekamer and H. L. Mitchell. Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 133:3260–3289, 2005.

T. Janjic and S. E. Cohn. Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review*, 134:2900–2915, 2006.

M. Jardak, I. M. Navon, and M. Zupanski. Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation. *International Journal for Numerical Methods in Fluids*, 62:374–402, 2000.

A. H. Jazwinski. *Stochastic Processes and Filtering Theory.* Accademic Press, 1970.

R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering (Seried D)*, 82:35–45, 1960.

R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Transactions of the ASME-Journal of Basic Engineering (Seried D)*, 83:95–108, 1961.

E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2002.

A. Kassam and L. Trefethen. Fourth-order time-stepping for stiff pdes. *SIAM J. Sci. Computing*, 25:1214–1233, 2005.

E. Kendon, N. Roberts, C. Senior, and M. Roberts. Realism of rainfall in in a very high-reolution regional climate model. *Journal of Climate*, 25:5791–5806, 2012.

Y. Kuramoto. Diffusion-induced chaos in reaction systems. *Progress of Theoretical Physics*, 64:346–367, 1978.

W. Lahoz, B. Khattatov, and R. Menard, editors. *Data Assimilation Making Sense of Observations*, chapter 2. Springer, 2010.

H. Lean, P. Clark, M. Dixon, N. Roberts, A. Fitch, R. Forbes, and C. Halliwell. Charictaristics of high-resolution versions of the met office unified model for forecasting convection over the united kingdom. *Monthly Waether Review*, 136:3408–3424, 2008.

J. M. Lewis, S. Lakshmivarahan, and S. K. Dhall. *Dynamic Data Assimilation*. Cambridge University Press, 2006.

H. Li, E. Kalnay, and T. Miyoshi. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 128:1367–1386, 2009.

Z.-Q. Liu and F. Rabier. The interaction between model resolution observation resolution and observation density in data assimilation: A one dimensional study. *Quarterly Journal of the Royal Meteorological Society*, 128:1367–1386, 2002.

Z.-Q. Liu and F. Rabier. The potential of high-density observations for numerical weather prediction: A study with simulated observations. *Quarterly Journal of the Royal Meteorological Society*, 129:3013–3035, 2003.

D. Livings. Aspects of the ensemble Kalman filter. Master's thesis, University of Reading, 2005.

D. M. Livings, S. L. Dance, and N. K. Nichols. Unbiased ensemble square root filters. *Physica D*, 237:1021–1028, 2008.

A. C. Lorenc. A global three-dimensional multivariate statistical interpolation scheme. *Monthley Weather Review*, 109:701–721, 1981.

A. C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112:1177–1194, 1986.

A. C. Lorenc. The potential of the ensemble Kalman filter for NWPa comparison with 4d-var. *Quarterly Journal of the Royal Meteorological Society*, 129:3183–3203, 2003.

A. C. Lorenc, R. Bell, and B. Macpherson. The Meterological Office analysis correction data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 117: 59–89, 1991.

A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126:2991–3012, 2000.

R. Mènard, S. Cohn, L. Chang, and P. Lyster. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: Formulation. *Monthly Weather Review*, 128, 2000.

R. Mènard, Y. Yang, and Y. Rochon. Convergence and stability of estimated error variances derived from assimilation residuals in observation space. *In Proceedings of Workshop on diagnostics of data assimilation system performance, 15-17 June 2009*, 2009.

T. Miyoshi, Y. Sato, and T. Kadowaki. Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. *Monthly Weather Review*, 138:2846–2866, 2010.

T. Miyoshi, E. Kalnay, and H. Li. Estimating and including observation-error correlations in data assimilation. *Inverse Problems in Science and Engineering*, 21, 2013. Early View. doi: 10.1080/17415977.2012.712527.

R. H. Nevanlinna. *Introduction to complex analysis*. AMS, second edition, 1969.

P. R. Oke and P. Sakov. Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and oceanic technology*, 25:1004–1017, 2007.

E. Pavelin, S. English, and F. J. Bornemann. MTG-IRS simulation project. Technical report, Met Office, UK, 2009. www.eumetsat.int/groups/pps/documents/document/pdf_mtg_rep36.pdf.

B. Protas. Adjoint-based optimization of PDE systems with alternative gradients. *Journal of Computational Physics*, 227:6490–6510, 2008.

F. Rabier, H. Jrvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126:1143–1170, 2000.

F. Rawlins, S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne. The met office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 133:347–362, 2007.

W. J. Reichmann. *Use and Abuse of statistics*. Oxford University Press, 1962.

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

I. Schur. Bemerkungen zur theorie der beschrnkten bilinearformen mit unendlich vielen vernderlichen. *J. reine angew. Math.*, 140:1–28, 1911.

R. S. Seaman. Absolute and differential accuracy of analysis achievable with specified observational network charictaristics. *Monthly Weather Review*, 105:1211–1222, 1977.

G. Sivashinsky. Non-linear analysis of hydrodynamic instability in laminar flames. *Acta Astronautica*, 4:1177–1206, 1977.

M. Small and C. K. Tse. Applying the method of surrogate data to cyclic time series. *Physica D*, 164:187–201, 2002.

L. M. Stewart. *Correlated observation errors in data assimilation*. PhD thesis, University of Reading, 2010.

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

L. M. Stewart, J. Cameron, S. L. Dance, S. English, J. R. Eyre, and N. K. Nichols. Observation error correlations in IASI radiance data. Technical report, University of Reading, 2009. Mathematics reports series, www.reading.ac.uk/web/FILES/maths/obs_error_IASI_radiance.pdf.

L. M. Stewart, S. L. Dance, and N. K. Nichols. Data assimilation with correlated observation errors: analysis accuracy with approximate error covariance matrices. Technical report, University of Reading, 2012a. Department of Mathematics and Statistics Preprint MPS-2012-17, www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx.

L. M. Stewart, S. L. Dance, N. K. Nichols, J. R. Eyre, and J. Cameron. Estimating interchannel observation error correlations for IASI radiance data in the met office system. Technical report, University of Reading, 2012b. Department of Mathematics and Statistics Preprint, www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx.

R. Swinbank, V. Shutyaev, and W. A. Lahoz. *Data Assimilation for the Earth System*. Kluwer Academic Publishers, 2003.

O. Talagrand. Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 75:191–209, 1997.

Y. Tang, H. Lean, and J. Bornemann. The benefits of the Met Office variable resolution NWP model for forecasting convection. *Meteorological Applications*, 2012. Early View. doi: 10.1002/met.1300.

J. Theiler, S. Eubank, A. Longtin, G. B., and J. D. Farmer. Testing for non-linearity in time series: the method of surrogate data. *Physica D*, 58:77–94, 1992.

M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamil, and J. S. Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131:1485–1490, 2003.

L. Trefethen. *Spectral methods in MATLAB*. SIAM, 2000.

UK Met Office. http://www.metoffice.gov.uk/research/modelling-systems, Last accessed 18^{th} February 2013.

J. Waller (née Pocock), S. L. Dance, A. Lawless, N. K. Nichols, and J. R. Eyre. Representativity error for temperature and humidity using the Met Office UKV model. Technical report, University of Reading, 2012. Department of Mathematics and Statistics Preprint, www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx.

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

J. S. Whitaker, T. M. Hamill, X. Wei, Y. Song, and Z. Toth. Ensemble data assimilation with the NCEP global forecast system. *Monthly Weather Review*, 136:463–482, 2008.