Accounting for Model Error
in Four-Dimensional Variational
Data Assimilation

Katherine E. Howes

University of
Reading

DEPARTMENT OF MATHEMATICS AND STATISTICS

Thesis submitted for the degree of

Doctor of Philosophy

February 2016

Abstract

Data assimilation is used to update the state of a system. Four-dimensional variational data assimilation (4DVar) updates an a priori estimate of a state by making use of observational data throughout a time window. 4DVar requires the use of model equations to describe the evolution of the state up to each observation time. The strong constraint formulation of 4DVar is commonly used within the meteorological community, which assumes the model equations perfectly describe the true evolution of the model state. However, models are representations of true dynamical systems and often contain errors, where the statistics of these errors are often unknown. A current objective of operational weather centres is to better account for model errors within the data assimilation process, both for the purpose of improving the accuracy of atmospheric and oceanic forecasts and for the purpose of more accurately reanalysing historical weather events. This gives us the motivation for the work in this thesis.

We develop diagnostics to verify the specification of an estimated model error covariance matrix. Once verified the matrix can subsequently be used in the weak formulation of 4DVar. However, often little is known about the model error statistics and we identify the need for model error to be accounted for within the data assimilation process without the requirement for the explicit specification of the model error statistics. Therefore, we next develop a combined model error and observation error covariance matrix, to replace the observation error covariance matrix, within the strong constraint 4DVar framework. We formulate a method to estimate the combined error statistics which does not require the explicit specification of model error statistics. The aim of this 4DVar method is to achieve a more accurate estimation of the initial model state when model error is present. Lastly, we next develop a coupled 4DVar technique that simultaneously estimates coupling parameters along with a coupled atmosphere-ocean model state. We show using an idealized coupled model that this method can be successful in compensating for model errors and hence improve the coupled model forecast accuracy.

Declaration

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

Katherine Howes

Acknowledgements

Firstly I would like to thank my supervisors Dr. Amos Lawless and Dr. Alison Fowler for the continuous support and encouragement throughout this PhD. I am very grateful for the opportunity that they have given me and for all the useful discussions we have had. I am extremely grateful for the funding that I have received from the Natural Environmental Research Council (NERC) that made this work possible.

I would also like to thank other members of staff in the Mathematics department that have made my time studying for my PhD an enjoyable experience, including Mrs Peta-Ann King who went above and beyond her role to help out in whatever way possible. Many thanks to the fantastic group of friends I have made at the university, the self-proclaimed 'lovely people', whose energy and enthusiasm together with the odd darts break has made the office environment a great place to be. Thanks also to the fabulous friends and neighbours I have outside the mathematics world for all your support and encouragement.

Thanks to my mum for always supporting me in following my dreams and always being my rock. Thanks to my dad, firstly for passing on the Maths gene, secondly for your patience in teaching me how to improve my grammar, but most of all for bringing me up with a 'work hard, play hard' ideology which without I would have not had all the fantastic experiences I have had in my life so far, including doing this PhD. Thanks to my gorgeous little sisters Jo and Em, that are also my best friends, for all your support and love. I am eternally grateful to my husband to be Dan, who always sees the potential in me and encourages me in life to pursue my dreams.

Contents

9 Conclusions

List of Figures

xiv

List of Tables

Acronyms

ECMWF	European Centre for Medium-Range Weather Forecasts
EKF	Extended Kalman Filter
ENSO	El Niño and Southern Oscillation
IFS	Integrated Forecast System
FOAM	Forecast Ocean Assimilation Model
KF	Kalman Filter
NEMO	Nucleus for European Modelling of the Ocean
NWP	Numerical Weather Prediction
ODE	Ordinary Differential Equation
OSTIA	Operational Sea Surface Temperature and Sea Ice Analysis
PDE	Partial Differential Equation
PDF	Probability Density Function
RMSE	Root Mean Square Error
SOAR	Second-order Auto-regressive

Acronyms continued

SSH	Sea Surface Height
SST	Sea Surface Temperature
Var	Variational data assimilation
3D−FGAT	3D-First Guess at Appropriate Time
3DVar	Three Dimensional Variational data assimilation
4DVar	Four Dimensional Variational data assimilation

Notation

A   Analysis error covariance matrix

B   Background error covariance matrix

$\widetilde{\mathbf{B}}$ Estimated background error covariance matrix

C   Correlation matrix

D   Covariance matrix of the errors in the model error analysis

$\mathbf{d}^a{}_b$ Difference between analysis and background in observation space

$\mathbf{d}^o{}_a$ Difference between observations and analysis in observation space

$\mathbf{d}^o{}_b$ Innovation vector

$E[\cdot]$ Expectation operator

F   Background coupling parameter error covariance matrix

H   Linearised observation operator

$\mathcal{H}$ Nonlinear observation operator

I   Identity matrix

$\mathcal{J}$ Cost function

Notation continued

$\mathcal{J}_b$ Background term of the cost function

$\mathcal{J}_{ob}$ Observation term of the cost function

$\mathbf{K}$ Gain matrix

$\mathcal{L}$ Correlation length scale

$\mathbf{M}$ Linearised model operator

$\mathcal{M}$ Nonlinear model operator

$\mathbf{M}^e$ Erroneous linear model matrix

$\mathcal{M}^e$ Erroneous nonlinear model operator

$\mathbf{M}^T$ Adjoint model operator

$\mathbf{Q}$ Model error covariance matrix

$\widetilde{\mathbf{Q}}$ Estimated model error covariance matrix

$\mathbf{R}$ Observation error covariance matrix

$\widetilde{\mathbf{R}}$ Estimated observation error covariance matrix

$\mathbf{R}^*$ Combined model error and observation error covariance matrix

$\widetilde{\mathbf{R}}^*$ Estimated combined model error and observation error covariance matrix

Notation continued

$\mathbf{x}$	Model state vector
$\mathbf{x}^a$	Analysis model state vector
$\mathbf{x}^b$	Background model state vector
$\mathbf{x}^t$	True model state vector
$\mathbf{y}$	Vector of observations
$\boldsymbol{\alpha}^b$	Vector of background coupling parameters
$\boldsymbol{\alpha}^t$	Vector of true coupling parameters
Δt	Model time-step
Δx	Model spatial-step
$\boldsymbol{\epsilon}_b$	Background error vector
$\boldsymbol{\epsilon}_{ob}$	Observation error vector
$\boldsymbol{\epsilon}_{ob}^*$	Combined model error and observation error vector
$\boldsymbol{\epsilon}_\alpha$	Vector of errors in background coupling parameters
$\boldsymbol{\eta}$	Model error vector
$\boldsymbol{\eta}^a$	Analysis of model error vector

Notation continued

ρ SOAR function

${\sigma_a}^2$ Analysis error variance

$\|\cdot\|$ L_2 norm

Chapter 1

Introduction

Data assimilation is used to best estimate the state of a dynamical system. Data assimilation methods require *observations* and an *a priori* estimate of the state, often known as the *background* state. Data assimilation is used within the meteorological community, for example to estimate the conditions for initialising forecasts of the atmosphere and ocean [99] [6]. However, the mathematical formulations of data assimilation methods are non-specific to a particular scientific field and other applications include image processing and oil reservoir modelling [14] [100]. Data assimilation methods account for the fact that both background states and observations contain errors. The best estimate of the state, known in the data assimilation community as the *analysis*, is obtained by weighting the background and observations according to the confidence in the accuracy of the information provided.

A commonly used data assimilation method is *four-dimensional variational* data assimilation (4DVar), where the four-dimensions consist of the three spatial dimensions and time. This method aims to best estimate the state of a system at the start of a specified time window, known as the *assimilation window*. The process compares both an estimate of the initial state with the background state, and a model evolved

estimated initial state with observations at specific times throughout the assimilation window. The *strong constraint* formulation of 4DVar assumes that the model used in the process perfectly describes the true dynamics of the system.

In this thesis we acknowledge that models are representations of true dynamical systems and often contain error. In particular, we focus on the errors present in models describing the dynamics of the atmosphere and ocean, where the statistics of the model errors present are often unknown. Strong constraint 4DVar is a method commonly used at operational meteorological centres, for example to initialise Numerical Weather Prediction (NWP) forecasts [99] [98]. A current objective of operational weather centres is how to better account for model error within the data assimilation process, both for the purpose of improving the accuracy of atmospheric and oceanic forecasts and for the purpose of more accurately reanalysing historical weather events. This gives us the motivation for work in this thesis, which will begin with the review of variational data assimilation methods developed to account for errors present in models, the types of errors present in operational atmosphere and ocean models and the current data assimilation processes used at operational centres. We will next detail the aims of work in this thesis.

1.1   Aims

Work in this thesis focuses on how to deal with model error within the data assimilation process. Model error statistics are often unknown and this prevents use of generic data assimilation methods developed to account for model error, such as the weak formulation of 4DVar. Specifically, work in this thesis aims to:

1. Develop a method to verify the specification of model error statistics. Subsequently investigate the potential use of such a method to refine estimated model

error statistics.

2. Amend the strong constraint 4DVar method in such a way that the effect model error has on the estimation of the initial conditions is mitigated. This developed data assimilation method should account for model errors without the requirement for explicit specification of the model error statistics.

3. Develop a strong constraint 4DVar method with the specific objective of improving the accuracy of a coupled atmosphere-ocean forecast.

1.2   Summary of new results

A summary of the new results in the thesis are as follows:

1. The derivation of diagnostic equations that account for errors of a random nature present in a model. These equations can be used as quality checks for the specification of model error statistics. The first set of diagnostic tools derived are for use with the strong constraint formulation of 4DVar, whereas the second set of diagnostic tools are for use with the weak constraint formulation of 4DVar. We show how, under certain conditions, the strong constraint 4DVar diagnostic tools have the potential to refine estimated model error statistics.

2. The development of a combined model error and observation error covariance matrix. This matrix accounts for the error statistics in the comparison of observations with a model evolved state mapped to observation space, where the model error is of a random nature at each time-step. When these combined error statistics replace the observation error statistics in the strong constraint formulation of 4DVar, a statistically more accurate estimate of the initial state is obtained. Further to this, the development of a method to estimate the

combined model error and observation error covariance matrix, which does not require explicit specification of the model error statistics.

3. The formulation of a strong constraint 4DVar method, to improve an erroneous coupled model forecast, by simultaneously estimating coupling parameters along with the atmosphere and ocean model state initial conditions. Estimation of a coupling parameter can compensate for both model bias and time invariant atmosphere and ocean parameter errors present in an idealized coupled model and hence improve the accuracy of the coupled forecast.

1.3   Thesis Outline

In Chapter 2, we introduce variational data assimilation (Var) methods used to best estimate the state of a system. We firstly describe Var methods developed for use with a perfect model and subsequently describe Var methods that remove the perfect model assumption. We recognise that the Var methods that account for errors present in a model often require the specification of model error statistics. Next in Chapter 2, we give an introduction to the use of diagnostic tools as quality checks for the specification of the background and observation error statistics required in data assimilation schemes [38]. These equations have not been formulated with model evolution, hence when used with model evolution do not account for model errors present [2]. Subsequently, we present one diagnostic equation that has been derived with model evolution and accounts for the presence of random error in a model [33] [2].

In Chapter 3, we describe the formation of coupled atmosphere-ocean models and acknowledge that these models are best representations of the true coupled dynamics and contain error [104] [54] [34]. We discuss the potential origins of the model errors

and the subsequent effects they have on the model state evolution, which can be of a random or systematic nature. Next in Chapter 3, we outline the current data assimilation methods used at operational meteorological centres for both; the purpose of estimating initial conditions to run forecasts and for reanalysis of past weather events. We also discuss the operational use of diagnostic tools as quality checks for the specification of background and observation error statistics and highlight that caution is advised if these tools are used with a model of an erroneous nature [2].

In Chapter 4, we firstly discuss the current difficulties that operational NWP centres have in specifying model error statistics. We then describe methods that have been developed to estimate model error statistics [118] [24] [30] [116] and report the problems that arose when using them. Coupled atmosphere-ocean models are used at operational meteorological centres to produce seasonal to decadal forecasts. In Chapter 4, we examine data assimilation methods that have been formulated to compensate for bias in coupled atmosphere-ocean models and hence improve coupled model forecasts [130] [131] [114] [83]. We highlight the potential for developing such methods in a 4DVar context. Lastly in Chapter 4, we outline techniques designed to account for representativity error present in the data assimilation process [56] [79] and discuss the relationship between representativity error and model error. We acknowledge that these techniques have the potential to be further developed to account for model error.

In Chapter 5, we introduce two dynamical systems that will be used to demonstrate methods developed in this thesis. The first of these is the linear advection equation which can be used to represent the transportation of a passive tracer in the atmosphere, for example water vapour carried along by a constant light breeze. The second of the dynamical systems is an idealized coupled atmosphere-ocean model [87]. Both of these dynamical systems consist of governing differential equations, for which we will use numerical schemes to provide approximate time-stepping solutions. In

5

Chapter 5, to avoid repetition, we outline certain properties of the data assimilation set up which will be used in numerical experiments throughout this thesis.

In Chapter 6, we develop the first of our results, which are two sets of diagnostic equations that account for random error present in a model. These equations can be used as quality checks, in observation space, for the specification of a model error covariance matrix with both background and observation error covariance matrices. The first set of diagnostic tools we derive consists of three equations and are for the case where an erroneous model is used in strong constraint 4DVar. The second set of diagnostic tools consists of four equations and are specifically for the case where an erroneous model is used in the weak formulation of 4DVar. Work in Chapter 6 also involves investigation of when the diagnostic tools can be used to refine an inaccurately estimated model error covariance matrix.

In Chapter 7, we develop a combined model error and observation error covariance matrix. Observations contain error and we consider a model of an erroneous nature with random error present at each time-step. This combined error matrix accounts for the errors in the comparison of observations with a model evolved initial state mapped to observation space. The replacement of the observation error covariance matrix with the combined error covariance matrix ensures the strong constraint 4DVar problem is formulated to be mathematically correct, in order to best estimate the initial state with use of an erroneous model of this nature. The formula we derive for the combined error statistics requires the specification of model error statistics, which are often unknown. Therefore, work in Chapter 7 involves developing a method, with use of diagnostic tools, to estimate the combined model error and observation error matrix without the need for explicit specification of the model error statistics. We subsequently demonstrate the successful application of our developed 4DVar method, with estimated combined statistics, to produce an improved analysis.

In Chapter 8, we develop a data assimilation method to account for errors present

6

in a coupled atmosphere-ocean model and hence improve the accuracy of the coupled forecast. We extend the idea of coupled atmosphere-ocean model state estimation in the 4DVar framework to also estimating the models coupling parameters, with the aim that the coupling parameter estimation compensates for the errors in the model. We demonstrate the success of this method, in compensating for both model bias and static atmosphere and ocean parameter errors present in an idealized coupled model, and hence in producing forecasts of a higher accuracy. We investigate the requirements for the developed scheme to be successful in compensating for the errors present in the coupled model and the conditions under which this method provides significant improvements to coupled forecasts.

Finally in Chapter 9, we summarise the work conducted in this thesis and draw conclusions. We present suggestions for further work that could be carried out following on from work conducted to date.

Chapter 2

Data Assimilation

In this chapter we introduce variational data assimilation methods (Var), which aim to best estimate the state of a system. Practical applications of Var include both atmospheric and oceanic forecasting. We acknowledge in this thesis that models are representations of true dynamical systems and often contain error. In Section 2.1 we introduce general nonlinear perfect model equations and subsequently general nonlinear erroneous model equations. In Section 2.2 we describe Var methods formulated for use firstly with a perfect model and secondly with an erroneous model. We shall then give a brief introduction to diagnostic tools in Section 2.3 which can be useful quality checks for the specification of statistics required in the data assimilation process.

2.1   Model system equations

In this section we introduce the notation and terminology for discrete dynamical models and model error that we will use throughout this thesis. The notation and terminology we use is commonly used in the data assimilation community. However,

the meaning of the term 'error' can be interpreted in various ways and therefore we will firstly discuss our interpretation of a 'perfect model' and an 'erroneous model'.

When considering dynamics of a 'true' system, such as the atmosphere and ocean, a forecast model is always going to be an approximation of the truth. This is because models can only represent a finite number of variables at a finite number of spacial points and only be evolved over a finite number of temporal points. Therefore even if the 'true' dynamics of a system are well represented at these finite points with a forecast model, this does not represent the whole dynamical system. Therefore what we consider a 'perfect model' to be is a 'deterministic model' that evolves a finite number of model state variables at finite points in space, over a finite number of time-steps. Further to this, if the variables input into the 'perfect model' match the 'true' state values, then the outputs from the 'perfect model' produce values that match the 'true' state at future points in time. However, often this is not the case, and when this is not the case we describe the forecast models as 'erroneous models'. We can consider that the error is 'stochastic', where either the 'erroneous' model is a 'stochastic model' or the 'truth' is stochastic from the perspective of the forecast model. In these cases the 'error' is of a random nature. We can also assume that the 'erroneous' model differs from the 'perfect model' with systematic correlations in time.

2.1.1   Perfect model

We consider the discrete nonlinear dynamical model equations of the form,

$$\mathbf{x}_i = \mathcal{M}_{\{i-1\}\to i}(\mathbf{x}_{i-1}) \qquad i = 1, 2, ... \tag{2.1}$$

to be perfect, where the column vector $\mathbf{x}_i \in \mathbb{R}^m$ is known as the state vector and contains each of the model variables at time t_i. The model operator $\mathcal{M}_{\{i-1\}\to i} : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ defines the true evolution of the state vector from the previous time t_{i-1} to time t_i.

2.1.2 Erroneous model

With error present in the model operator, we consider model equations of the form,

$$\mathbf{x}_i = \mathcal{M}^e_{\{i-1\}\to i}(\mathbf{x}_{i-1}) \qquad i = 1, 2, ... \tag{2.2}$$

where the model operator $\mathcal{M}^e_{\{i-1\}\to i} : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ is the best known description of the evolution of the state vector from the previous time t_{i-1} to time t_i. In this thesis we assume that to acquire the true model state $\mathbf{x}^t_i$ at time t_i,

$$\mathbf{x}^t_i = \mathcal{M}^e_{\{i-1\}\to i}(\mathbf{x}^t_{i-1}) + \boldsymbol{\eta}_i \qquad i = 1, 2, ... \tag{2.3}$$

where $\boldsymbol{\eta}_i \in \mathbb{R}^m$ is a column vector containing the effect of 'model errors' at time t_i. We are not in fact stating that all errors in a generic model are of an additive nature, but instead stating that the errors in the model can be 'corrected' for by addition of a vector $\boldsymbol{\eta}_i$ at each time-step.

2.1.3 Glossary of Model Error

This thesis will consider the two cases;

- **Random model error**: A 'stochastic model' where $\boldsymbol{\eta}_i$ (2.3) contains entries that are random Gaussian distributed uncorrelated in time with a zero mean and covariance matrix $\mathbf{Q}_i$.

- **Systematic model error**: The vector $\boldsymbol{\eta}_i$ (2.3) is considered to be of a systematic nature that can have a non-zero mean and allow for correlations in time.

  - Types of systematic errors considered in this thesis are model bias and static parameter errors.

2.2 Data assimilation methods

Data assimilation aims to improve the accuracy of a prior estimate of a state, known as the background. The data assimilation process uses observational data and often a model, to provide a best estimate of the state, known as the analysis. There are a range of data assimilation techniques. We will concentrate on variational data assimilation methods (Var) within this thesis. The other main class of data assimilation techniques are known as sequential methods. These are direct methods, that use information either at one time only or information propagated forward in time, that explicitly provide an updated estimate of a state [77]. These sequential methods specifically include schemes such as Optimal Interpolation [50] and the Kalman Filter method [62]. Var differs from these sequential methods; the estimation of the state is sought iteratively through minimisation of a cost function [77]. Var specifically includes three-dimensional variational data assimilation (3DVar) where information is used at one time only, and four-dimensional variational data assimilation (4DVar) which uses information propagated both forward and backwards in time. Var is used widely within the scientific community, including at operational Numerical Weather

Prediction (NWP) centres including the Met Office [99], ECMWF [98], the Japanese Met Agency [61] and the Canadian Met Service [51]. This is because Var can provide estimates of a huge number of state variables at a computationally feasible cost within the time frames required for operational use [70].

We assume we have observations $\mathbf{y}_i$ at time t_i, which are related to the true model state by,

$$\mathbf{y}_i = \mathcal{H}_i(\mathbf{x}_i{}^t) + \boldsymbol{\epsilon}_{obi}, \qquad i = 0, 1, ... \qquad (2.4)$$

where the column vector $\mathbf{y}_i \in \mathbb{R}^{p_i}$ contains p_i observations at time t_i. The observation errors at time t_i are represented in the column vector $\boldsymbol{\epsilon}_{obi} \in \mathbb{R}^{p_i}$ and are assumed to be of a random nature. The nonlinear observation operator $\mathcal{H}_i : \mathbb{R}^m \longrightarrow \mathbb{R}^{p_i}$ maps the state from model space to observation space. In this thesis we assume we have the same number of observations in $\mathbf{y}_i$ at each time t_i and let $p_i = p \ \forall$ times t_i. The background $\mathbf{x}^b$ is related to the true state $\mathbf{x}^t{}_0$ at time t_0 such that,

$$\mathbf{x}^b = \mathbf{x}^t{}_0 + \boldsymbol{\epsilon}_b \qquad (2.5)$$

where the column vector $\mathbf{x}^b \in \mathbb{R}^m$ contains a prior estimate to each of the model state variables. The column vector $\boldsymbol{\epsilon}_b \in \mathbb{R}^m$ contains errors in the background model state and is assumed to be of a random nature.. Each observation error vector $\boldsymbol{\epsilon}_{obi}$ has a corresponding observation error covariance matrix $\mathbf{R}_i \in \mathbb{R}^{p \times p}$ describing the correlations and variances of the observation errors at time t_i. Similarly the background error vector $\boldsymbol{\epsilon}_b$ has a corresponding error covariance matrix $\mathbf{B} \in \mathbb{R}^{m \times m}$. We next specify assumptions widely used in the data assimilation community, which are fundamental

for use of variational techniques. Firstly, both the background errors and observation errors are unbiased Gaussian distributed and have corresponding symmetric positive definite error covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$ respectively and secondly, errors in the background are uncorrelated to errors present in observations.

2.2.1 3DVar

Three-dimensional variational data assimilation (3DVar) aims to best estimate the state of a system at a particular time. The three-dimensions in 3DVar simply represent the three spacial dimensions and therefore no model evolution through time is required. The approach is to improve a background state approximation with use of observations present at that time to acquire the best estimate of the model state, known as the analysis [122]. Current operational applications of such methods include ocean data assimilation at meteorological centres around the world [78] [132] [49] [59]. This method can be applied at any point in time where there is both a background state and observations available. For ease of comparison with other Var techniques we will use the notation t_0 for the particular point in time chosen. The use of Bayes Theorem in conjunction with the assumed Gaussian distributed data and uncorrelated errors enables an expression for the posterior probability density function (PDF) of the state given the observations [80], with the maximum likelihood at the analysis. This state with the maximum likelihood is equivalent to the state with the minimum variance and can be determined by minimising the cost function,

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}_0 - \mathcal{H}_0(\mathbf{x}_0))^T \mathbf{R}_0^{-1}(\mathbf{y}_0 - \mathcal{H}_0(\mathbf{x}_0)), \quad (2.6)$$

with respect to the state vector $\mathbf{x}_0$. The model state is compared both to the background and observations which are weighted according to the accuracy specified by their respective error covariance matrices. A gradient descent algorithm is an iterative

13

process often used to minimise (2.6) which requires the gradient of the cost function with respect to the model state,

$$\nabla \mathcal{J}(\mathbf{x}_0) \;=\; \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) - \mathbf{H}_0{}^T \mathbf{R}_0{}^{-1}(\mathbf{y}_0 - \mathcal{H}_0(\mathbf{x}_0)), \tag{2.7}$$

where $\mathbf{H}_0 \in \mathbb{R}^{p \times m}$ is the linearised observation operator. However, when a linear observation operator is used in the cost function (2.6), the analysis can be explicitly given by,

$$\mathbf{x}^a{}_0 \;=\; \mathbf{x}^b + \mathbf{K}\mathbf{d}^o_b, \tag{2.8}$$

where $\mathbf{K} = \mathbf{B}\mathbf{H}_0{}^T(\mathbf{H}_0\mathbf{B}\mathbf{H}_0{}^T + \mathbf{R}_0)^{-1}$ which is commonly referred to as the gain matrix [91] and $\mathbf{d}^o_b = \mathbf{y}_0 - \mathbf{H}_0\mathbf{x}^b$ which is known as the innovation vector [38].

3DVar is a useful tool to estimate conditions of a state vector at a particular time. However, in practice this scheme may not be the most appropriate variational data method to use in all cases, for example when there is a lack of observational data at the analysis time. This 3DVar method can be extended to include a fourth dimension, time, with the data assimilation technique known as four-dimensional variational data assimilation (4DVar).

2.2.2 4DVar

4DVar uses observations at multiple times across a specified window, known as the assimilation window. The objective of 4DVar is to best estimate the initial conditions at the start of the assimilation window, known as the analysis [76] [40] [28]. The process involves the comparison of the initial conditions model state vector with a background state and the comparison of the initial conditions evolved through time

using a model with the observations. 4DVar techniques are currently used at many NWP centres to find initial conditions of the atmosphere [99] [98]. 4DVar provides an affordable method to make use of vast amount of observational data to enhance the accuracy of the analysis, as opposed to sequential techniques such as the Kalman Filter (KF) which are of a higher computational expense [45]. The 4DVar technique aims to minimise a four-dimensional cost function with respect to the initial state vector $\mathbf{x}_0$,

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)), (2.9)$$

subject to satisfying the nonlinear model equations (2.1). This formulation of the minimisation problem assumes that the model (2.1) used in the assimilation is perfect. We pictorially represent the 4DVAR process in Figure 2.1, which clearly shows the aim to minimise the square of the difference between $\mathbf{x}_0$ and the background and also between the model trajectory of $\mathbf{x}_0$ and the observations. Running the model from the analysis past the last observation time can then be used as a model forecast.

Often an iterative gradient descent algorithm is used to search for the minimum of the cost function (2.9). To do so, the minimisation algorithms require the input of both the cost function and its gradient evaluated at the most recent approximation to the model state initial conditions. The total value of the cost function (2.9) can be calculated by running the forward nonlinear model (2.1) to acquire the model state values at all observation times in the window. However, to acquire the gradient of the cost function (2.9) with respect to the initial state $\mathbf{x}_0$, subject to the nonlinear model equations (2.1) requires a further process. Firstly, we split the cost function (2.9) into its background and observation parts respectively,
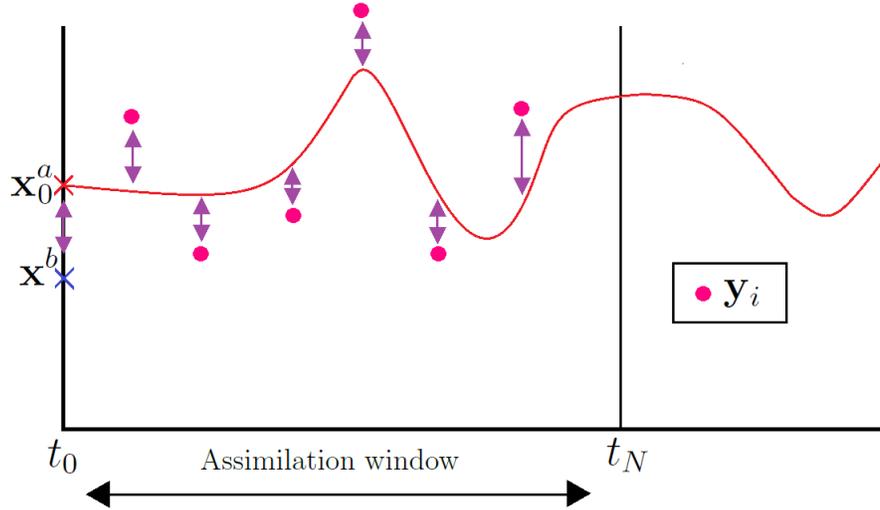
Figure 2.1: Pictorial representation of the 4DVAR minimisation process. The aim is to minimise both the difference (purple arrow) between the background state $\mathbf{x}^b$ (blue cross) and the initial state $\mathbf{x}_0$ and the differences (purple arrows) between the model trajectories of the initial state $\mathbf{x}_0$ and the observations $\mathbf{y}_i$ (pink circles). The initial state $\mathbf{x}_0$ that provides the total minimal square of the differences (purple arrows), subject to the weighting provided by the background error and observation error covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$ respectively, is the analysis $\mathbf{x}^a$ (red cross). The red line shown is the analysis trajectory which has been produced by running the model from the analysis.

$$\mathcal{J} = \mathcal{J}_b + \mathcal{J}_{ob}, \tag{2.10}$$

where $\mathcal{J}_b = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b)$ and $\mathcal{J}_{ob} = \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))$. The gradient of the background part of the cost function $\mathcal{J}_b$ is model independent and therefore can be explicitly evaluated,

$$\nabla \mathcal{J}_b = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b). \tag{2.11}$$

We now redefine the observational part of the cost function, which is subject to the model equations. We use a Lagrangian functional [65] which enables the problem to be unconstrained,

$$\mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i) = \mathcal{J}_{obi} + \boldsymbol{\lambda}_{i+1}^T (\mathbf{x}_{i+1} - \mathcal{M}_{i\rightarrow\{i+1\}}(\mathbf{x}_i)), \tag{2.12}$$

where $\boldsymbol{\lambda}_i \in \mathbb{R}^m$ is a column vector of Lagrange multipliers at time t_i, in this application known as the adjoint variables [65]. We take the derivative of the functional (2.12) and require this to be zero [40],

$$\frac{\partial \mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i)}{\partial \boldsymbol{\lambda}_i} = \mathbf{0}, \tag{2.13}$$

$$\frac{\partial \mathcal{L}(\mathbf{x}_i, \boldsymbol{\lambda}_i)}{\partial \mathbf{x}_i} = \mathbf{0}. \tag{2.14}$$

which are the conditions for an extremum of (2.12), where the first of the equations ensures the nonlinear model equations (2.1) are upheld and the latter gives us the set of adjoint equations,

$$\boldsymbol{\lambda}_i = \mathbf{M}_{i\rightarrow\{i+1\}}^T \boldsymbol{\lambda}_{i+1} + \mathbf{H}_i^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)), \tag{2.15}$$

where $\mathbf{M}_{i\rightarrow\{i+1\}} = \frac{\partial(\mathcal{M}_{i\rightarrow\{i+1\}})}{\partial \mathbf{x}_i}$ is the linearised model operator and $\mathbf{H}_i = \frac{\partial \mathcal{H}_i(\mathbf{x}_i)}{\partial \mathbf{x}_i}$ is the linearised observation operator. The term $\mathbf{M}_{i\rightarrow\{i+1\}}^T$ is the transpose of the linearised model $\mathbf{M}_{i\rightarrow\{i+1\}}$, commonly known in the data assimilation community as the adjoint model operator [65]. The adjoint equations (2.15) are run backwards in time with

the initial input $\boldsymbol{\lambda}_{N+1} = \mathbf{0}$ and resulting output $\boldsymbol{\lambda}_0$. The gradient of $\mathcal{J}_{ob}$ with respect to x_0 is given by [25],

$$\nabla \mathcal{J}_{ob} = -\boldsymbol{\lambda}_0. \tag{2.16}$$

Therefore, the total gradient of the cost function (2.9) with respect to the initial state $\mathbf{x}_0$ is,

$$
\begin{aligned}
\nabla \mathcal{J} &= \nabla \mathcal{J}_b + \nabla \mathcal{J}_{ob}, & (2.17) \\
&= \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) - \boldsymbol{\lambda}_0, & (2.18)
\end{aligned}
$$

where $\boldsymbol{\lambda}_0$ is calculated by running the adjoint equations (2.15) backwards in time.

The analysis can be explicitly calculated when the model equations (2.1) and observation operators (2.4) are linear. Let us assume we have observations at $N+1$ times throughout the assimilation window and rewrite the 4D cost function (2.9) with linear operators in the form,

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0)^T \hat{\mathbf{R}}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0), \tag{2.19}$$

where,

$$\hat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{y}_N \end{pmatrix} \qquad \hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_{0\rightarrow 1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{0\rightarrow N} \end{pmatrix}$$

and

$$\hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{R}_1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{R}_N \end{pmatrix}.$$

The minimum of the cost function (2.19) is evaluated by differentiating with respect to the initial model state $\mathbf{x}_0$ and setting equal to a vector of zeros,

$$\nabla \mathcal{J}(\mathbf{x}_0) \;\; = \;\; \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) - \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0) = \mathbf{0}, \qquad (2.20)$$

leading to the analysis,

$$\mathbf{x}^a{}_0 \;\; = \;\; \mathbf{x}^b + \hat{\mathbf{K}}\hat{\mathbf{d}}^o_b, \qquad (2.21)$$

with the gain matrix $\hat{\mathbf{K}} = \mathbf{B}\hat{\mathbf{H}}^T(\hat{\mathbf{H}}\mathbf{B}\hat{\mathbf{H}}^T + \hat{\mathbf{R}})^{-1}$ and the innovation vector $\hat{\mathbf{d}}^o_b = \hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}^b$. Either by using the equation for the analysis (2.21) or by calculating the Hessian (second derivative) of the cost function (2.19) and subsequently taking the inverse [119], it can be shown the analysis error covariance matrix is given by [28],

$$\mathbf{A} \;\; = \;\; (\mathbf{B}^{-1} + \hat{\mathbf{H}}^T\hat{\mathbf{R}}^{-1}\hat{\mathbf{H}})^{-1} = (\nabla^2\mathcal{J}(\mathbf{x}_0))^{-1}. \tag{2.22}$$

The analysis error covariance matrix provides statistical information on the accuracy of the model state analysis. Equation (2.22) shows that the statistical accuracy of the model state analysis (given by $\mathbf{A}$) has improved from the statistical accuracy of the background model state (given by $\mathbf{B}$) and the significance of this improvement is dependent on the accuracy of the observations (specified in $\hat{\mathbf{R}}$).

2.2.3 Incremental 4DVar

The incremental formulation of 4DVar replaces the minimisation of the strong constraint 4DVar cost function (2.9) subject to the nonlinear model equations (2.1) with a process of multiple minimisations of the following cost function,

$$
\begin{aligned}
\mathcal{J}(\delta\mathbf{x}_0{}^{(k)}) \;\; = \;\; & \frac{1}{2}(\delta\mathbf{x}_0{}^{(k)} - (\mathbf{x}^b - \mathbf{x}_0{}^{(k)}))^T\mathbf{B}^{-1}(\delta\mathbf{x}_0{}^{(k)} - (\mathbf{x}^b - \mathbf{x}_0{}^{(k)})) \\
& + \; \frac{1}{2}\sum_{i=0}^{N}((\mathbf{H}_i\delta\mathbf{x}_i{}^{(k)} - \mathbf{d}_{b\,i}^{o\,(k)})\mathbf{R}^{-1}(\mathbf{H}_i\delta\mathbf{x}_i{}^{(k)} - \mathbf{d}_{b\,i}^{o\,(k)}), \tag{2.23}
\end{aligned}
$$

where $\mathbf{d}_{b\,i}^{o\,(k)} = \mathbf{y}_i - \mathcal{H}(\mathbf{x}_i{}^{(k)})$, subject to the linearised model $\delta\mathbf{x}_i = \mathbf{M}_i\delta\mathbf{x}_{i-1}$. This linearised model operator $\mathbf{M}_i$ which evolves the model state increments $\delta\mathbf{x}_{i-1}$ at time t_{i-1} to time t_i, uses a lower resolution spatial grid and further simplifications in the parameterisations than the full nonlinear model [60]. Operational NWP centres, such as the Met Office and ECWMF, use the incremental version of strong constraint 4DVar method to estimate the initial conditions of the atmospheric model state vector. The incremental version of 4DVar requires less computational resource and time than

the standard 4DVar method, due to the low resolution inner loop, and therefore is feasible for use at NWP centres [28]. The incremental 4DVar technique involves calculations in both an 'outer loop' and 'inner loop' as the flow chart describing the process shows in Figure 2.2. In the outer loop; the nonlinear model is used to evolve the current estimate of the initial model state up to the observation times and then the nonlinear observation operator is used to evaluate the innovations. In the inner loop; the linearised model is used to evolve the model state increment and the linearised observation operator is applied to the model evolved increment. The number of outer and inner loops conducted can be predefined, however often the inner-loop minimisation will halt when a stopping criteria is reached, from which the analysis of the model state $\mathbf{x}^a_0$ is produced. An example of a stopping criterion is that the gradient of the cost function at the current iteration has reduced by a set tolerance in comparison to its value on the first iteration [72]. The incremental method treats the minimisation problem as a sequence of quadratic problems, as can be seen in the pictorial representation of incremental 4DVAR in Figure 2.3, where the inner loops can be efficiently minimised using the gradient descent algorithms. The minimisation of the incremental 4DVar cost functions will converge to the solution of the original 4DVar problem, as long as the tangent linear approximations hold [28] and each loop is solved to sufficient accuracy [71]. This condition enables us to use the original 4DVar formulation as described in Section 2.2.2 in methods we develop in this thesis.

The 4DVar methods described in this section have been formulated upon the basis that the perfect model equations are known. The cost function is constrained by these perfect model equations and the method is known as strong constraint 4DVar. Next, we will review data assimilation methods that remove the perfect model assumption and consider how the model error should be dealt with.
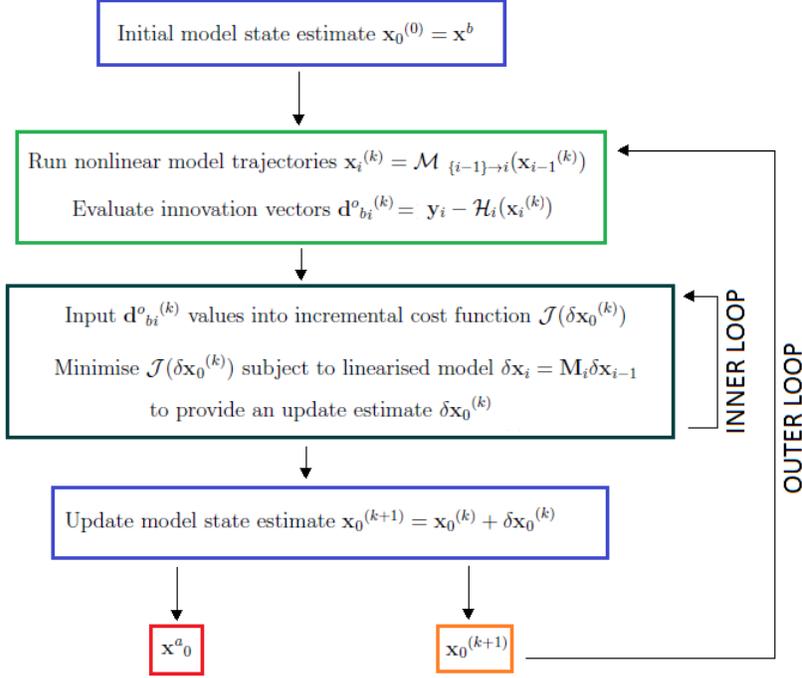
Figure 2.2: Flow chart representation of the incremental 4DVAR minimisation. The initial estimate is chosen to be the background model state. The nonlinear model trajectories are run and the innovations calculated. The inner loop consists of the minimisation of the incremental cost function subject to the linearised model, to estimate the incremental model state update. This incremental model state update is then added to the estimate of the model state. This process is repeated until the specified number of inner and outer loops have been run, with the analysis then output from the process.

2.2.4   Weak constraint 4DVar

We will take the situation that the best known model describing the evolution of the model state is erroneous and of the form (2.2). As it is known that these model equations contain error, we can let them be satisfied approximately and become weak constraints in the data assimilation problem [103] [117]. The general formulation of weak constraint 4DVar assumes an unknown vector of additive model error at each time t_i denoted by $\boldsymbol{\eta}_i$ (2.3). Weak constraint 4DVar can be formulated upon the assumption that each model error vector $\boldsymbol{\eta}_i$ at time t_i is not correlated with any other
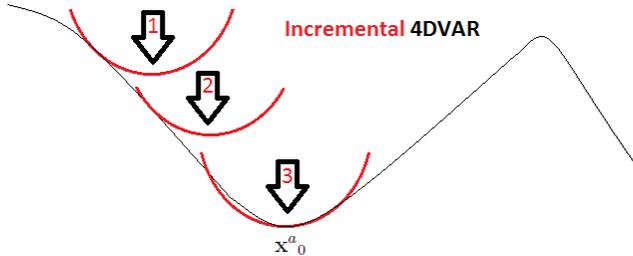
Figure 2.3: Pictorial representation of the incremental 4DVAR minimisation technique. Each inner loop procedure consists of minimising a quadratic problem in order to best estimate the model state increment. This figure illustrates a fictional case where three outer loops have been used (red curves) to best estimate the minimum of a nonlinear cost function (black curve).

model error vectors at any other times and is of a random nature at each time-step. The distribution of the model errors $\boldsymbol{\eta}_i$ is assumed to be Gaussian around a zero mean and has a corresponding error covariance matrix $\mathbf{Q}_i$ at each time t_i. The model errors are considered to be uncorrelated with both the errors in the background and errors in the observations. There are two formulations of weak constraint 4DVar [117]; the objective of the first formulation is to estimate both the initial conditions of the model state and the model error vectors at each time-step in the assimilation window, and the objective of the second formulation is to estimate the model state at each time-step throughout the assimilation window. The two formulations are mathematically similar, however the sensitivities of the cost functions to small changes in the input data differ [44]. We will describe the first of these formulations of weak constraint 4DVar, as this is the formulation most often considered [117]. The process again involves the comparison of the initial conditions with a background state and the comparison of the initial conditions evolved through time using a model with the observations. However, when using this data assimilation technique, the model used within the cost function allows for model error at each time step,

$$\mathbf{x}_i = \mathcal{M}^e{}_{\{i-1\}\to i}(\mathbf{x}_{i-1}) + \boldsymbol{\eta}_i \qquad i = 1, 2, ... \qquad (2.24)$$
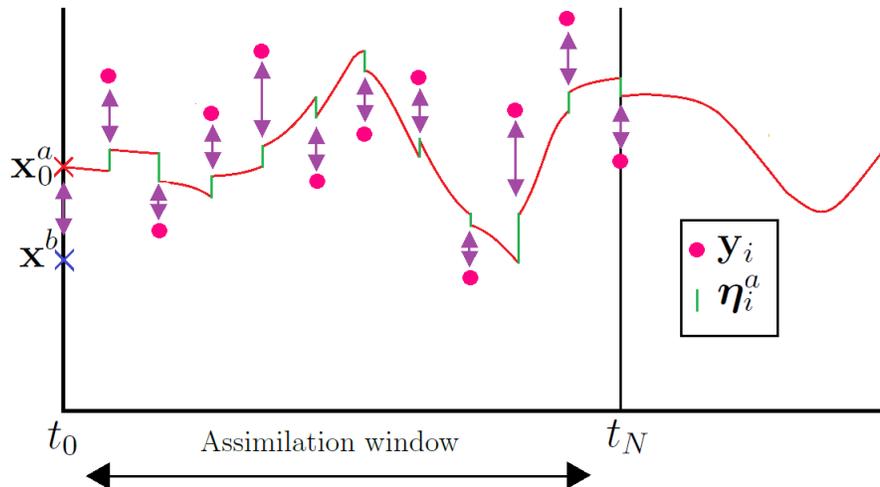
Figure 2.4:   Pictorial representation of the weak constraint 4DVAR minimisation process. The aim is to minimise both the difference (purple arrow) between the background state $\mathbf{x}^b$ (blue cross) and the initial state $\mathbf{x}_0$ and the differences (purple arrows) between the model trajectories of the initial state $\mathbf{x}_0$ and the observations $\mathbf{y}_i$ (pink circles). The estimation of the model errors $\boldsymbol{\eta}_i$ enables the relaxed model equations to better fit the observations. The initial state $\mathbf{x}_0$ and model errors $\boldsymbol{\eta}_i$ that provide the total minimal differences (purple arrows) and model errors (green lines), when subject to the weighting provided by the background error, observation error and model error covariance matrices $\mathbf{B}$, $\mathbf{R}_i$ and $\mathbf{Q}_i$ respectively, provide the analysis $\mathbf{x}^a{}_0$ and $\hat{\boldsymbol{\eta}}^a$

A pictorial representation of weak constraint 4DVar is shown in Figure 2.4. Let us define the column vector containing the model errors at all time-steps in the

24

assimilation window by,

$$\hat{\boldsymbol{\eta}} \in \mathbb{R}^{Nm} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \vdots \\ \vdots \\ \boldsymbol{\eta}_N \end{pmatrix}.$$

This method aims to minimise the weak constraint 4DVar cost function with respect to both the initial state vector $\mathbf{x}_0$ and to the model errors $\hat{\boldsymbol{\eta}}$,

$$
\begin{aligned}
\mathcal{J}(\mathbf{x}_0, \hat{\boldsymbol{\eta}}) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2} \sum_{i=0}^{N} (\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)) \\
&+ \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{\eta}_i{}^T \mathbf{Q_i}^{-1} \boldsymbol{\eta}_i,
\end{aligned}
\tag{2.25}
$$

which is subject to the relaxed model equations (2.24). The output from this formulation of weak constraint 4DVar is both the analysis of the model state initial conditions $\mathbf{x}^a{}_0$ and the analysis of the model errors at each time-step throughout the assimilation window $\hat{\boldsymbol{\eta}}^a$. Hence the minimisation of the cost function (2.25) requires differentiation both with respect to $\mathbf{x}_0$ and $\hat{\boldsymbol{\eta}}$. An iterative gradient descent algorithm is required to attain the analysis, where modified adjoint equations are used to evaluate the gradient of the cost function. A key point that should be highlighted is that for use of this weak constraint 4DVar technique the model error covariance matrix $\mathbf{Q}_i$ is to be specified for each time t_i in the assimilation window. Little is often known about the model error statistics and therefore in some cases this method is impractical for operational use. There is also a significant increase in the number of variables estimated in weak constraint 4DVar as opposed to strong constraint 4DVar, specifi-

cally $(N+1)m$ instead of m. Therefore, even if the model error covariance matrix $\mathbf{Q}_i$ is known, this is a computationally expensive problem to solve. Operational weather centres, such as ECMWF [118], have conducted investigations into whether the use of weak constraint 4DVar could improve the quality of the atmospheric analysis and forecast. These investigations involved assumptions such as a static $\mathbf{Q}$ and this is further discussed in Chapter 4.

The model error considered in this section has been of a random nature at each time-step. Next we describe a data assimilation technique that can account for model error of a systematic nature, correlated in time.

2.2.5 Systematic error correction scheme

We now investigate a method that compensates for error of a systematic nature in a model by including a correction term [36]. We denote this correction term $\boldsymbol{\eta}_i$ (2.3) and introduce notation to describe the evolution of the error through time [54] [92],

$$
\begin{aligned}
\boldsymbol{\eta}_i &= \mathcal{T}_i(\mathbf{e}_i), \\
\mathbf{e}_{i+1} &= \mathcal{G}_{i\to\{i+1\}}(\mathbf{x}_i, \mathbf{e}_i),
\end{aligned}
\tag{2.26}
$$

where $\mathbf{e}_i \in \mathbb{R}^r$ is a column vector that contains systematic model error components, the operator $\mathcal{T}_i : \mathbb{R}^r \longrightarrow \mathbb{R}^m$ distributes the model error components to the appropriate model equations. The evolution of the model error components from time t_i to time t_{i+1} is defined using the function $\mathcal{G}_{i\to\{i+1\}} : \mathbb{R}^m \times \mathbb{R}^r \longrightarrow \mathbb{R}^r$, which allows for the model error to be state dependent. Examples of model errors that can be represented using this method include; a constant model bias and model error that fluctuates smoothly on a time-scale such as a daily cycle [54]. The model equations,

$$\mathbf{x}_i = \mathcal{M}^e{}_{\{i-1\}\rightarrow i}(\mathbf{x}_{i-1}) + \boldsymbol{\eta}_i \qquad i = 1, 2, ... \tag{2.27}$$

together with the model error equations (2.26) are known as the augmented state system model. The aim of this data assimilation technique is to solve the augmented problem, to best estimate both the initial conditions of the model state and the model error components at the initial time. It is assumed a prior estimate of the model error components $\mathbf{e}^b$ is known along with a combined background model state error and background model error components covariance matrix $\mathbf{W} \in \mathbb{R}^{(m+r)\times(m+r)}$ which allows for cross correlations between the errors in the background model state and background model error components. The observation errors are assumed to be uncorrelated to both the model state background and model error background.

The process again involves the comparison of the initial conditions with a background model state and the comparison of the model evolved initial conditions with observations. However, when using this data assimilation technique, the model used within the cost function allows for systematic model error at each time step and the initial conditions of the model error components $\mathbf{e}_0$ are estimated along with $\mathbf{x}_0$. The cost function for this augmented data assimilation approach is as follows,

$$
\begin{aligned}
\mathcal{J}(\mathbf{x}_0, \mathbf{e}_0) \;=\; & \frac{1}{2}\begin{pmatrix} \mathbf{x}_0 - \mathbf{x}^b \\ \mathbf{e}_0 - \mathbf{e}^b \end{pmatrix}^T \mathbf{W}^{-1} \begin{pmatrix} \mathbf{x}_0 - \mathbf{x}^b \\ \mathbf{e}_0 - \mathbf{e}^b \end{pmatrix} \\
& + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)),
\end{aligned}
\tag{2.28}
$$

which is to be minimised, subject to the augmented state system model (2.26) and (2.27), to acquire the analysis of both $\mathbf{x}^a{}_0$ and $\mathbf{e}^a{}_0$. The minimisation of the cost

function (2.28) requires differentiation with respect to both $\mathbf{x}_0$ and $\mathbf{e}_0$. An iterative gradient descent algorithm is often used in the minimisation process, where modified adjoint equations are required to evaluate the gradient of the cost function.

Often little is known about the errors present in a model. A key point that should be highlighted is that for use of this scheme: the evolution of the model error and the background model error components $\mathbf{e}^b$ along with the corresponding model error statistics need to be specified. We will describe current operational use of this scheme in the next chapter.

2.3   Consistency diagnostics

Data assimilation methods require the specification of both background error and observation error covariance matrices. These statistics are often extremely hard to specify, due to lack of knowledge about the both the background errors and observation errors and due to the huge size the matrices can be [22]. The idea of studying innovation statistics between forecasts and observations has been around for many years [57]. These have been further developed into consistency checks for the specification of both the background error and observation error covariance matrices, known in the data assimilation community as the Desroziers diagnostics [38]. They can be thought of as quality control checks for approximated background error and observation error statistics. These diagnostics consist of combinations of background model states, observations and model state analysis, all in observation space. Hence, all information used in these diagnostics is readily available after an assimilation run. Next, we present diagnostics tools similar to those derived by Desroziers et al. [38]. The actual Desroziers diagnostics [38] do not include a model matrix, however we include model evolution within the diagnostic equations, which can be verified with the Desroziers diagnostics [38] when all observations are at the start of the time window.

The first of the diagnostic equations we present was previously stated by Andersson [2].

2.3.1 Consistency diagnostics with a perfect model

We consider a vector of observations $\mathbf{y}_1$, of the form (2.4), available at time t_1 with a corresponding observation error covariance matrix $\mathbf{R}$. The nonlinear observation operator $\mathcal{H}$ takes the model state from state space to observation space. We assume we have a background model state $\mathbf{x}^b$, of the form (2.5), with a corresponding background error covariance matrix $\mathbf{B}$. The perfect nonlinear model operator $\mathcal{M}_{0\rightarrow1}$ evolves the initial state to the observation time as follows,

$$\mathbf{x}_1 = \mathcal{M}_{0\rightarrow1}(\mathbf{x}_0). \tag{2.29}$$

We simplify the model notation, as we are only considering the model evolution over one time-step, by letting $\mathcal{M}_{0\rightarrow1} = \mathcal{M}$. We assume the tangent linear hypothesis holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}$, which states that; the first-order Taylor expansion of a nonlinear model is sufficient to describe the behaviour of the model for small perturbations along the nonlinear solution trajectories [17]. The linear observation operator $\mathbf{H}$ is the first order term in the expansion of the Taylor series of $\mathcal{H}(\mathbf{x} + \delta\mathbf{x})$ and the tangent linear model $\mathbf{M}$ is the first order term in the expansion of the Taylor series of $\mathcal{M}(\mathbf{x} + \delta\mathbf{x})$. This allows the evaluation of the explicit incremental 4DVar analysis with one 'outer-loop',

$$\mathbf{x}^a{}_0 \;\; = \;\; \mathbf{x}^b + \mathbf{K}\mathbf{d}^o_b, \tag{2.30}$$

29

where the gain matrix $\mathbf{K} = \mathbf{BM}^T\mathbf{H}^T(\mathbf{HMBM}^T\mathbf{H}^T + \mathbf{R})^{-1}$ and the nonlinear innovation vector $\mathbf{d}_b^o = \mathbf{y}_1 - \mathcal{H}(\mathcal{M}(\mathbf{x}^b))$ [43]. Note that incremental 4DVar with one 'outer-loop' is equivalent to the best linear unbiased estimate (BLUE) method and therefore gives the same analysis solution (2.30). This explicit gain matrix $\mathbf{K}$ enables us to show that the update to the background state is dependent on the specification of both $\mathbf{B}$ and $\mathbf{R}$. Let us now define the following differences in observation space,

$$\mathbf{d}_b^o = \mathbf{y}_1 - \mathcal{H}(\mathcal{M}(\mathbf{x}^b)) \approx \epsilon_{ob} - \mathbf{HM}\epsilon_b, \tag{2.31}$$

$$\mathbf{d}_b^a = \mathcal{H}(\mathcal{M}(\mathbf{x}^a{}_0)) - \mathcal{H}(\mathcal{M}(\mathbf{x}^b)) \approx \mathbf{HMK}\mathbf{d}_b^o, \tag{2.32}$$

$$\mathbf{d}_a^o = \mathbf{y}_1 - \mathcal{H}(\mathcal{M}(\mathbf{x}^a{}_0)) \approx (\mathbf{I} - \mathbf{HMK})\mathbf{d}_b^o, \tag{2.33}$$

which assume the tangent linear hypothesis holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}$. Note that equations (2.31)-(2.33) each hold exactly when both the observation operator and model equations are of a linear nature. With $\mathbf{M} = \mathbf{I}$ these differences in observation space (2.31)-(2.33) are as described in the Desroziers et al. paper [38]. Diagnostics are then derived by taking statistical expectations in observation space. By taking the statistical expectation of the product of the innovation vector $\mathbf{d}_b^o$ (2.31) with itself we can show that,

$$E[\mathbf{d}_b^o(\mathbf{d}_b^o)^T] \approx \mathbf{R} + \mathbf{HMBM}^T\mathbf{H}^T. \tag{2.34}$$

By taking the statistical expectation of the product of the difference vector $\mathbf{d}_b^a$ (2.32) with the innovation vector $\mathbf{d}_b^o$ (2.31) we can show that,

$$E[\mathbf{d}_b^a(\mathbf{d}_b^o)^T] \approx \mathbf{HMBM}^T\mathbf{H}^T. \tag{2.35}$$

By taking the statistical expectation of the product of the difference vector $\mathbf{d}_a^o$ (2.33) with the innovation vector $\mathbf{d}_b^o$ (2.31) we can show that,

$$E[\mathbf{d}_a^o(\mathbf{d}_b^o)^T] \approx \mathbf{R}. \tag{2.36}$$

Finally, by taking the statistical expectation of the product of the difference vector $\mathbf{d}_b^a$ (2.32) with the difference vector $\mathbf{d}_a^o$ (2.33) we can show that,

$$E[\mathbf{d}_b^a(\mathbf{d}_a^o)^T] \approx \mathbf{HMBM}^T\mathbf{H}^T(\mathbf{HMBM}^T\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R} \tag{2.37}$$

$$= \mathbf{HMAM}^T\mathbf{H}^T, \tag{2.38}$$

where $\mathbf{A}$ is the analysis error covariance matrix corresponding to the analysis of the model state (2.30). These equations (2.34)-(2.38) are consistent with the Desroziers et al. diagnostics [38] when all observations are at the start of the time window. The first of the diagnostic equations involving model evolution (2.34) was previously deduced by Andersson [2]. Note that in the presence of both a linear observation operator and linear model matrix, the equations (2.34)-(2.37) will be exactly equal.

Practical use of the diagnostics involves calculation of both sides of the equations independently and then comparing the results. The left-hand sides (LHS) of the diagnostic equations involve the differences $\mathbf{d}_b^o$ (2.31), $\mathbf{d}_b^a$ (2.32) and $\mathbf{d}_a^o$ (2.33) in observation space. To compute the expectations in the LHS of (2.34)-(2.37), samples

Difference vector	Data required
$\mathbf{d}_b^o$	Sample of background vectors at time t_0
	Sample of observation vectors at time t_1
$\mathbf{d}_b^a$	Sample of background vectors at time t_0
	Sample of analysis vectors at time t_0
$\mathbf{d}_a^o$	Sample of analysis vectors at time t_0
	Sample of observation vectors at time t_1

Table 2.1: Obtaining sample vectors of differences in observation space

of the vectors $\mathbf{d}_b^o$ (2.31), $\mathbf{d}_b^a$ (2.32) and $\mathbf{d}_a^o$ (2.33) are required. Table 2.1 details what data is required to obtain samples of each of the differences in observation space. For each sample innovation vector $\mathbf{d}_b^o$ a background sample vector is selected and evolved using the forecast model and then mapped to observation space using the observation operator. This resulting vector is then subtracted from a sample vector of observations. For each sample innovation vector $\mathbf{d}_b^a$ both a background sample vector and an analysis sample vector are selected and separately evolved using the forecast model and then mapped to observation space using the observation operator. This resulting evolved background vector is subtracted from the evolved analysis vector. For each sample innovation vector $\mathbf{d}_a^o$ an analysis sample vector is selected and evolved using the forecast model and then mapped to observation space using the observation operator. This resulting vector is then subtracted from a sample vector of observations. Once a sufficient sample size has been collected for each of the differences in observation space $\mathbf{d}_b^o$, $\mathbf{d}_b^a$ and $\mathbf{d}_a^o$, a sample of each of the products $\mathbf{d}_b^o(\mathbf{d}_b^o)^T$, $\mathbf{d}_b^a(\mathbf{d}_b^o)^T$, $\mathbf{d}_a^o(\mathbf{d}_b^o)^T$ and $\mathbf{d}_b^a(\mathbf{d}_a^o)^T$ is evaluated. Subsequently the mean of the samples of products are taken to obtain the expectations as described by the LHS of (2.34)-(2.37).

If the error covariance matrices have been specified correctly, then the diagnostic equations will hold, approximately with nonlinear operators and exactly with linear operators (for a significantly large sample size of difference vectors). These diagnostics

(2.34)-(2.38) assume a perfect model is available for use in the data assimilation process. We next consider the effect that the use of an erroneous model in data assimilation has on the diagnostics.

2.3.2   Diagnostics with model error present

Let us consider the best known representation of a true dynamical system over one time step to be a nonlinear erroneous model operator $\mathcal{M}^e$, with the unknown model error vector at time t_1 given by $\boldsymbol{\eta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$. When using the nonlinear erroneous model operator $\mathcal{M}^e$ the innovation vector is given by,

$$\mathbf{d}_b^{o*} \;=\; \mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)) \approx \epsilon_{ob} - \mathbf{HM}^e \epsilon_b + \mathbf{H}\boldsymbol{\eta}_1. \tag{2.39}$$

Taking the statistical expectation of the innovations gives,

$$E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T] \;\approx\; \mathbf{R} + \mathbf{HM}^e \mathbf{BM}^{eT} \mathbf{H}^T + \mathbf{HQH}^T, \tag{2.40}$$

assuming the model error, the errors in the background and the errors in the observations are uncorrelated. Note that this equation (2.40) will hold exactly with both a linear model matrix and linear observation operator. Equation (2.40) clearly has an additional term when compared to the corresponding diagnostic calculated with a perfect model (2.34) [33] [2]. Practical use of (2.40) involves the calculation of the mean product of a sample of innovation vectors $\mathbf{d}_b^{o*}$ (2.39). In Section 2.3.1 we described how this would be performed in the absence of model error. However, here each sample innovation vector $\mathbf{d}_b^{o*}$ would require a sample model error vector. Possible methods to gain such data are discussed in Section 6.1.4 of this thesis.

A similar result to (2.40) has also been derived by Todling [116] using an Extended Kalman Filter (EKF) data assimilation technique, which we will discuss further in Chapter 4. The term $\mathbf{HQH}^T$ represents the covariance of the model errors in observation space. Interestingly this diagnostic (2.40) gives us a consistency check in observation space not only for the background error and observation error covariance matrices $\mathbf{B}$ and $\mathbf{R}$ respectively, but also for the model error covariance matrix $\mathbf{Q}$. This relates to the first of the thesis objectives set out in Section 1.1 and will be further investigated in Chapter 6. Clearly care should be taken to use the appropriate diagnostic equation; either (2.34) or (2.40) depending on whether the model to be used in strong constraint 4DVar is perfect or erroneous. This will be discussed further in the next chapter.

2.4   Summary

In this chapter we have introduced variational data assimilation (Var) methods, which use background, observation and model information to provide the best estimate of the state of a system. We have given a brief overview of Var methods that assume model equations to be true representations of dynamical systems. This thesis will focus on the fact that often models contain error and how this should be best dealt with in the data assimilation process. Hence, we also described Var methods that have been formulated to deal with additive error in a model. For use of all data assimilation methods detailed in this thesis, background error and observation error covariance matrices need to be specified. We have discussed how diagnostic tools can be very useful to assess the consistency of these matrices in observation space and shown for the first of the diagnostic equations that the presence of random error in a model influences the diagnostic result. The next chapter will describe current use of Var methods at operational NWP centres, when estimating initial conditions for

atmosphere-ocean forecasts and when estimating the conditions for the purpose of reanalysis. Operational models are often erroneous and we will highlight potential areas of improvement in the current data assimilation processes and in the diagnostic tools used.

Chapter 3

Atmosphere and ocean data assimilation

In Section 3.1 of this chapter we provide a description of atmosphere-ocean models. The models are best representations of the true atmosphere-ocean dynamics and contain model error [104] [83] [106] [120]. In Section 3.1 we also discuss the potential origins of the model errors and the subsequent effects they have on the model state evolution. Scientists at meteorological centres use data assimilation to initialise atmosphere forecasts, ocean forecasts, coupled forecasts and for the purpose of reanalysis, where the objective is to seek the best estimate of historical weather events. In Section 3.1 we outline current data assimilation methods used operationally for both forecasting and reanalysis. In Section 3.2 we then discuss the operational use of consistency diagnostics as quality control checks for background error and observation error covariance matrices [38]. Finally, in Section 3.3 we highlight future aims of operational weather centres, in particular the objectives which involve how to best deal with model error within the data assimilation process.

3.1 Coupled models, forecasts and reanalysis

Reliable forecasts of both the atmosphere and ocean are highly desirable for use by scientists, governments, meteorological companies and other businesses, as well as to the general public. The ability to be able to produce good quality forecasts help with planning and risk assessment, for example to be able to predict the impact of the seasonal weather on crops [89].

The importance of the interaction between the atmosphere and the ocean has been recognised for many years [68]. It has been shown that coupling atmosphere and ocean models enhances the seasonal to interannual forecasts for both the atmosphere and the ocean [114]. The sun is the source of energy that contributes towards motion in the atmosphere and the ocean [52]. Over half of this solar radiation is absorbed at the surface of the earth, of which the ocean and seas cover 70.8% [111]. Therefore ocean conditions, in particular sea surface temperature (SST), are very influential on the atmospheric conditions, such as air temperature. Both the ocean and atmosphere store and exchange energy in the form of heat, moisture, and momentum. Hence, circulation in the atmosphere and ocean should be considered as a coupled system due to the continuous interactions. The influences the atmosphere and ocean have on one another vary from instantaneous, for example surface ocean currents affecting the atmospheric wind speed, to longer, slower processes which include the upper layers of the ocean storing heat in the summer that is later released into the atmosphere in the winter months [52].

Within the meteorological community, the study and prediction of weather extremes such as El Niño and Southern Oscillation (ENSO), which is a coupled atmosphere-ocean phenomena, are of great interest. El Niño is a term that defines periods of anomalously warm tropical Pacific sea temperatures, which occur every 3-7 years

[108]. The term Southern Oscillation defines atmospheric changes that are related to El Niño which can be described as a see-saw of atmospheric pressure between Indonesia and the South Pacific Ocean [108]. ENSO is known to affect patterns of temperature and rainfall across the globe [13] and is a good example of the significant impact atmosphere-ocean interactions can have. Next we introduce coupled atmosphere-ocean models and outline how these are derived.

3.1.1 Atmosphere, ocean and coupled models

At operational weather centres both atmosphere forecasting models and ocean forecasting models are derived from key governing equations [52] [96] [107]. These governing laws form a set of partial differential equations (PDEs) for the atmosphere and ocean circulation, known in the meteorology community as the fundamental equations of fluid dynamics [12]. Atmospheric variables present in the PDEs can include potential temperature, pressure, wind and density [121]. Examples of ocean variables that can be present in the PDEs include potential temperature, salinity, sea surface height (SSH) and the horizontal components of velocity [5]. It would be impossible to define these variables at every point in the atmosphere and ocean, as this problem would be of an infinite size. Therefore, the atmosphere-ocean domain is discretized. For example, an atmospheric domain can be discretized spatially into a longitude−latitude grid and vertically divided into intervals by pressures [121] [128]. An example of an ocean domain discretization consists of a longitude−latitude grid which is stretched near the poles and has varying vertical levels which are closer together near the surface [112]. Operational centres use numerical schemes to approximate the solutions of the PDE's, such as a semi-Lagrangian scheme [96], resulting in discretized nonlinear equations. These discretized nonlinear equations require the specification of a model time-step Δt. All operational numerical models of the atmosphere and ocean are based on

the same set of governing equations, though may differ in the approximations and assumptions made in the application of these equations.

To form operational coupled atmosphere-ocean models, the atmosphere and ocean models are combined, with both the atmosphere and ocean variables forming a coupled model state vector $\mathbf{x}$ as described in equation (2.2). For example, at the ECMWF the atmosphere forecasting model used is the Integrated Forecast System (IFS) and the ocean forecasting model is the Nucleus for European Modelling of the Ocean (NEMO) [96] [5], which are combined to produce a coupled atmosphere-ocean model. The exchange of heat, momentum and water at the atmosphere-ocean interface is an important part of the coupled dynamics, allowing the atmospheric and ocean variables to influence one another. Bulk formula are used to represent these fluxes in the coupled model equations [86], which require the specification of coupling parameters. The coupled system equations are often known as coupled atmosphere-ocean general circulation models, such as GloSea developed at the Met Office for global seasonal forecasting [53].

3.1.2 Errors in coupled atmosphere-ocean models

Very little is known about the characteristics of the errors present in atmosphere models, ocean models and hence coupled atmosphere-ocean models [89]. The possible sources of model errors we discuss in this section are relevant to all three types of dynamical models. Therefore, to avoid repetition, we will only discuss the potential origins of the errors present in the context of coupled atmosphere-ocean models.

Coupled models require specification of atmospheric parameters, ocean parameters and coupling parameters. These are often assumed to be constant in time and are best estimates which contain inaccuracies [64]. For example, one parameter required is the thermal roughness length, which is the height above ground where the mean

wind speed is zero and is required in the calculation for the momentum flux [109]. This parameter varies depending on the physical surface in the area, for instance over smooth water this value can be less than 10^{-4}m while built up areas, such as cities, can have a roughness length of up to 5m. Parameters such as this, need to be specified for each spacial model grid point. Significant errors in model parameters can cause the model run to produce different forecast predictions to those produced if the true model parameters were used [64].

The coupled model vector comprises the atmosphere and ocean variables at all points on the model grid. Operational coupled model vectors currently consist of an atmospheric component of dimension $\mathcal{O}(10^8 - 10^9)$ and an ocean component of dimension $\mathcal{O}(10^7)$ [85]. Lack of model resolution leads to coupled atmosphere-ocean models using parameterisations of physical processes that are on scales too small to be directly resolved by the model. These physical processes can be parameterised inadequately or even be absent entirely from coupled atmosphere-ocean models, introducing further error into the model. For example, shallow cloud features are not parameterised [67], but are thought to have a significant role in the climate system and are important in local weather predictions [9].

In order to solve the atmosphere and ocean PDEs, numerical schemes are used to provide time-stepping models. It should be mentioned that these discretized models are approximate solutions to the PDEs and therefore inevitably contain error [88]. Boundary conditions at the edge of the atmosphere-ocean spacial domain under consideration are required for use in a coupled model and are often estimates and therefore are also a source of model error [54]. For example, regional coupled models require boundary conditions at the borders of the horizontal domain [63].

Of course, if the exact sources of model error could be identified and resolved, the true dynamical coupled system could be modelled perfectly. However, this is not the case. Although multiple possible origins of model error in coupled atmosphere-ocean

models can be described, the specific identification of the error in the model equations is highly complex. As stated in the second of the thesis aims in Section 1.1, work in this thesis will focus on how to reduce the model error influence within the data assimilation process and this work will be conducted in Chapter 7. Work in this thesis will also focus on how to reduce the model error influence on a coupled model forecast, as stated in the third of the thesis aims in Section 1.1, and this work will be conducted in Chapter 8.

In this thesis we will assume that multiple sources of error lead to a vector of additive model error $\boldsymbol{\eta}_i$ at each time t_i, where the true state could be acquired if the model error vector was known, as in equation (2.3). Broadly speaking, the model error can have two effects on the model state variables. The first of which has a random nature at each time-step and the second of which is of a systematic nature correlated in time. Both additive random model error and additive systematic model error are thought to be present in coupled atmosphere-ocean models. We will consider errors of a random nature in work conducted in Chapter 6 and Chapter 7 and subsequently consider errors of a systematic nature in work conducted in Chapter 8. An example of the presence of systematic model error is that seasonal to decadal predictions tend to drift away from observed states, due to imperfect model equations [130] which suggests some form of model bias is present. Bias in a model may be constant in time or vary, for instance with the seasons [84]. Next we will outline current data assimilation methods used at centres such as the Met Office and ECMWF, to estimate conditions for both the atmosphere and ocean, including use of a systematic correction scheme to mitigate errors in part of the ocean model [84] [11].

3.1.3 Atmosphere, ocean and coupled forecasts

Observations of the atmosphere and ocean are collected from both in-situ instruments and satellites. In-situ observations are made at the location and therefore can measure small scale features, but have poor global coverage. In-situ measuring instruments include ground based instruments, balloons, aircraft, buoys and ships which take observations such as temperature, pressure, wind speed and wind direction [65]. Global observation coverage significantly increased from 1960 when the first weather satellite was launched [101]. Since the 1990's satellite observations have been used to greatly improve the NWP forecasts [16], however, many observations are not directly of the atmospheric or oceanic model variables. For example, satellites are often used to measure radiances, which can subsequently be related to model variables such as temperature and moisture [1].

Observation errors are taken into consideration in the data assimilation process. Within the data assimilation community we define $\mathbf{y}_i$ to be the vector of observations at time t_i, with the corresponding error vector $\boldsymbol{\epsilon}_{obi}$, as defined in equation (2.4). Observation errors can originate from multiple sources including instrument error, pre-processing errors (such as cloud detection), errors in observation operators $\mathcal{H}_i$ and representativity error (the model not correctly representing small scale features that observations are measuring) [31]. There are a vast number of observations, for example the number of atmospheric observations considered is of the order $\mathcal{O}(10^7)$. In practice, observation errors are likely to include correlations [126] [127] [110], however operational centres often only specify the diagonal elements of the observation error covariance matrix $\mathbf{R}_i$ at each time t_i [85]. This is because the correlations of the observation errors are often unknown and even if they were known, the matrices $\mathbf{R}_i$ are too large to specify all the entries. However, recent work with the Met Office 4DVar assimilation scheme has introduced correlations into the observation error covariance

matrices for certain atmospheric observations from the high resolution sounder IASI, which in turn has shown improvements in forecast accuracy [127].

Both a background model state for the atmosphere and a background model state for the ocean are required, with corresponding background error covariance matrices, for subsequent use in the atmosphere data assimilation and ocean data assimilation methods. The background model state $\mathbf{x}^b$ of the form (2.5) is usually taken from a previous forecast. The specification of background error covariance matrices $\mathbf{B}$ is of much importance as they determine the relative weight that is given to the background state as opposed to the observations and also influence how the information from the observations is spread to the model state variables [70]. The calculation of background error covariance matrices is a complex task for both the atmosphere and ocean background model states. One method used is to evaluate information about the errors in the background by calculating the statistics of the differences between 24 hour and 48 hour forecasts [7]. Operational background matrices are often so large that they are impossible to store explicitly. The use of control variable transforms at operational meteorology centres, such as the Met Office, enables $\mathbf{B}$ to be implicitly included in the data assimilation cost function [8]. The method involves letting the background error covariance matrix $\mathbf{B} = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L}$ is the change of variable operator [3] [8]. A new variable is introduced $\boldsymbol{\chi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ such that,

$$\mathbf{L}\boldsymbol{\chi} \;\; = \;\; \mathbf{x}_0 - \mathbf{x}^b, \tag{3.1}$$

allowing the background term of the cost function (2.10) to be rewritten,

$$\mathcal{J}_b = \frac{1}{2}\boldsymbol{\chi}^T\boldsymbol{\chi}, \tag{3.2}$$

to implicitly include the background error covariance matrix [47].

Currently operational centres use data assimilation to estimate the initial conditions of the atmosphere and ocean separately. We will firstly outline methods used to estimate initial conditions of the atmosphere and then subsequently describe how initial conditions are estimated for the ocean.

Operational NWP centres, such as the Met Office and ECWMF, use an incremental version of the strong constraint 4DVar method, as described in Section 2.2.3, to estimate the initial conditions of the atmospheric model state vector. The incremental version of 4DVar requires less computational resource and time than the standard 4DVar method, due to the lower resolution used in the inner loop, and therefore is feasible for use at NWP centres [28]. When operational NWP centres use incremental 4DVar to best estimate initial conditions of the atmospheric model state vector, the assimilation window length needs to be specified. For example, the Met Office currently use an assimilation window length of 6 hours [99] and ECMWF currently use a 12 hour window [96]. Atmospheric models are updated with the most currently available SST, for example the Met Office Unified Model is updated with the latest SST update from the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system [41]. Available observations are selected for use in the specified time window and incremental 4DVar applied to best estimate the initial conditions. The initial conditions at distinct weather centres will vary for multiple reasons including differences in the following: model equations, model resolution, number of outer and inner loops, available observations, specification of the observation operators and assimilation window length. Incremental 4DVar is performed periodically to best estimate initial conditions for the purpose of running NWP forecasts, for example at the Met Office every 6 hours [99].

Operational NWP centres, such as the Met Office and ECMWF, best estimate the initial conditions for ocean model state vectors using a different data assimilation

technique, known as incremental 3D-FGAT (first guess at appropriate time) [18] [124]. This 3D-FGAT method has similarities with both incremental 3DVar and incremental 4DVar. Observations are used across a specified assimilation window as in 4DVar. However, no model evolution is conducted in the 3D-FGAT inner loop, this alleviates the 4DVar requirement for a full tangent linear and adjoint model [69] and this makes the 3D-FGAT method similar to 3DVar [124]. Prior to use within the data assimilation process, operational ocean models are prescribed best estimates of surface momentum, heat and freshwater fluxes at the atmosphere-ocean interface. These surface forcing fields can be prescribed from global NWP model estimates [10]. Data assimilation window lengths have to also be specified by operational centres. These can be significantly longer for ocean data assimilation than atmosphere data assimilation, due to the slower nature of ocean dynamics in comparison to the faster chaotic nature of atmospheric dynamics. For example, ECMWF use an assimilation window length of 10 days for ocean data assimilation [6], compared with a notably shorter 12 hour assimilation window length for the atmosphere [96].

The topography of the ocean surface reacts to the wind just above the surface. Therefore the wind affects the pressure gradient in the ocean. Wind stress is a model input and if specified incorrectly can lead to an imbalance with the pressure gradient estimated in the data assimilation process [11]. It has been shown that systematic errors are present in the wind-stress inputs [58]. A systematic error correction method is used in the ocean data assimilation process, as described in Section 2.2.5, known in this application as the pressure correction method. This estimates temperature and salinity bias fields which 'correct' the pressure gradient. The resulting pressure gradients oppose those originally imposed and restore balance between pressure gradients and wind stress [11].

NWP centres, such as the Met Office and ECMWF, use coupled atmosphere-ocean models for both seasonal and decadal forecasts and climate reanalysis [55] [75] [21]

[66]. The use of coupled models to also produce forecasts on shorter length time-scales, for example NWP, has been a recent area of work at operational centres including the Met Office, ECMWF, the Naval Research Laboratory and Meteo-France [21]. Currently initial conditions of the atmosphere and ocean are estimated separately, using the respective data assimilation schemes, and subsequently used to initialise coupled atmosphere-ocean forecasts [66]. Current research at operational centres is exploring the use of coupled data assimilation techniques to simultaneously estimate both the atmosphere model state and ocean model state [75] [55] [114] [21], further details are given in Section 3.3. Next we discuss the operational use of reanalysis.

3.1.4 Reanalysis

Reanalysis uses modern data assimilation methods on past time periods, by recalculating the analysis. Reanalysis can be conducted over short or long time periods. The purpose of reanalysis is to improve on the most current estimates of past weather events. Examples include, atmospheric reanalysis at centres such as ECMWF (ERA-Interim reanalysis) to produce data products spanning back decades to investigate past weather events [35]. To produce the ERA-Interim reanalysis a 12 hourly 4DVar cycle is used for the upper-air atmospheric state [35]. Whereas to produce the reanalysis of the surface of the atmosphere an Optimal Interpolation scheme is used [35].

The Ocean Reanalysis System 4 is implemented at ECMWF, providing reanalysis of the ocean from 1958 onwards [5]. This system uses an incremental version of 3DVar-FGAT [5]. The ocean reanalysis is used, along with the ocean initial conditions, to produce coupled seasonal forecasts at at ECMWF [5].

A comparison of reanalysis conducted with different models can be useful to assess the increase in analysis and forecast accuracy obtained though implementation of a

new model. For example, the Met Office conducted a 2 year reanalysis with the FOAM (Forecast Ocean Assimilation Model) using an updated NEMO model, and compared the forecast accuracy with that from the previous FOAM system [15]. The data assimilation method used was a 3DVar incremental FGAT scheme. This resulted in observing improvements to near surface ocean fields, but a degradation to some sub-surface fields which highlighted areas for future improvements in the model equations [15].

The second aim of this thesis, as detailed in Section 1.1, is to investigate how to better estimate a model state by allowing for model error in the data assimilation process, this work is conducted in Chapter 7. This work should be applicable to improve the reanalysis of atmosphere and ocean conditions.

3.2 Use of Desroziers diagnostics

We will now discuss use of diagnostics tools; firstly the Desroziers diagnostics [38] in which model evolution is unaccounted for, and secondly the use of the diagnostic equation (2.34) that accounts for model evolution, but not model error. The diagnostics themselves should not be relied on to specify background error or observation error covariance matrices, but can be useful as quality checks and to identify the areas of mis-specification of the variances and correlation information [39] [37].

Observational information from IASI was assimilated at the Met Office with a diagonal observation error covariance matrix until January 2013 [127]. However, recent work involving diagnostic tools allowed improvements to be made to the observation error covariance matrix, which we will now outline. The right-hand side of the third of the diagnostic equation $E[\mathbf{d}^o_a(\mathbf{d}^o_b)^T]$ (2.36), both when not accounting for model evolution [38] and when accounting for model evolution, is equal to the observation error covariance matrix $\mathbf{R}$. Work using the third of the Desroziers diagnostic equa-

tions $E[\mathbf{d}^o{}_a(\mathbf{d}^o{}_b)^T]$ [38], has enabled the recognition of inter-channel correlations in the errors of certain IASI radiance observations assimilated in 4DVar at the Met Office [110]. The variances on the diagonal of the observation error covariance matrix, for certain channels of the IASI observations, were over specified to account for the omittance of the inter-channel error correlation information [127] [110]. Use of the Desroziers et al. diagnostic equation $E[\mathbf{d}^o{}_a(\mathbf{d}^o{}_b)^T]$ [38] enabled specification of inter-channel error correlations for IASI data in the observation error covariance matrix, for example specification of strong correlations between water-vapour sensitive channels [127]. The use of this improved observation error covariance matrix in 4DVar at the Met Office lead to an improvement in the forecast accuracy, due to more weight given to these IASI observations [127].

The other three diagnostics, specified in equations (2.34),(2.35) and (2.37) with the inclusion of model evolution, include the term $\mathbf{HMBM}^T\mathbf{H}^T$. The evaluation of $\mathbf{HMBM}^T\mathbf{H}^T$ is required in order to evaluate the right-hand side of the equations, to then compare with the calculations of the expectations on the left-hand side. For an operational sized NWP system, this transformation of the background error covariance matrix is impossible to evaluate directly, primarily as operational background error covariance matrices are huge and therefore cannot even be specified explicitly. Therefore, a randomization technique has been developed to approximate the standard deviations in a background error covariance matrix, providing a low rank estimate of $\mathbf{B}$ [48]. This randomization method has been extended to approximate the diagonal of $\mathbf{HBH}^T$ [3]. A further development of the randomization technique allowed the elements of $\mathbf{HMBM}^T\mathbf{H}^T$ to be approximated. We will now outline this method. Firstly rewrite equation (3.1) in terms of the model increment,

$$\mathbf{L}\boldsymbol{\chi} \quad = \quad \delta\mathbf{x}_0, \tag{3.3}$$

48

where $\delta\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$. It follows that $\mathbf{HM}\delta\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{HMBM}^T\mathbf{H}^T)$. A random sample of n vectors is generated, each independently drawn from $\boldsymbol{\chi}^i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ for $1 \leq i \leq n$ [3]. The term $\mathbf{HMBM}^T\mathbf{H}^T$ can subsequently be estimated as follows [2],

$$\mathbf{HMBM}^T\mathbf{H}^T \approx \frac{1}{n}\sum_{i=1}^{n} \mathbf{HML}\boldsymbol{\chi}^i(\mathbf{HML}\boldsymbol{\chi}^i)^T, \tag{3.4}$$

using the fact that $\mathbf{B} = \mathbf{LL}^T$ [3] [8].

This method of approximating $\mathbf{HMBM}^T\mathbf{H}^T$ enabled the first diagnostic equation (2.34), that accounts for model evolution, to be used in operational experiments [2]. For example, experiments were conducted using the ECMWF atmospheric system, operational in October 2003, with vast amounts of data, approximately $3,500,000$ observations every 12 hour assimilation period. A six day period was selected in February 2003 and the large number of observations were binned into hourly time intervals by observation type and geographical area, for example wind data from aircraft over North America. The randomisation method (3.4) with a sample size of $n = 100$ was used to approximate $\mathbf{HMBM}^T\mathbf{H}^T$. The diagonal elements of the diagnostics (2.34) were calculated and the left and right-hand sides compared for the binned data. Assuming a perfect model, the results suggested that the background error variances were underspecified in $\mathbf{B}$ [2]. However, the atmospheric model used in the experiments contained model error and this could be the explanation for the resulting inconsistencies in the evaluation of the diagnostic (2.34) [2]. The expectations computed with the innovation data were larger than the direct computations using the error covariance matrices. The expectations that were computed with the innovation data also increased with time. It was suggested by Andersson [2] that this could have been due to the growth of the model error $\mathbf{Q}$ over time. Let us refer to

the corresponding diagnostic that accounts for error in the model (2.40), where there is the presence of the additional term $\mathbf{HQH}^T$ due to the model error, which could possibly have made up for the inaccuracies in these experiments.

3.3   Future aims of operational weather centres

Weather centres, such as the Met Office and ECMWF, are currently researching potential benefits of simultaneously estimating the initial conditions for both the atmosphere and ocean using a coupled model in a coupled data assimilation process [75] [55]. The main objectives of this work are to improve coupled forecasts and increase the accuracy of coupled reanalysis [74] [66] [75] [55] [102]. Estimation of a coupled analysis should improve the balance between the atmosphere and ocean initial conditions and hence reduce initialisation shock at the start of the coupled forecast [4]. Initialisation shock is a fast drift, even a jump, at the start of a forecast window. This is primarily caused by the fact initial conditions are estimated separately for the atmosphere and ocean using separate respective atmosphere and ocean models, leading to imbalances in these initial conditions which can be far from the natural state of the coupled model. Coupled data assimilation should also allow coupled model errors that grow in the first few hours of the forecast to be better understood, by comparing the original forecast with the later updated coupled analysis trajectory [75].

Weakly coupled data assimilation is an incremental technique that uses a coupled atmosphere-ocean model in the outer loops, but performs separate atmosphere and ocean inner loops. This is the coupled data assimilation technique formulated at operational centres, such as the Met Office and ECMWF, for current research due to the relative ease of implementation, compared to a fully coupled technique [75]. The computer code for the inner loops of the atmosphere and ocean increment updates

require little alteration and no atmosphere-ocean cross covariance error information is required. For example, the Met Office uses the atmospheric Met Office Unified Model and the ocean NEMO model coupled every hour in the outer loop, however the respective models remain uncoupled in the inner loops [75]. Results from simultaneously estimating the atmosphere and ocean initial conditions using weakly coupled data assimilation as opposed to estimating the atmosphere and ocean initial conditions separately have shown improvements to both coupled forecasts and reanalysis [55] [102]. For example, experiments at the Met Office, using a 6 hour assimilation window, have shown a reduction in forecast error for atmospheric temperature in the northern hemisphere [75].

The Japan Agency for Marine-Earth Science and Technology (JAMSTEC) has developed a coupled data assimilation method with results showing improvements in the representation of major seasonal to interannual dynamical processes [114]. This coupled system uses long assimilation windows of length 9 months, course spatial grids and an approximated adjoint, with the key objective to better estimate climate processes. This coupled 4DVar technique estimates ocean initial conditions and bulk adjustment factors (latent heat, sensible heat and momentum fluxes) [114], but does not explicitly estimate the atmospheric initial conditions. Therefore, this type of coupled data assimilation system is not suited to best estimate atmospheric and oceanic conditions for shorter time-scales, in particular it cannot be used for NWP initialisation [55].

Future work at operational centres involves continued research using weakly coupled data assimilation and investigating the potential operational implementation of strongly coupled data assimilation [55]. Fully coupled data assimilation requires specification of coupled covariances. The Australian Bureau of Meteorology has undertaken preliminary research in this area using ensemble techniques [94]. Although coupled forecasts and reanalysis can be improved through coupled data assimilation

51

techniques, the errors in the coupled models, as described in Section 3.1.2, limit the accuracy.

The weakly coupled data assimilation process at the Met Office uses the pressure correction method [11], as described in Section 3.1.3, to correct the imbalance between wind stress and the pressure gradient. However, further use of this scheme is restricted, largely due to unknown model error statistics and unknown model evolution required for specification in this method. A key objective of weather centres is therefore to extend coupled data assimilation research to investigate how to best deal with model error in coupled models during the coupled data assimilation process [55] [75] which gives us motivation for work in this thesis.

The first of our objectives, as stated in Section 1.1, is to derive diagnostic equations that account for model errors to verify and refine the specification of a model error covariance matrix. This work is conducted in Chapter 6 of this thesis. The second of the thesis objectives, as stated in Section 1.1, is to develop a strong constraint 4DVar method to mitigate the effect model error has on the estimation of the initial conditions. This work is conducted in Chapter 7 of this thesis. The third aim of this thesis, as stated in Section 1.1, is to develop a strong constraint 4DVar method to improve the accuracy of a coupled atmosphere-ocean forecast. This work is conducted in Chapter 8 of this thesis.

3.4   Summary

In this chapter we outlined the derivation of coupled atmosphere-ocean models and described the multiple sources of model errors present. We then discussed the operational use of data assimilation for the purpose of atmosphere-ocean forecasts and reanalysis. We stated the use of a pressure correction method to mitigate the effect of error in part of the ocean model. We described the use of consistency diagnostics

as quality checks for the specification of both the background error and observation error covariance matrices and highlighted that not accounting for model error can lead to the diagnostics not equating. Lastly in this chapter, we have presented future objectives of operational weather centres and outlined how our work in this thesis is motivated by those objectives, in particular, how to account for model error in the operational data assimilation process. In the next chapter we will review methods that have been developed for this purpose.

Chapter 4

Model error in data assimilation

In Chapter 2 we described Var methods for use firstly with a perfect model and secondly with an erroneous model. In Chapter 2 we also introduced diagnostic tools which can be useful as quality checks for the specification of the statistics required in Var methods. We then described the operational application of data assimilation methods and the operational use of diagnostic tools at NWP centres in Chapter 3. We identified that a key objective of operational NWP centres is to better account for model error within data assimilation operationally. Therefore, in this chapter we outline work that has been conducted in this area including; highlighting the problems that have arisen and identifying ideas that we can develop further. In Section 4.1 we discuss the difficulties in specifying model error covariance matrices for both atmosphere and ocean models at operational NWP centres. In Section 4.2 we outline methods proposed to estimate a model error covariance matrix. In Section 4.3 we examine data assimilation methods that have been formulated to compensate for bias present in coupled models with the objective to improve coupled model forecasts. Finally, in Section 4.4 we discuss when the statistics of the errors in the comparison of observations with the model state are not properly accounted for in data assimilation.

4.1 Model error covariance matrices in weak constraint variational data assimilation

Operational weather centres, such as ECMWF [118], have conducted investigations into whether the use of weak constraint 4DVar could improve the quality of the atmospheric analysis and forecast. The general formulation of weak constraint 4DVar, as described in Section 2.2.4, assumes a vector of random model error $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$ is present at each time-step t_i. One of the reasons that is currently preventing implementation at operational NWP centres is the difficulty in specifying the required model error statistics in $\mathbf{Q}_i$. We will outline suggested methods to prescribe $\mathbf{Q}_i$ with the required model error statistics for operational systems and highlight the fundamental issues with these methods of specification.

To specify all entries in an atmospheric model error covariance matrix $\mathbf{Q}_i$ would require prescribing $\mathcal{O}(10^{18})$ elements [118]. Very little is known about the characteristics of error in forecast models and a better understanding of model error statistics is desired [89] [117]. For these reasons, methods to approximate simplified model error covariance matrices have been investigated. A common simplification is to assume the model error covariance matrix $\mathbf{Q}_i$ is constant over the assimilation window, denoted by $\mathbf{Q}$.

It has been suggested that $\mathbf{Q}$ should be proportional to the background error covariance matrix $\mathbf{B}$ [118] due to the fact the background error, also known as short-term forecast error, will include a component of model error. The relationship between the background error covariance matrix and model error covariance matrix is explicitly defined by one of the Kalman Filter data assimilation equations. Specifically the background error covariance matrix is re-evaluated at each sequential time-step and is composed of the propagation of the background error covariance matrix from the

last analysis time plus the model error covariance matrix [45]. At operational NWP centres information in the background error covariance matrix $\mathbf{B}$ is implicitly available through the change of variable operator $\mathbf{L}$ [3] [8], as discussed in Section 3.1.3, so this would be a possible choice for use at operational centres. Experiments performing incremental weak constraint 4DVar using the ECMWF atmospheric IFS system were conducted using $\mathbf{Q} = c\mathbf{B}$ [118], where c is a scalar. This choice of $\mathbf{Q}$ was found to restrict the ability of the weak constraint 4DVar method to estimate the model error appropriately. The lack of independence between $\mathbf{B}$ and $\mathbf{Q}$ led to both the model state analysis increment and the model error analysis increment being in the same direction [118]. Therefore, a method of approximating $\mathbf{Q}$ that would give weak constraint 4DVar more freedom to explore different directions, with the model error analysis increment, was next sought.

The differences between the tendencies of an ensemble of forecasts was used to approximate a static $\mathbf{Q}$ [118]. These forecasts were run from atmospheric analyses produced from an ensemble of 4DVar assimilations. The spread of the ensemble members can be thought to represent the probability distribution of the true model state. The differences between the tendencies can be interpreted as the ensemble of the model error [118]. ECMWF performed experiments using this method where the ensemble of the model error vectors were then used to produce an estimate of $\mathbf{Q}$. Note that $\mathbf{Q}$ and $\mathbf{B}$ are not necessarily proportional in this case [118]. Simple numerical experiments using weak constraint 4DVar were performed using the ECMWF IFS system. The model error control variable was taken to be constant over the assimilation window. The weak constraint 4DVar analysis was found to be more accurate than the strong constraint 4DVar analysis. However when applying the model error forcing correction to ten day forecasts there was a negative effect on the forecast performance. The assimilation window length of 12 hours was used in these experiments, which may be too long a period for a model error vector to be assumed constant. It

was also noted that caution should be taken in areas prone to observation bias, for example aircraft temperature taken when aircraft are ascending or descending, as weak constraint 4DVar may estimate model error in these areas inaccurately if the bias is not first removed from the observations [118].

Weak formulations of Var have not just been applied to atmospheric operational models, we next describe the application of a weak constraint Var method to an operational ocean model. Operational ocean models are likely to have substantial model biases [73] [6]. Work has been conducted using the FOAM system at the Met Office to mitigate the effect of both observation bias and model forecast bias [73]. The methodology used was an amended 3D-FGAT technique, a *weak* formulation of 3D-FGAT, used to simultaneously estimate observation bias and model bias along with the ocean model state. In this method the model state is assumed to be biased and requires a correction term. The bias in the model was assumed to be in the SSH (sea surface height) field. Investigations centred on the Atlantic and Arctic oceans and involved the assimilation of altimeter data, in situ temperature and salinity data and SST (sea surface temperature) data. The amended cost function required specification of the model error covariance matrix $\mathbf{Q}$. In experiments conducted with the FOAM system, the model error covariance matrix was specified as spatially uniform with the amplitude of 9×10^{-3} cm^2 day^{-2} and had a large correlation scale of 400km. The results from running this amended 3D-FGAT scheme showed improvements to the innovation statistics [73]. Together the estimation of observation bias and model bias helped mitigate the effect of the total bias, but it is unknown whether the resulting separation of the estimated observation bias and model bias was successful. Further work at the Met Office includes estimating an improved model error covariance matrix $\mathbf{Q}$ [73]. Further to this, future work also includes combining this developed scheme with the pressure correction method, as described in Section 3.1.3.

Next we introduce a method that has been developed to estimate a model error

covariance matrix. Note that this work has only involved investigations with an idealized model, not an operational model. A method to estimate a model error covariance matrix where the error in the model is only due to incorrectly specified parameter values has been developed [24]. We will now outline this technique for the case where the parameter errors are uncorrelated in time. Consider the model $\mathbf{x}_i = \mathcal{M}(\mathbf{x}_{i-1}, \boldsymbol{\phi}_{i-1}{}^m)$ which takes the model state from time t_{i-1} to t_i. The vector $\boldsymbol{\phi}_i{}^m$ contains the erroneously specified model parameters at time t_i. We denote $\boldsymbol{\epsilon}_{\phi_i} = \boldsymbol{\phi}_i{}^m - \boldsymbol{\phi}_i{}^t$ to be the error in the parameters at time t_i, where the vector $\boldsymbol{\phi}_i{}^t$ contains the true model parameters. The error in the parameters $\boldsymbol{\epsilon}_{\phi_i}$ at time t_i has the corresponding error covariance matrix $\mathbf{S}_i$. The model error covariance matrix $\mathbf{Q}_i$ can be approximated as follows [24],

$$\mathbf{Q}_i \approx \mathbf{M}_{\phi_{i-1}{}^m} \mathbf{S}_{i-1} (\mathbf{M}_{\phi_{i-1}{}^m})^T, \tag{4.1}$$

where $\mathbf{M}_{\phi_{i-1}{}^m} = \left. \frac{\partial \mathcal{M}}{\partial \phi_{i-1}} \right|_{\phi_{i-1} = \phi_{i-1}{}^m}$ is evaluated with the true model state values at time t_{i-1}. However $\mathbf{M}_{\phi_{i-1}{}^m}$ can be approximated using the background model state. If the covariance matrix of the errors in the parameters $\mathbf{S}_{i-1}$ is known a priori, then the model error covariance matrix $\mathbf{Q}_i$ for each time t_i can be approximated using (4.1). Results using the Lorenz 63 model [81] with an amended weak constraint 4DVar cost function showed improvements to the analysis trajectory, as opposed to using strong constraint 4DVar [24]. However, we emphasise that this method was tested using a toy model where the only source of model error were parameter errors. Operationally there are multiple sources of model error, as described in Section 3.1.2. Although this method was shown to improve the forecast of a toy model with erroneous parameter values, it does not account for the multiple sources of model error that are present in operational models. Work in this thesis will not concentrate purely on one origin

of model error, but will focus on dealing with the effect multiple origins of model error have on a system, i.e. the random or systematic additive model error vector as described in Section 2.1.2.

Due to the difficulty in prescribing model error statistics, work in this thesis will involve formulating methods to account for model error in variational data assimilation that do not require the specification of model error statistics explicitly. When deriving techniques to account for model error, both the nature of the model error, random or systematic, and whether the primary objective is to improve the accuracy of the analysis or subsequent forecast, will need to be considered. Specifically, in Chapter 7 we will work on the second aim of the thesis, as described in Section 1.1, to develop a 4DVar method to account for random model error present in a model in order to improve the accuracy of the analysis. Later in this thesis, in Chapter 8 we will work on the third aim of this thesis, as described in Section 1.1, to develop a method to mitigate the effect systematic error can have on a coupled forecast. Methods developed in both Chapter 7 and Chapter 8 will not require explicit specification of a model error covariance matrix.

We do not want to dismiss the use of methods, such as weak constraint 4DVar, which require the explicit specification of a model error covariance matrix. Therefore, work in Chapter 6 of this thesis will involve developing methods to verify, and under certain circumstances refine, the specification of a model error covariance matrix. This objective is as stated in the first aim in Section 1.1. We next outline how consistency diagnostics have been used to estimate model error covariance matrices and the associated problems that arose.

4.2 Consistency diagnostics to estimate model error covariance matrices

The estimation of model error covariance matrices has been an area of research for many years. We discuss the work of R. Daley [30], who in 1992 published a technique based on Kalman filter data assimilation theory to estimate a model error covariance matrix. The technique developed aimed to estimate the stationary component of a model error covariance matrix $\mathbf{Q}$ by subtracting an estimate for the predictability (background) error covariance matrix $\mathbf{P}^p$ from an estimate for the total forecast error covariance matrix $\mathbf{P}^f$. The Kalman Filter innovation vector is the difference between observations $\mathbf{y}_i$ at time t_i and the forecast state mapped to observations space $\mathbf{H}\mathbf{x}_i{}^f$. The estimate of the total forecast error covariance matrix $\mathbf{P}^f$ can be estimated in observation space by taking the expectation of the product of Kalman Filter innovation vectors and subsequently subtracting the observation error covariance matrix $\mathbf{R}$. The predictability error covariance matrix $\mathbf{P}^p$ is estimated by selecting an ensemble of Kalman filter analysis values $\mathbf{x}_i{}^a$ and subsequently perturbing these initial conditions with an ensemble of random perturbations. The covariance matrix $\mathbf{P}^p$ is evaluated using the statistics from the differences between the model evolved analysis values and model evolved perturbed analysis values. Then the stationary components of the model error covariance matrix $\mathbf{Q}^s \approx \mathbf{P}^f - \mathbf{P}^p$. However, there are significant disadvantages of estimating a model error covariance matrix using this technique which include; firstly, the method only estimates the stationary components of a model error covariance matrix and secondly, the fact that the approximate total forecast error covariance matrix $\mathbf{P}^f$ is obtained in observation space. In addition to these issues, due to the use of diagnostic residuals in this method, an accurate approximation to the model error covariance matrix in observation space is dependent on accurate specifica-

tion of both the predictability (background) error covariance matrix and observation error covariance matrix.

A method proposed by R. Todling [116] involves using residual statistics from sequential data assimilation methods to estimate a model error covariance matrix. There are two sequential data assimilation techniques used in this method, firstly the sequential Kalman filter and secondly the fixed lag-1 Kalman smoother. The sequential Kalman filter uses observations only at the analysis time to estimate the analysis, where as the fixed lag-1 Kalman smoother also uses observations at the subsequent time-step to the analysis time in order to improve the analysis [27]. The vector $\mathbf{w}_i$ is the difference between the observation operator applied to the smoother and filter analysis at time t_i. The vector $\mathbf{r}_i$ is the difference between the observations and lag-1 forecast mapped to observation space. An approximation of the model error covariance matrix in observation space is calculated by taking the expectation $< \mathbf{w}_i \mathbf{r}_i^T > \approx \mathbf{HQH}^T$. The use of these diagnostics in simple experiments with toy models and simple observation operators proved to be successful in estimating a stationary model error covariance matrix. However, while R. Todling [115] conducted further work in this area, problems with this method were identified when the background error and observation error covariance matrices were inaccurately specified, which led to inaccurate estimation of model error covariance matrices. The estimated statistics for errors in the background, observations and model were found to be inseparable. As previously mentioned in Section 3.1, both the background error and observation error covariance matrices are approximated for use in operational atmosphere and ocean data assimilation procedures and therefore caution would be required if a method such as this was to be used operationally to estimate a model error covariance matrix. Another consideration that would have to be made, if this approach were to be used operationally, is that the model error covariance matrix estimated using this method is static, whereas in reality it is likely to evolve with

time. Using this technique provides an estimate of the model error covariance matrix in observation space only, which prevents use of these model error statistics in weak constraint 4DVar, as described in Section 2.2.4. Also all of the information required for use of this method is not readily available to NWP centres using variational data assimilation, as this technique has been derived using sequential data assimilation methods.

We have outlined techniques developed to obtain an estimate of a model error covariance matrix using sequential data assimilation methods and have highlighted areas of caution and reasons that these methods may not be suitable for use at NWP centres. The techniques described have involved the use of sequential data assimilation methods. In Chapter 6 we develop methods to verify the specification of an estimated model error covariance matrix using variational data assimilation techniques. Specifically, we will derive diagnostic tools in the 4DVar framework which is often used at operational NWP centres.

In Section 4.1 we stated that a current limitation for operational use of weak constraint variational data assimilation is the specification of model error covariance matrices. One of the objectives in this thesis, see aim number 1 in Section 1.1, is to develop a method to verify the specification of a model error covariance matrix and subsequently investigate the potential use of such a method to refine estimated model error statistics. In Chapter 6 we further develop the strong constraint 4DVar consistency diagnostics to account for model error, similar to that of equation (2.40), with the objective to verify the consistency of an estimated model error covariance matrix, prior to operational implementation of the weak constraint 4DVar formulation. Subsequently, if an estimated model error covariance matrix passed these quality checks and weak constraint 4DVar was operationally implemented, there is the issue of whether these model error statistics would still be valid at a later date. Therefore, in Chapter 6 we also derive consistency diagnostics for weak constraint 4DVar.

Next we discuss data assimilation methods developed to account for bias in coupled atmosphere-ocean models without the requirement for the explicit specification of a model error covariance matrix.

4.3   Compensating for bias in a coupled atmosphere-ocean model

We outline data assimilation methods that have been formulated to compensate for bias in coupled atmosphere-ocean models, with the objective to improve coupled forecasts. These methods do not require model error covariance matrices to be specified, which is a key advantage to potential operational use due to the problems discussed in Section 4.1 and Section 4.2.

It has been shown that the effect of bias present in a simple coupled atmosphere-ocean decadal system has on the forecast can be reduced through estimation of model parameters along with the model state using an ensemble Kalman Filter method [130] [131]. The biased simple coupled model in these investigations was set to have erroneous atmospheric and oceanic parameter values. The values of all 16 atmospheric and oceanic parameters were estimated within the data assimilation process. Estimating model parameters allows for flexibility in the model so that the data assimilation process can better fit the model state to the observations. No prior knowledge about the model bias had to be specified to the data assimilation scheme for the method to be successful in reducing the effect the model bias had on the forecast.

Work by Lu et al. [82] using a simple coupled equatorial atmosphere-ocean model involved the simultaneous estimation of 6 damping and coupling parameters along with 3 ocean initial state variables. In total there were 10 model parameters. Erroneous parameter values, for all 10 model parameters, were used in the coupled model

in the data assimilation process. The simultaneous estimation of the damping and coupling parameters along with the ocean initial conditions was conducted with an amended 4DVar method. We highlight that the resulting simultaneous analysis of the ocean initial conditions and parameters was generally less accurate than when either the ocean initial conditions or parameters were estimated separately. However, what should be considered here is that when only the ocean initial conditions were estimated the true model parameters were assumed to be known and when the parameters were estimated the true model state initial conditions were assumed to be known. This knowledge of the respective true model parameters and true ocean initial conditions would not be applicable in an operational setting. Also, a key message that resulted from this work was that the initial conditions and parameters should be observable, meaning that the data supplied by the observations needs to be linked to the initial conditions and parameters estimated [82].

Another study which is of interest is that estimating adjustment factors in bulk formulae (for example latent heat, sensible heat or momentum fluxes) together with oceanic model state variables using 4DVar has been shown to improve the forecasting ability of an operational coupled atmosphere-ocean model, including partial compensation for bias in the model [114]. Yet again, no prior information about the bias in the model was specified to the data assimilation scheme for the benefits in the forecast accuracy to be obtained.

Adjustments to the momentum flux (at the atmosphere-ocean interface) in the ECMWF coupled model to correct for model bias, have shown improvements to ENSO (El Niño Southern Oscillation) forecasts at seasonal time-scales [83]. In this case the wind stress seen by the ocean was modified by the momentum flux correction. The momentum flux correction term was calculated for unique use in the coupled forecast as a separate process to estimating the model state initial conditions. This method pushes the model solution towards the observed climatology and hence nearer the true

model state trajectories. The flux correction calculation used is not straight-forward which makes this method unfavourable for continual implementation operationally.

The third of our thesis aims, as stated in Section 1.1, is to develop a strong constraint 4DVar method with the specific objective; to improve the accuracy of a coupled atmosphere-ocean forecast. We can use the fact that atmosphere-ocean models contain coupling parameters which help define the behaviour of the system at the atmospheric-oceanic interface. The true values of these coupling parameters are often unknown, with experts currently specifying best guesses to be used in coupled models. The estimation of coupling parameters along with the coupled atmosphere-ocean initial conditions has not yet been formulated in a 4DVar framework with the objective to mitigate the effect of model bias from a coupled forecast. In Chapter 8 we develop this amended 4DVar method and conduct investigations using this developed scheme with an erroneous idealized coupled atmosphere-ocean model.

Next we outline data assimilation methods that have been formulated to account for representativity error in the data assimilation process and discuss how these methods could be developed to instead account for model error.

4.4   Improper comparison of observations with the model state

Representativity error stems from the fact that observations can resolve scales that a model cannot [31]. In this section we describe data assimilation methods developed to account for representativity error. Further to this, we discuss how methods to account for representativity error in data assimilation have the potential to be developed to account for model error.

Let us recall the aim of 4DVar; to best estimate an initial state of a system with

use of a background state, observations and a model. The term in the cost function $\mathcal{J}_{ob}$ as previously defined (2.10) compares observations to an initial state evolved through time with use of a model. It is this term $\mathcal{J}_{ob}$ that we will be focusing on in this section.

The study of how to account for the effect of representativity error in the data assimilation process has been an area of interest for many years [32] [93] [56] [79]. The objective of work presented in this section is not to give a review of all the different methodologies developed to deal with representativity error, but to outline two methods specifically of interest due to the potential of adapting these ideas to deal with model error.

We firstly outline work of D. Hodyss and N. Nichols [56] where representativity error is defined to be the error that arises from the distinct nature of an atmospheric forecast model attractor and the true atmospheric attractor. This work extends the Kalman filter data assimilation technique [62], however the issues and results we discuss are also applicable to variational data assimilation. The true atmospheric attractor is assumed to be biased from the perspective of the forecast model. Here the forecast model is considered to be a truncated version of the truth with fewer degrees of freedom. Observations at time t_i are of the form $\mathbf{y}_i$, as previously defined (2.4), with the error in the observation $\boldsymbol{\epsilon}_{obi}$ assumed in this case to be only instrument error with corresponding error covariance matrix $\mathbf{R}_i$. When the true physical system is not represented by the forecast model, the error statistics in $\mathbf{R}_i$ that describe the Gaussian distributed observation instrument errors around the true model state, do not fully correspond to the comparison of the observations and forecast model evolved state in the cost function term $\mathcal{J}_{ob}$. In fact this work shows that in the presence of representativity error, the appropriate covariance matrix for use in the cost function term $\mathcal{J}_{ob}$ is a sum of both $\mathbf{R}_i$ and a covariance error term that involves representativity error statistics. Instead of replacing the covariance matrix $\mathbf{R}_i$ to account for

this, D. Hodyss and N. Nichols [56] amend the observation operator $\mathbf{H}_i$ to map the forecast model state to the observation, as opposed to the true model state to the observation. This enables a like for like comparison of the observation and model state and enables the covariance matrix $\mathbf{R}_i$ to be the appropriate matrix to be used in the cost function term $\mathcal{J}_{ob}$. There are two key outcomes of this technique, firstly consistency diagnostics, as described in Section 2.3.1, will now hold with this amended observation operator (assuming correctly specified $\mathbf{B}, \mathbf{R}_i$ and the exclusion of model error). Secondly, this method produces a best estimate of the model state on the forecast attractor. It should be noted that this best estimate is not necessarily closer to the true conditions than when the conditions are estimated without change to the observation operator. So the use of this method would depend on the objective of the data assimilation and would be more suited to forecasts than reanalysis applications. Application of this method to operational NWP systems would require the specification of the mapping between the true and and forecast model attractors, which is a current area of research [56].

We next discuss work by Li et al. [79] which involves use of a multi-scale data assimilation scheme to account for representativity error. This method amends the incremental 3DVar cost function to account for the different resolutions of the observations and the model grid respectively. Theory is derived for the case where high resolution observations are used with a low resolution model and corresponding model state vector. The model state increment can be split into the sum of two vectors $\delta\mathbf{x} = \delta\mathbf{x}_L + \delta\mathbf{x}_S$, the first representing large scale features $\delta\mathbf{x}_L$ and the second representing small scale features $\delta\mathbf{x}_S$. This separation of the large and small scale features is also applied to the background error $\boldsymbol{\epsilon}_b = \boldsymbol{\epsilon}_{bL} + \boldsymbol{\epsilon}_{bS}$. The background error covariance matrix $\mathbf{B}$ is divided into the error statistics for the large scale background components $\mathbf{B}_L$ and the error statistics for the small scale background components $\mathbf{B}_S$, assuming large scale and small scale background errors are

uncorrelated. The incremental 3DVar cost function is minimised with respect to the low resolution model state. The background term of the cost function $\mathcal{J}_b$ is specified with the low resolution background model state and corresponding low resolution background error covariance matrix $\mathbf{B}_L$. This multi-scale data assimilation method accounts for missing smaller scale features, otherwise known as representativity errors, by replacing $\mathbf{R}_0$ with $\mathbf{R}_0 + \mathbf{H}_0 \mathbf{B}_S \mathbf{H}_0{}^T$. The statistics in $\mathbf{R}_0$ are formulated upon the comparison of observations with the true model state vector, where the resolution of both the observations and model state variables are the same. However, in this case, the observational cost function term $\mathcal{J}_{ob}$ compares high resolution observations and a low resolution model state. The addition of $\mathbf{H}_0 \mathbf{B}_S \mathbf{H}_0{}^T$ to the covariance matrix $\mathbf{R}_0$, accounts for the uncertainty in the comparison due to representativity error. Results using multi-scale data assimilation on a mesoscale convective system showed significant improvements to the analysis [79].

We consider representativity error and model error to be separate types of error. We assume representativity error originates from the observation operator and model error originates from the model. However both representativity error and model error can affect the comparison of observations with the model state mapped to observation space. What we can identify from the work of both D. Hodyss and N. Nichols [56] and Li et al. [79] is that when representativity error is present, there is an improper comparison between the observations and the model state and the cost function should be amended to account for this. It was identified both in the work of D. Hodyss and N. Nichols [56] and Li et al. [79] that representativity error could be accounted for by amending the observation error covariance matrix. This gives us the motivation to develop these ideas and amend the observation error covariance matrix used in the cost function for the presence of model error, as opposed to representativity error. The methods we outlined which were developed to account for representativity error have not included model evolution and only used data at one time. However, as we

68

will be accounting for model error, we require model evolution, and therefore we will develop a 4DVar scheme to make use of observations at multiple times.

We refer back to the second aim of this thesis, as defined in Section 1.1, to amend the strong constraint 4DVar method in such a way that the effect model error has on the estimation of the initial conditions is mitigated. Work in Chapter 7 will involve accounting for the improper comparison of observations with the model state, due to model error, where the model error is of a random nature as defined in equation (2.3). We will investigate the effect of replacing the observation error covariance matrix with a combined error covariance matrix that includes not only observation error statistics, but also model error statistics. This derived combined error matrix changes over time to allow for the accumulation of the model errors over the assimilation window. The specification of model error statistics is of a complex nature, as discussed in Section 4.1 and Section 4.2. Therefore in Chapter 7, we develop a method to estimate the combined model error covariance matrix which does not require the explicit specification of the model error statistics.

4.5 Summary

In Section 4.1 we outlined the difficulties in specifying model error statistics for operational models. We recognise the need for a 4DVar method to account for model error without the requirement for the specification of model error statistics. This will be the focus of our work in Chapter 7. Current weak formulations of data assimilation do require the specification of a model error covariance matrix and we do not wish to dismiss the use of these methods. Therefore in Section 4.2 we outlined methods proposed to estimate model error covariance matrices and detailed the problems resulting when using these techniques. We acknowledged that the development of 4DVar diagnostic tools to verify, and if possible help refine, an estimated model

error covariance matrix will be useful and this is the focus of our work in Chapter 6. In Section 4.3 we examined data assimilation methods that have been formulated to compensate for bias present in coupled models with the objective to improve coupled model forecasts. This gave us the idea for the development of a 4DVar method to simultaneously estimate coupling parameters along with the coupled model state in a 4DVar framework, to compensate for model bias, and this work is covered in Chapter 8 of this thesis. Lastly, in Section 4.4 we outlined methods formulated to account for representativity error and discussed the potential development of such methods to account for model error in 4DVar. This gave us the motivation for the formulation of the 4DVar method developed in Chapter 7 to account for random error present in a model. In the next chapter we introduce dynamical models which will be used to test mathematical methods developed within this thesis.

Chapter 5

Dynamical models

In this chapter we introduce two dynamical models which will be used to demonstrate methods we develop in this thesis. In Section 5.1 we present the linear advection equation, followed by an idealized coupled atmosphere-ocean model in Section 5.2. Both of these dynamical systems consist of governing differential equations. We will explain how numerical schemes provide approximate time-stepping solutions to these dynamical systems. Finally, in Section 5.3 of this chapter we will outline certain aspects of the data assimilation set up which will be used in numerical experiments throughout this thesis.

5.1   Linear advection equation

The one dimensional linear advection equation is used to model the transportation of a scalar quantity $u(x, t)$ carried along by a flow with constant speed v in one direction,

$$\frac{\partial u}{\partial t} + v \frac{\partial u}{\partial x} = 0, \tag{5.1}$$

71

where x and t are independent variables representing space and the time respectively. Although this model is far more simplistic than any operational model representing the dynamics of the atmosphere or ocean, the model dynamics (5.1) can be interpreted to represent an idealized description of a passive tracer transported in the atmosphere in one direction by a constant flow, for example water vapour carried along by a constant light breeze. Another reason for the selection of this model is the linear nature. In this thesis we derive diagnostic tools that involve linear model matrices. We wish to demonstrate the use of these tools with a linear model, prior to considering models of a nonlinear nature. The partial differential equation (PDE) (5.1) does not describe the shape of the passive tracer, only its movement in space and time. We use the spacial domain $x \in [0, 10)$. For this problem to have a unique solution, a function f to describe the shape of the passive tracer at the initial time $t_0 = 0$ is required as follows,

$$u(x, 0) = f(x). \tag{5.2}$$

We take the function f from [26] where,

$$f(x) = \begin{cases} \exp\left\{-(x - x_0)^2\right\} & : 2.5 \leq x \leq 7.5 \\ 0 & : \text{otherwise} \end{cases}$$

which is an exponential function centred around $x_0 = 5$.

The PDE (5.1) can be solved analytically, subject to the initial conditions (5.2), with the resulting solution,

$$u(x,t) = f(x - vt). \tag{5.3}$$

As time increases, with $v > 0$ the distribution of u is shifted vt in the positive x direction, however with $v < 0$ the distribution of u is shifted vt in the negative x direction. Note than when $v = 0$ there is no movement in the distribution of u over time. For numerical experiments conducted in this thesis we specify $v = 2$. Operational NWP centres make use of numerical schemes to provide approximate solutions to atmosphere and ocean governing PDEs, resulting in spacially discretized time-stepping models of the form (2.2). To be consistent with this methodology, we apply numerical schemes to provide approximate time-stepping solutions to the differential equations in the models we use.

5.1.1   Numerical approximation scheme

We firstly select the time domain $t \in [0, 1]$. We specify the boundary conditions to be periodic so that the value of u at $x = 10$ is equal to the value of u at $x = 0$. The Crank-Nicolson scheme [29] is an implicit method that computes a second order approximation of the spacial derivative and a first order approximation of the time derivative. We use this unconditionally stable method, with the spacial step $\Delta x = 0.1$ and time-step $\Delta t = 0.1$ resulting in a time-stepping approximation of the PDE (5.1). The analytical solution is very well represented with use of this numerical scheme, as can be seen in Figure 5.1 comparing the analytical solution and numerical solution both near the start of the time window at $t_1 = 0.1$ and at the end of the time window when $t_{10} = 1$. Figure 5.1 clearly shows as time increases the distribution of u is shifted $2t$ in the positive x direction. We next introduce an idealized coupled atmosphere-ocean model.
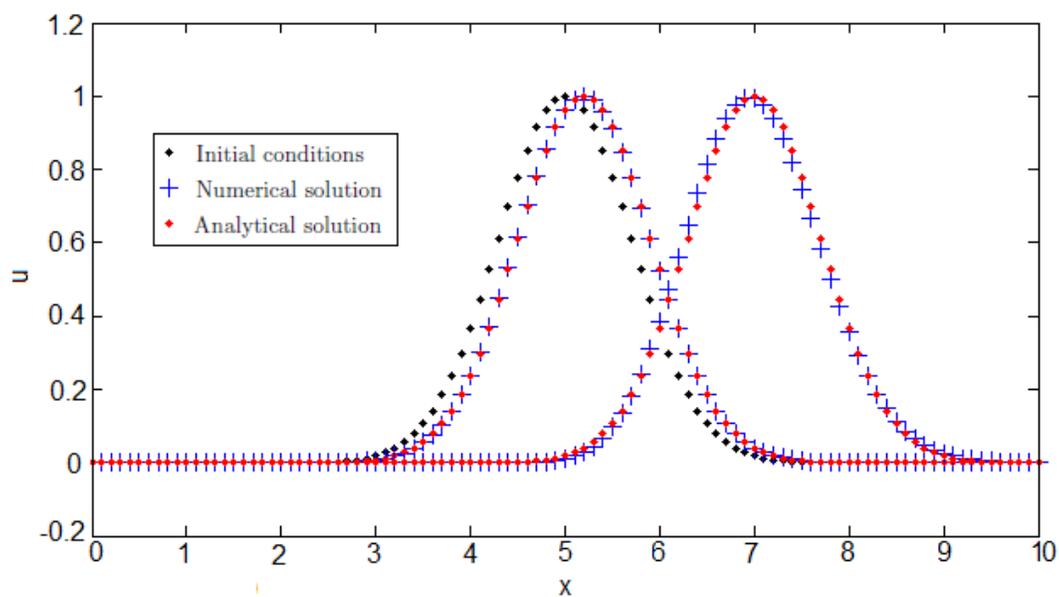
Figure 5.1: Comparison of analytical solution (red dots) and numerical solution (blue crosses) to the linear advection equation at times $t_1 = 0.1$ and $t_{10} = 1$, with the initial conditions shown at time $t_0 = 0$ (black dots). Solutions for u evaluated at every spacial grid step $\Delta x = 0.1$ in the spacial domain $x \in (0, 10)$.

5.2   Idealized coupled atmosphere-ocean model

The idealized coupled atmosphere-ocean model we use in this thesis was developed by Molteni et al. [87] and is given by the equations,

$$
\begin{aligned}
\frac{dx}{dt} &= -\sigma x + \sigma y + \alpha v, \\
\frac{dy}{dt} &= -xz + rx - y + \alpha w, \\
\frac{dz}{dt} &= xy - bz, \\
\frac{dw}{dt} &= -\Omega v - k(w - w^*) - \alpha y, \\
\frac{dv}{dt} &= \Omega(w - w^*) - kv - \alpha x, \tag{5.4}
\end{aligned}
$$

with parameter values $\sigma = 10$, $r = 30$, $b = \frac{8}{3}$, $k = 0.1$, $\Omega = \frac{\pi}{10}$, $w^* = 2$ and coupling parameter $\alpha = 1$ as specified by M.A. Dubois and P. Yiou [42]. With the coupling parameter $\alpha = 0$ the atmosphere is represented by the Lorenz 63 equations [81] with the state variables x, y and z and the ocean is represented by two linear equations with the state variables w and v. However with the coupling parameter $\alpha = 1$ the atmosphere variables x and y are coupled with the ocean variables w and v. The coupled model (5.4) describes the relationship between convection in the atmosphere and the influences that the SST (sea surface temperature) has on the convection in the atmosphere, as interpreted by M.A. Dubois and P. Yiou [42]. Figure 5.2 shows a pictorial representation of the dynamics described by this idealized coupled system, with the convective motion in a layer of fluid interacting with the SST. The atmospheric variables describe properties in a layer of fluid of uniform depth in the atmosphere. The atmospheric state variable x is proportional to the intensity

of convective motion. The temperature difference between ascending and descending currents is proportional to the atmospheric state variable y. The atmospheric state variable z is proportional to the distortion of the vertical temperature profile from linearity. The model variables w and v represent equatorial SST anomalies' influence on the global system. The coupling parameter is represented by α and in this model can be interpreted as heat flux at the atmosphere-ocean interface.

The reasons for choosing this model include firstly, the fact this model is of a nonlinear nature and we wish to perform numerical experiments to demonstrate the ability of methods developed in this thesis to successfully work on systems of a nonlinear nature. Secondly, work within this thesis aims to improve coupled atmosphere-ocean forecasts. Although the model selected is of a much more simplistic nature than those used operationally, key characteristic properties of the atmosphere, ocean and interactions at the interface are present in this model. This will be discussed further in the Section 5.2.2.

5.2.1 Numerical approximation scheme

The model state vector $\mathbf{x}_i$ consists of all atmospheric and oceanic state variables at time t_i. We apply the Runge-Kutta second order scheme [113] to provide a time-stepping approximate solution to the coupled ordinary differential equations (ODEs) (5.4). We let the initial time $t_0 = 0$ and use the fixed time step $\Delta t = 0.01$. Figure 5.3 and Figure 5.4 show plots for the atmospheric variables x and z and ocean variables w and v respectively. In both figures the resulting trajectories are the time-stepping approximate solutions run for 2000 time-steps, therefore a total of 20 time units, initialized using the conditions $x_0 = -3.4866$, $y_0 = -5.7699$, $z_0 = 18.3410$, $w_0 = -10.7175$ and $v_0 = -7.1902$. Figure 5.3 shows the model trajectory of z against x, where as Figure 5.4 shows the model trajectories of w and v respectively against
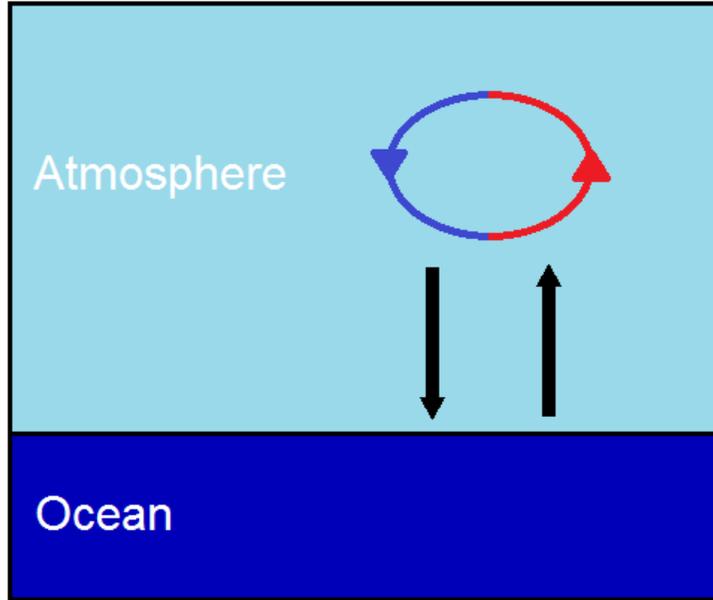
Figure 5.2: Pictorial representation of processes represented in the idealized coupled atmosphere ocean model (5.4). The blue arrow represents descending currents, the red arrow represents ascending currents, which together represent convective motion in a layer of fluid of uniform depth in the atmosphere. The influences the SST anomalies of the ocean have on the atmosphere are represented by the black arrow pointing upwards. The influences the atmospheric convective motion have on the SST anomalies of the ocean are represented by the black arrow pointing downwards

time. The characteristics of the coupled model (5.4) are further analysed in the next section.

5.2.2   Characteristics of the model

Next we describe key characteristic features that the idealized coupled atmosphere-ocean model (5.4) shares with operational coupled atmosphere-ocean dynamical systems used at NWP centres, as described in Section 3.1.1. Although the coupled model we use (5.4) is an idealized model, the distinct nature of atmosphere and ocean variables respectively are represented.

In reality ocean dynamics are on a slower time scale than atmospheric dynamics. For example, due to the large heat capacity of the ocean, the ocean surface temper-
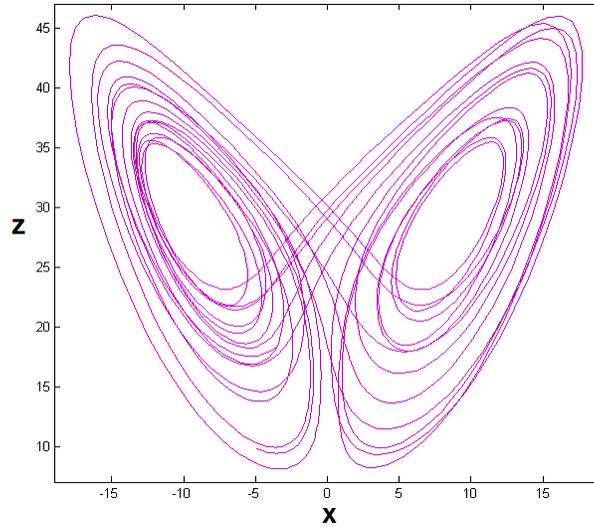
Figure 5.3: Coupled atmosphere: Runge-Kutta second order scheme applied to the coupled model equations (5.4), from the initial time $t_0 = 0$ with the fixed time step $\Delta t = 0.01$. The plot shows the model trajectory of z against x from running the coupled model for 2000 time-steps, therefore a total of 20 time units, from the initial conditions $x_0 = -3.4866$, $y_0 = -5.7699$, $z_0 = 18.3410$, $w_0 = -10.7175$ and $v_0 = -7.1902$.
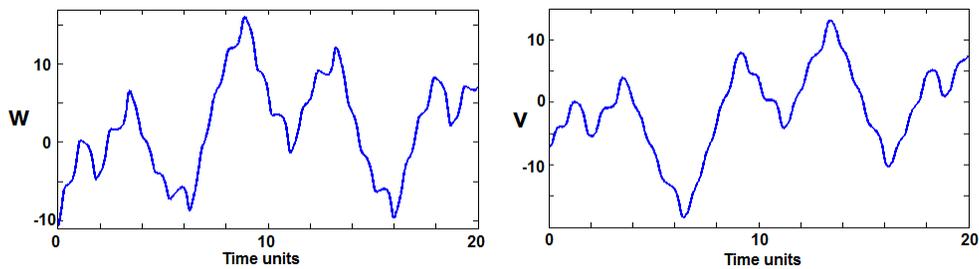
Figure 5.4: Coupled ocean: Runge-Kutta second order scheme applied to the coupled model equations (5.4), from the initial time $t_0 = 0$ with the fixed time step $\Delta t = 0.01$. The plots show the model trajectories for the ocean variables w and v from running the coupled model for 2000 time-steps, therefore a total of 20 time units, from the initial conditions $x_0 = -3.4866$, $y_0 = -5.7699$, $z_0 = 18.3410$, $w_0 = -10.7175$ and $v_0 = -7.1902$.

78

ature changes by much smaller amounts and on a much slower time-scale than that of the land surface [52]. Uncoupled, the atmospheric equations of the coupled model (5.4) form the Lorenz 63 model, which has a transition time between the two sides of the model attractor of one to two time units. Uncoupled, the ocean variables in the coupled model (5.4) are on a much slower time-scale with an oscillation period of 20 units. Figure 5.5 and Figure 5.6 show plots for the atmospheric variables x and z and ocean variables w and v respectively. In both figures the resulting trajectories are from the Runge-Kutta second order scheme applied, with the fixed time step $\Delta t = 0.01$, to the coupled model (5.4) with $\alpha = 0$, therefore uncoupling the atmosphere and ocean models. Figure 5.5 shows the model trajectory of z against x from running the model for 2000 time-steps, therefore a total of 20 time units. Figure 5.6 shows the model trajectories for the ocean variables w and v also from running the model for 2000 time-steps, therefore a total of 20 time units. This creates a realistic time-scale difference between the pace of atmospheric and oceanic model variables. For example, currently ECMWF use a 12 hour assimilation window for the atmosphere data assimilation process and a 10 day ocean data assimilation window, due to the difference in time-scales of the respective atmosphere and ocean [96] [6].

Irregular oscillations are very common in atmospheric systems, this property is reflected in the Lorenz 63 model, with the inclusion of both dissipative processes and external forcing that prevent the system from reaching a state of rest. The use of nonlinear equations in the Lorenz 63 ODEs leads to a non-periodic flow, which is a recognised feature of the true atmospheric dynamics. The Lorenz 63 model is renowned for its chaotic nature, where a small change in the initial conditions can cause significant changes to the trajectories of the model variables. When plotting the evolution of atmospheric variables z against x, an attractor is formed, resembling butterfly wings, as can be seen in Figure 5.5. For more information refer to the work of E. Lorenz [81]. This property is sustained when coupled to the linear ocean
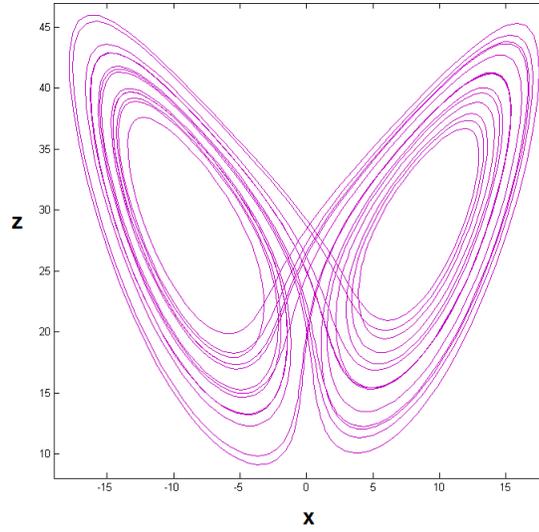
Figure 5.5: Uncoupled atmosphere: Runge-Kutta second order scheme applied to the Lorenz 63 model equations [81], from the initial time $t_0 = 0$ with the fixed time step $\Delta t = 0.01$. The plot shows the model trajectory of z against x from running the model for 2000 time-steps, therefore a total of 20 time units, from the initial conditions $x_0 = -3.4866$, $y_0 = -5.7699$ and $z_0 = 18.3410$. The atmosphere parameter values $\sigma = 10$, $r = 30$ and $b = \frac{8}{3}$ as defined in the coupled model equations (5.4).
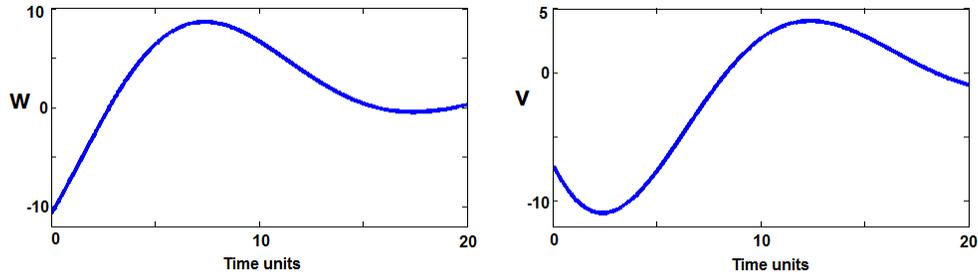
Figure 5.6: Uncoupled ocean: Runge-Kutta second order scheme applied to the linear ocean model equations (5.4) with $\alpha = 0$, from the initial time $t_0 = 0$ with the fixed time step $\Delta t = 0.01$. The plots show the model trajectories for the ocean variables w and v from running the model for 2000 time-steps, therefore a total of 20 time units, from the initial conditions $w_0 = -10.7175$ and $v_0 = -7.1902$. The ocean parameter values $k = 0.1$, $\Omega = \frac{\pi}{10}$ and $w^* = 2$ as defined in the coupled model equations (5.4).

equations in (5.4), as can be seen in Figure 5.3. We can interpret switching between the two sides of the attractor as the change in direction of the convective motion. True ocean dynamics are known to be of a less chaotic nature than atmospheric dynamics, with many large scale features of tropical circulation described by linear systems [87]. Uncoupled, the ocean part of the idealized coupled model (5.4) has non chaotic behaviour, periodically oscillating with decreasing amplitude, as can be seen in Figure 5.6. However once coupled the ocean variables are of a slightly more chaotic nature, as can be seen in Figure 5.4. This is representative of true ocean dynamics, as not all ocean features are of a slow periodic nature.

5.2.3 Sensitivity of an idealized coupled model to a coupling parameter

We next demonstrate that the specification of the coupling parameter α can have a significant impact on the behaviour of both the atmospheric and oceanic model state trajectories over time. We firstly initialise the idealized coupled model (5.4) with $\alpha = 1$ and run over 750 time-steps with the resulting model state trajectories shown in Figure 5.7 (black lines). The same intial conditions have been used as in the previous section. Next we take the same idealized coupled model (5.4) with the same initial conditions and again run the model for 750 time-steps, but alter the value of the coupling parameter to $\alpha = 0.97$ in the first run (turquoise) and secondly $\alpha = 1.03$ in the subsequent run (dark green), with the resultant model state trajectories shown in Figure 5.7. Here we observe that with only small perturbations, namely 3% increase or decrease, to the value of the coupling parameter, the model's sensitivity to the coupling parameter causes significant departures from the model state trajectories for all atmospheric and oceanic variables throughout the time window (in particular subsequent to time-step 300).
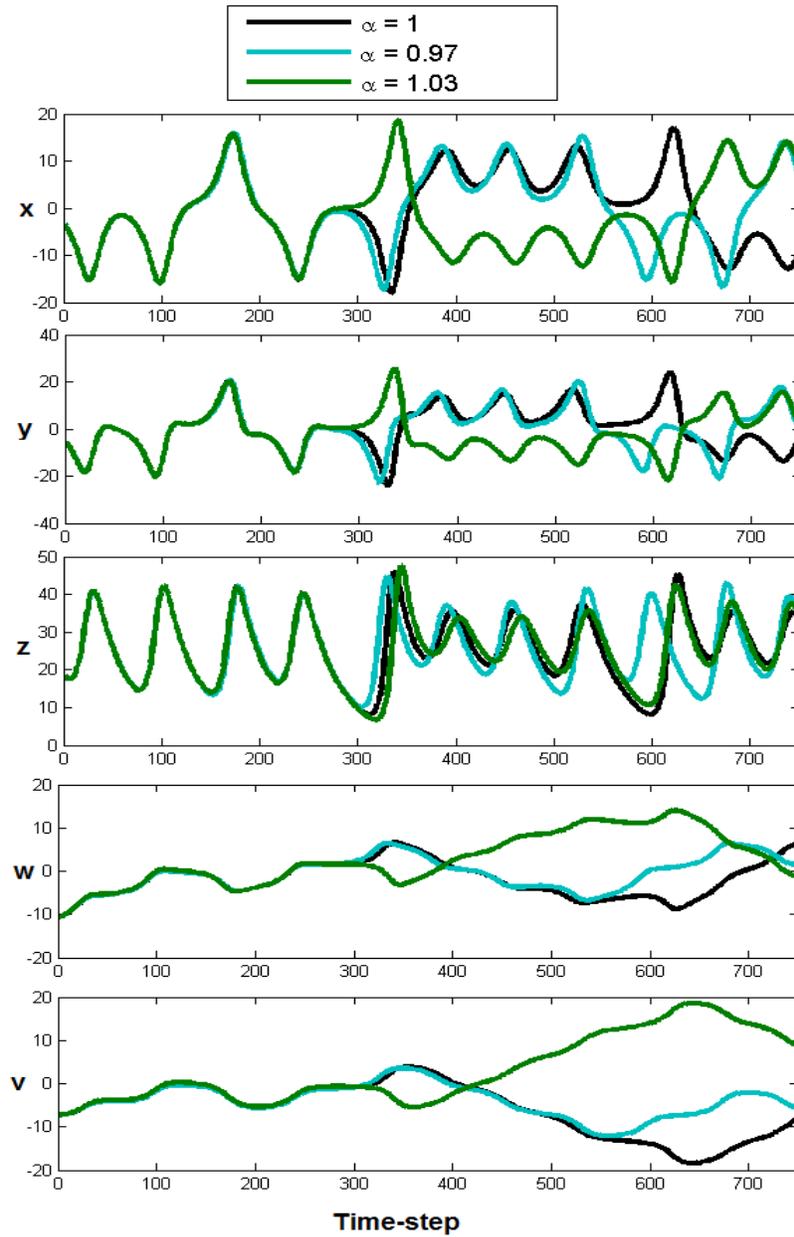
Figure 5.7: Model state trajectories produced using the idealized coupled model with: $\alpha = 1$ (black), $\alpha = 0.97$ (turquoise), $\alpha = 1.03$ (dark green). Model trajectories shown for all atmospheric variables x, y and z and all ocean variables w and v. All three model runs initialised with the same initial conditions $x_0 = -3.4866$, $y_0 = -5.7699$, $z_0 = 18.3410$, $w_0 = -10.7175$ and $v_0 = -7.1902$.

Note that even the trajectory of the atmospheric variable z is dependent on the coupling parameter specification, even though α is not present in the model equation for this particular variable. This is due to the interactions z has with the other atmospheric variables.

5.3 Data assimilation set up

We next define the structure of background error, observation error and model error covariance matrices used in numerical experiments shown in this thesis. Subsequently, we outline the minimisation algorithm selected and then explain how we will evaluate the accuracy of results from experiments conducted in this thesis.

5.3.1 Structure of error covariance matrices

Let $\boldsymbol{\epsilon}_i \in \mathbb{R}^\psi$ be a generic error vector at time t_i. We define the structure of the corresponding generic error covariance matrix $\mathbf{G}_i$ at time t_i as follows,

$$\mathbf{G}_i = \sigma_i{}^2 \mathbf{C}, \tag{5.5}$$

where $\sigma_i{}^2$ is the variance of the elements in $\boldsymbol{\epsilon}_i$ and $\mathbf{C} \in \mathbb{R}^{\psi \times \psi}$ is a correlation matrix. The assumptions made are that the error variances $\sigma_i{}^2$ are equal for every element in the error vector $\boldsymbol{\epsilon}_i$. Since the correlation matrix $\mathbf{C}$ is symmetric positive definite, then $\mathbf{G}_i$ will also have this characteristic, which is necessary as we require the error covariance matrices $\mathbf{B}$, $\mathbf{R}_i$ and $\mathbf{Q}_i$ to be invertible. In work in this thesis we either set $\mathbf{C} = \mathbf{I}$ and hence include no error correlation information, or we set the correlations in $\mathbf{C}$ with the SOAR (second-order auto-regressive) function. The reasons for setting

$\mathbf{C} = \mathbf{I}$ in our work includes the fact that operationally at NWP centres often only the diagonal elements of observation error covariance matrices $\mathbf{R}_i$ are specified, although recent work has involved specification of correlations for specific observations as detailed in Section 3.1.3. Also, the use of $\mathbf{C} = \mathbf{I}$ in our work allows for the results from numerical experiments to be clearly interpreted without the correlation information influencing the results.

Note that the structure of the covariance matrices as described by (5.5) are only used in experiments with the linear advection equation. For experiments with use of the coupled model we do not include correlation information, but do allow variances to differ for each variable.

In this thesis the SOAR correlation function is only used in the construction of error covariance matrices in experiments involving the linear advection equation. Consider the case where we have a total of D grid points. We let the element in the kth row and lth column of the correlation matrix be defined by,

$$\mathbf{C}_{(k,l)} = \rho(x_k, x_l). \tag{5.6}$$

The SOAR function can be used to calculate the correlation between the two grid points x_k and x_l $(k, l = 1, ..., D)$ as follows,

$$\rho(x_k, x_l) = \left(1 + \frac{|r|}{\mathcal{L}}\right) \exp\left\{-\frac{|r|}{\mathcal{L}}\right\}, \tag{5.7}$$

where $\mathcal{L}$ is the correlation length scale and r is the distance between the two grid points [95]. When equation (5.7) is used on a periodic domain the distance of the chord, as opposed to the straight line, is evaluated by $r = 2a\sin\left(\frac{\theta}{2}\right)$, where a is

the radius and θ is the angle between the two points. This is necessary to allow correlations valid on the real line to also be valid on the circle and for a finite domain [125] [129]. Use of the SOAR correlation function (5.7) forms a circulant covariance matrix, with each row the same as the one above except shifted to the right by one element. The correlation matrix $\mathbf{C}$ produced using a SOAR function is an idealised correlation structure, often used at operational centres, such as the Met Office [59], to specify background error correlations in the horizontal.

5.3.2 Data assimilation minimisation algorithm

Whether the variational data assimilation technique chosen is 3DVar, 4DVar, weak constraint 4DVar or any further developed methods in the variational framework, the objective is still to minimise a cost function $\mathcal{J}$ to obtain the respective analysis. There are multiple choices of minimisation algorithms. In this thesis we select the Polak-Ribiere conjugate gradient method [97]. This is a nonlinear conjugate gradient method, which requires an initial estimate of the solution. On each iterative loop, this algorithm evaluates both the cost function $\mathcal{J}$ and the gradient of $\mathcal{J}$ to set the descent direction, towards the minimum of the cost function $\mathcal{J}$, in order to update the estimate. Such conjugate gradient algorithms are used at operational NWP centres as described by M. Fisher [46] for application in atmospheric data assimilation and are used also for the minimisation of cost functions in ocean data assimilation, for example at the Met Office [123]. This Polak-Ribiere conjugate gradient method is an iterative procedure requiring stopping criteria which we will detail for the case where the control vector is $\mathbf{x}_0$. We define $\mathbf{x}_0{}^0$ to be the initial estimate and we define $\mathbf{x}_0{}^k$ to be the estimate of the initial conditions $\mathbf{x}_0$ on the kth iteration. The algorithm will stop when one of two cases occur. Let us define $\| \cdot \|$ to be the L_2 norm. The first of the stopping criteria is when,

$$\frac{\|\nabla \mathcal{J}(\mathbf{x}_0{}^k)\|}{\|\nabla \mathcal{J}(\mathbf{x}_0{}^0)\|} < 10^{-3}. \tag{5.8}$$

The second case is that the minimisation algorithm will perform a maximum of 200 iterations i.e. $k = 200$.

5.3.3 Evaluation of results

We compute the accuracy of the model state trajectories by calculating the RMSE (Root Mean Square Error) between the analysis model run and true model run as follows,

$$RMSE = \sqrt{\frac{\sum\limits_{i=m}^{n} (\mathbf{x}^a{}_i - \mathbf{x}^t{}_i)^2}{n - m + 1}}, \tag{5.9}$$

where m is the time-step at the beginning of the respective assimilation/forecast window and n is the corresponding time-step at the end of the time window under consideration.

5.4 Summary

In this chapter we described two dynamical models that will be used in numerical experiments in this thesis, the linear advection equation and an idealized coupled atmosphere-ocean model. We outlined the use of numerical schemes to provide approximate time-stepping solutions to the respective dynamical models. We identified the characteristics these idealized models share with true atmosphere and ocean dynamics. Later in the chapter we stated that we will either include no correlation

information in background error, observation error and model error covariance matrices used in numerical experiments in this thesis or we will use the SOAR function to produce correlations. We then stated the nonlinear conjugate gradient method which is chosen to be the minimisation algorithm for estimation of the analysis in numerical experiments. Lastly we detailed that RMSE (Root Mean Square Error) will be used to evaluate the accuracy of results in this thesis. The next chapter, is the first of our results chapters, where we develop diagnostic tools to verify the specification of an estimated model error covariance matrix.

Chapter 6

Verifying and refining model error statistics

In Section 3.2 we discussed the use of Desroziers diagnostics [39] as quality checks for the specification of both background error and observation error covariance matrices. Further to this we detailed work undertaken with the ECMWF atmospheric system, which highlighted that caution should be taken when using diagnostic tools with operational models when model error is not accounted for [2]. In this chapter we derive diagnostic equations that account for random model error. Model error statistics are often unknown and hence model error covariance matrices are estimates of the correct statistics. The diagnostic equations we develop can be used as quality checks to verify the consistency of an estimated model error covariance matrix with both background error and observation error covariance matrices in observation space. In Section 6.1 we derive the first set of diagnostics, which are for the case where an erroneous model is used in strong constraint 4DVar. Subsequently we investigate the potential use of these tools to refine an inaccurately estimated model error covariance matrix. In

Section 6.2 we then derive a further set of diagnostic tools, which are for use with the weak constraint formulation of 4DVar.

6.1 Strong constraint 4DVar diagnostics with model error

We now formulate differences in observation space which will later be used to derive diagnostic tools that account for model error. We assume a background model state $\mathbf{x}^b$ is available with the corresponding background error covariance matrix $\mathbf{B}$. We consider a vector of observations $\mathbf{y}$, of the form (2.4), with a corresponding observation error covariance matrix $\mathbf{R}$. The nonlinear observation operator $\mathcal{H}$ takes the model state from state space to observation space. The differences in observation space that we will consider involve two points in time, specifically the time a background model state vector is available and a future point in time when a vector of observations is available. Let us consider an erroneous model of the form,

$$\mathbf{x}_i = \mathcal{M}^e{}_{\{i-1\}\to i}(\mathbf{x}_{i-1}), \tag{6.1}$$

where the erroneous model operator $\mathcal{M}^e{}_{\{i-1\}\to i}$ evolves the model state vector from the background time t_{i-1} to the observation time t_i. To obtain the true model state at time t_i we assume,

$$\mathbf{x}^t{}_i = \mathcal{M}^e{}_{\{i-1\}\to i}(\mathbf{x}^t{}_{i-1}) + \boldsymbol{\eta}_i, \tag{6.2}$$

89

where the vector $\boldsymbol{\eta}_i$ contains random Gaussian distributed model error uncorrelated in time with a zero mean and covariance matrix $\mathbf{Q}_i$.

The following theory holds across a generic time interval with a background model state vector at time t_{i-1} and a vector of observations at time t_i. To simplify notation in this chapter we will denote these respective times by t_0 and t_1 and the erroneous model operator by $\mathcal{M}^e = \mathcal{M}^e_{\{i-1\}\to i}$. We also simplify the model error vector notation by letting $\boldsymbol{\eta} = \boldsymbol{\eta}_i$, as described in (6.2) with the corresponding model error covariance matrix $\mathbf{Q} = \mathbf{Q}_i$.

We assume the tangent linear hypothesis [17] holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}^e$, as described in Section 2.3.1. The linear observation operator $\mathbf{H}$ is the first order term in the expansion of the Taylor series of $\mathcal{H}(\mathbf{x} + \delta\mathbf{x})$ and the tangent linear model $\mathbf{M}^e$ is the first order term in the expansion of the Taylor series of $\mathcal{M}^e(\mathbf{x} + \delta\mathbf{x})$. This allows the evaluation of the explicit incremental strong constraint 4DVar analysis using one 'outer loop' (2.30),

$$\mathbf{x}^a_0 \;=\; \mathbf{x}^b + \mathbf{K}\mathbf{d}^{o*}_b, \tag{6.3}$$

where the gain matrix $\mathbf{K} = \mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T)^{-1}$ and the innovation vector,

$$\mathbf{d}^{o*}_b = \mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)). \tag{6.4}$$

Note that incremental 4DVar with one 'outer-loop' is equivalent to the best linear unbiased estimate (BLUE) method and therefore gives the same analysis solution (6.3). Note that in this case both the gain matrix $\mathbf{K}$ and innovation vector $\mathbf{d}^{o*}_b$

contain the erroneous model matrix $\mathbf{M}^e$. Note that as the error in the model is of a random nature the expected value of the analysis $E[\mathbf{x}^a{}_0]$ is unbiased and therefore equal to the true model state vector.

In Section 2.3.1 we presented four diagnostic equations that assume a perfect model (2.34)-(2.37). Further to this in Section 2.3.2, we presented the corresponding equation for the first of these diagnostic equations (2.34) that now accounted for random model error (2.40). We derive the further three corresponding diagnostic tools in the next section, specially to (2.35)-(2.37), that will account for random error present in a model. The computations will involve differences in observation space that we will now define. All the work conducted in this chapter uses the 4DVar framework. The evaluation of the analysis (6.3) and innovation vector (6.4) allows us to define the following differences in observation space,

$$\mathbf{d}_b^{o*} \;=\; \mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)) \approx \epsilon_{ob} - \mathbf{HM}^e\epsilon_b + \mathbf{H}\boldsymbol{\eta} \tag{6.5}$$

$$\mathbf{d}_b^{a*} \;=\; \mathcal{H}(\mathcal{M}^e(\mathbf{x}^a{}_0)) - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)) \approx \mathbf{HM}^e\mathbf{K}\mathbf{d}_b^{o*} \tag{6.6}$$

$$\mathbf{d}_a^{o*} \;=\; \mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^a{}_0)) \approx (\mathbf{I} - \mathbf{HM}^e\mathbf{K})\mathbf{d}_b^{o*}, \tag{6.7}$$

which assume the tangent linear hypothesis holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}^e$. Note that equations (6.5)-(6.7) are exactly equal when both the observation operator and model equations are of a linear nature. We can now use these differences in observation space (6.5)-(6.7) to derive diagnostic tools that account for random model error.

Firstly, we present the strong constraint 4DVar diagnostic equation that was previously derived [33] [2]. This uses the innovation vector $\mathbf{d}_b^{o*}$ as stated in (6.5). Taking the statistical expectation of the product of innovations leads to,

$$E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T] \quad \approx \quad \mathbf{R} + \mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{H}\mathbf{Q}\mathbf{H}^T, \qquad\qquad (6.8)$$

as stated in equation (2.40) in Section 2.3.2, which has the additional term $\mathbf{H}\mathbf{Q}\mathbf{H}^T$ [33] [2] when compared to the diagnostic calculated with a perfect model (2.34). The term $\mathbf{H}\mathbf{Q}\mathbf{H}^T$ is the model error covariance matrix in observation space and is present due to the presence of model error in the comparison of the observation vector and model evolved background vector in $\mathbf{d}_b^{o*}$. Specifically, the background vector has been evolved with the erroneous model $\mathbf{M}^e$ and hence the comparison of the observation vector and background vector in $\mathbf{d}_b^{o*}$ contains a vector of model error as stated in (6.5). The reason the vector of model error is in observation space in $\mathbf{d}_b^{o*}$ is because the background vector is mapped to observation space in order to be compared to the observation and hence the vector of model error is mapped to observation space. The term $\mathbf{H}\mathbf{Q}\mathbf{H}^T$ in the diagnostic equation (6.8) is computed as the expectation of the product of the model error vector in observation space.

This diagnostic (6.8) provides a consistency check in observation space not only for the background error and observation error covariance matrices $\mathbf{B}$ and $\mathbf{R}$ respectively, but also for the model error covariance matrix $\mathbf{Q}$. We will now re-evaluate the other three consistency diagnostics (2.35)-(2.37) for the case where an erroneous model of the form (6.1) is used in strong constraint 4DVar, as opposed to a perfect model.

The first of the consistency diagnostics we derive uses the analysis as stated in (6.3) and is conducted by taking the following statistical expectation,

$$
\begin{aligned}
E[\mathbf{d}_b^{a*}(\mathbf{d}_b^{o*})^T] \;&\approx\; \mathbf{HM}^e\mathbf{K}E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T], && (6.9)\\
&\approx\; \mathbf{HM}^e\mathbf{K}(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{HQH}^T),\\
&=\; \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\\
&+\; \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1}\mathbf{HQH}^T,\\
&=\; \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\boldsymbol{\Lambda}, && (6.10)
\end{aligned}
$$

where $\boldsymbol{\Lambda}=\mathbf{I}+(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1}\mathbf{HQH}^T$, using the diagnostic equation (6.8). We define this matrix $\boldsymbol{\Lambda}$ to be an inflation matrix. The coressponding diagnostic calculated with a perfect model (2.35) is essentially post multiplied by this inflation matrix $\boldsymbol{\Lambda}$ resulting in (6.10). The entries in the inflation matrix $\boldsymbol{\Lambda}$ depend on the the entries in $\mathbf{R}$, $\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T$ and $\mathbf{HQH}^T$. Specifically the amplitude of the inflation is dependent on the covariances of the errors in the observation vector and the covariances of the errors in the background vector evolved by the model and subsequently mapped to observation space, in comparison with the covariances of the errors in the model (mapped to observation space). The larger the errors in the model are (once mapped to observation space) in comparison with both the errors in the observation vector and the errors in the background vector (evolved by the model and mapped to observation space), the larger the inflation.

We next derive a further consistency diagnostic that accounts for random error present in a model by taking the statistical expectation of the following product of differences in observation space,

$$
\begin{aligned}
E[\mathbf{d}_a^{o*}(\mathbf{d}_b^{o*})^T] \;\approx\;& (\mathbf{I} - \mathbf{HM}^e\mathbf{K})E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T], && (6.11)\\
=\;& (\mathbf{I} - \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1})E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T],\\
=\;& (\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T - \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1}\\
& \times E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T],\\
=\;& \mathbf{R}(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1}E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T],\\
\approx\;& \mathbf{R}(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1}(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{HQH}^T),\\
=\;& \mathbf{R}+\mathbf{R}(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1}\mathbf{HQH}^T,\\
=\;& \mathbf{R}\boldsymbol{\Lambda}, && (6.12)
\end{aligned}
$$

using (6.8), with the inflation matrix $\boldsymbol{\Lambda} = \mathbf{I} + (\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)^{-1}\mathbf{HQH}^T$ as previously defined in (6.10) . The corresponding diagnostic that assumes a perfect model (2.36) can be post multiplied by the inflation matrix $\boldsymbol{\Lambda}$ to obtain the diagnostic equation that accounts for model error (6.12). Similarly, as with the last diagnostic equation that accounted for model error (6.10), the larger the covariances of the model errors (in observation space) in comparison with the observation errors and background errors (evolved with the model and mapped to observation space), the larger the inflation. Note that the sum of this diagnostic $E[\mathbf{d}_a^{o*}(\mathbf{d}_b^{o*})^T]$ (6.11) and the previous diagnostic $E[\mathbf{d}_b^{a*}(\mathbf{d}_b^{o*})^T]$ (6.9) add to make the first of the strong constraint diagnostic equations that accounts for model error $E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T]$ (6.8). This is a property that is shared with the model diagnostics (2.34)-(2.36) that assume a perfect model.

The final diagnostic that we derive in this section is deduced by taking the following expectation of the product of differences in observation space,

$$
\begin{aligned}
E[\mathbf{d}_b^{a*}(\mathbf{d}_a^{o*})^T] &\approx E[\mathbf{HM}^e\mathbf{K}\mathbf{d}^o{}_b^*(\mathbf{d}_a^{o*})^T], \\
&= \mathbf{HM}^e\mathbf{K}\left(E[\mathbf{d}_a^{o*}(\mathbf{d}^o{}_b^*)^T]\right)^T, \\
&\approx \mathbf{HM}^e\mathbf{K}\mathbf{\Lambda}^T\mathbf{R}, \\
&= \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R} \\
&\quad +\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1} \\
&\quad \times\mathbf{HQH}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R}, \\
&= \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R} + \mathbf{\Phi} \quad (6.13)
\end{aligned}
$$

where $\mathbf{\Phi} = \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{HQH}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R}$, using the earlier diagnostic result (6.12). This diagnostic (6.13) that accounts for error present in a model can effectively be obtained by taking the corresponding diagnostic that assumes a perfect model (2.37) and adding $\mathbf{\Phi}$. The term $\mathbf{\Phi}$ is the model error covariance matrix in observation space $\mathbf{HQH}^T$ pre multiplied by $\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}$ and post multiplied by $(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{R}$. The larger the model errors mapped to observation space are, the larger the amplitude of $\mathbf{\Phi}$. However the values in $\mathbf{\Phi}$ are also dependent on the pre and post multiplications. The pre multiplication arises from the effect model error has on the comparison of the analysis with the model evolved background in $\mathbf{d}_b^{a*}$. The larger the background error covariances (evolved using the erroneous model and mapped to observation space) in comparison with the observation errors, the larger the pre multiplication. The post multiplication arises from the effect model error has on the comparison of the observations with the model evolved analysis in $\mathbf{d}_a^{o*}$. The larger the observation error covariances in comparison with the background errors (evolved using the erroneous model and mapped to observation space), the larger the

post multiplication.

Having conducted comparisons of the diagnostics that account for the presence of random model error (6.8), (6.10), (6.12) and (6.13) with those that assume a perfect model (2.34)-(2.37), we have shown differences in all four of the diagnostic results. The model error covariance matrix $\mathbf{Q}$ is present in each of the diagnostic equations that account for random model error (6.8), (6.10), (6.12) and (6.13). The first of the diagnostic equations that accounts for model error (6.8) includes the additional term $\mathbf{HQH}^T$ when compared to the corresponding diagnostic that assumes a perfect model (2.34). However, the other three diagnostic equations (6.10), (6.12) and (6.13) that account for random error in the model include changes from the diagnostics that assume a perfect model (2.35)-(2.37) that involve both the background error covariance matrix $\mathbf{B}$ and observation error covariance matrix $\mathbf{R}$ along with the model error covariance matrix $\mathbf{Q}$. This is because these diagnostic equations use differences in observation space that involve the analysis (6.3), where the analysis has been deduced with use of both the background error and observation error statistics.

There are two key points to make from the work conducted in this section. Firstly, we highlight that caution should be taken if using any of the four diagnostic equations (2.34)-(2.37), that assume a perfect model, to verify the consistency of background error and observation error covariance matrices, if random model error is or could be present. If model error is not accounted for and an erroneous model of the form (6.1) is used with the diagnostic equations that assume a perfect model (2.34)-(2.37), this could lead to verification of both over specified background error variances and over specified observation error variances. For example, in Section 3.2 we discussed experiments that were conducted using the ECMWF atmospheric system, where the diagonal elements of the diagnostic (2.34) that assumes a perfect model were calculated and the left and right-hand sides compared. Assuming a perfect model, the results suggested that the background error variances were underspecified in $\mathbf{B}$ [2].

96

However, the atmospheric model used in the experiments contained model error and this could be the explanation for the resulting inconsistencies in the evaluation of the diagnostic (2.34) [2]. The presence of the additional term $\mathbf{HQH}^T$ in (6.8) is due to the model error, which could possibly have mitigated the inaccuracies in these experiments.

The second key point to be made in this section is that the set of the four diagnostic equations that account for random error in a model (6.8), (6.10), (6.12) and (6.13) are a set of diagnostic tools that can be used to verify the consistency of a model error covariance matrix $\mathbf{Q}$ with both background error and observation error covariance matrices in observation space. We should note that the diagnostic equations (6.8), (6.10), (6.12) and (6.13) rely on accurately specified background error and observation error covariance matrices in order to do this, which is often not the case in operational NWP centres as discussed in Section 3.1.3. In Section 6.1.3 we investigate the effect inaccurately specified background error and observation error statistics have on the diagnostic equations. However, for now we assume $\mathbf{B}$ and $\mathbf{R}$ can be accurately specified. Next we perform numerical experiments to demonstrate the use of the set of diagnostic tools that account for random model error (6.8), (6.10), (6.12) and (6.13) as quality checks for the specification of a model error covariance matrix.

6.1.1 Verifying an estimated model error covariance matrix

In this section we demonstrate how diagnostics can be used to verify the consistency of an estimated model error covariance matrix with both background error and observation error covariance matrices. We take our erroneous model (6.1) to be the time-stepping solution of the linear advection equation, as described in Section 5.1.1, over the spatial domain $x \in [0, 10)$ with the spatial step $\Delta x = 0.1$ and time-step

$\Delta t = 0.1$. The true initial conditions at time t_0 are defined with use of the exponential function f, as in Section 5.1. We run this erroneous model over one time-step from time t_0 to time t_1. The true model state at time t_1 differs from the erroneous model state by a vector of random error $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, as defined by equation (6.2). The structure of the model error covariance matrix $\mathbf{Q}$ is defined with use of the SOAR function, as described in Section 5.3.1, with correlation length scale $\mathcal{L} = 0.1$ and variance $\sigma_q{}^2 = 0.01$. We consider a background model state $\mathbf{x}^b$ at time t_0 with corresponding background error covariance matrix $\mathbf{B}$. We use the SOAR function, with correlation length scale $\mathcal{L} = 0.4$, to assign the correlations in the background error covariance matrix $\mathbf{B}$ and set the variance $\sigma_b{}^2 = 0.04$. We assume we have direct observations $\mathbf{y}$ at time t_1 of all spatial points with the linear observation operator $\mathbf{H} = \mathbf{I}$. The corresponding observation error covariance matrix is $\mathbf{R} = \sigma_{ob}{}^2\mathbf{I}$ where $\sigma_{ob}{}^2 = 0.04$. Note that the model error standard deviation is set to be smaller, specifically half, of both the background error and observation error standard deviations. In general, the standard deviation of random model error is thought to be less than the standard deviation of the error in the background. This is because the error in the background represents short term forecast error over a time window, whereas random model error is present at each individual time-step.

To evaluate the left-hand side of the consistency diagnostics (6.8), (6.10), (6.12) and (6.13), a sample of background vectors, model error vectors, observation vectors and 4DVar analysis vectors are required. We repeat the following steps $1,000$ times to produce a sample size of $1,000$ background vectors, model error vectors, observation vectors and 4DVar analysis vectors:

1. Produce a background vector $\mathbf{x}^b$ by adding noise to the true initial state $\mathbf{x}^t{}_0$ using the statistics specified in $\mathbf{B}$.

2. Produce a random model error vector $\boldsymbol{\eta}$ using the statistics specified in the

98

model error covariance matrix $\mathbf{Q}$.

3. Produce a vector of observations $\mathbf{y}$ by firstly evaluating the true model state vector $\mathbf{x}^t{}_1$ with equation (6.2) and subsequently adding noise in proportion to the error statistics specified in $\mathbf{R}$.

4. Conduct strong constraint 4DVar to produce a corresponding analysis (6.3).

We use the sample data to calculate the expectations as described by the left-hand side of the consistency diagnostics (6.8), (6.10), (6.12) and (6.13) that account for random error.

We evaluate the left-hand side of the consistency diagnostics that assume a perfect model (2.34)-(2.37), but use the erroneous model matrix $\mathbf{M}^e$ as defined above. We subsequently evaluate and subtract the right-hand side of the respective diagnostic equations (2.34)-(2.37) again with the erroneous model matrix $\mathbf{M}^e$. Due to the nature of the model we avoid use of an adjoint model in these RHS calculations by computing the evolution of the background and its transpose. The differences between these results are shown in Figure 6.1 - Figure 6.4 (left). Clearly, when not accounting for error in the model, using the consistency diagnostics that assume a perfect model (2.34)-(2.37), the expectations evaluated on the left-hand side do not match the evaluations of the right-hand side, with large inaccuracies caused by the error in the model. We re-emphasise the point that the diagnostic tools (2.34)-(2.37) that assume a perfect model should not be used to verify the consistency of background error and observation covariance matrices when random error is present in the model. In fact if the correct $\mathbf{B}$ and $\mathbf{R}$ were used in these diagnostic equations (2.34)-(2.37) you would conclude that both $\mathbf{B}$ and $\mathbf{R}$ had been specified incorrectly. As we are using the identity matrix for the observation operator, we in fact obtain an estimate of the model error covariance matrix $\mathbf{Q}$ when evaluating the inaccuracies in the first of the diagnostics, that assumes a perfect model, (2.34) as can be seen in

99

Figure 6.1 (left). We have specifically calculated and plotted the following in Figure 6.1,

$$E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T] - \left(\mathbf{R} + \mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T\right) \approx \mathbf{H}\mathbf{Q}\mathbf{H}^T. \tag{6.14}$$

When the observation operator is equal to the identity matrix, $\mathbf{H} = \mathbf{I}$, this gives us a direct estimate of the model error covariance matrix $\mathbf{Q}$. This estimation in Figure 6.1 (left) has captured the structure of the true model error covariance matrix $\mathbf{Q}$ and the errors in this estimation are random due to sample error. If we were to demonstrate this we would repeat the experiment multiple times and show the sample error is inversely proportional to the sample size.

We next evaluate the right-hand side of the consistency diagnostics that account for random model error (6.8), (6.10), (6.12) and (6.13) and subsequently use the sample data to subtract the calculations from the left-hand side of these consistency diagnostics (6.8), (6.10), (6.12) and (6.13), with the results shown in Figure 6.1 - Figure 6.4 (right). It is clearly shown that, subject to sample error, the left-hand side (LHS) expectations evaluated using the diagnostic equations are equal to the right-hand side (RHS) calculations of the diagnostic equations. The LHS expectations have captured the same structure as the calculations on the RHS, leaving the structure of the sample error to be random. Therefore, we can state that the diagnostics correctly imply the model error covariance matrix is consistent with both the background error covariance matrix and observation error covariance matrix in observation space, which was the expected outcome as we are using the correct model error covariance matrix. Note that the larger the sample size, the smaller the sample error. Here we are using a relatively small sample size of $1,000$ in comparison to the $10,000$ entries we are estimating. We next investigate whether the first of the diagnostic tools (6.8) can be

used to refine an estimated model error covariance matrix.
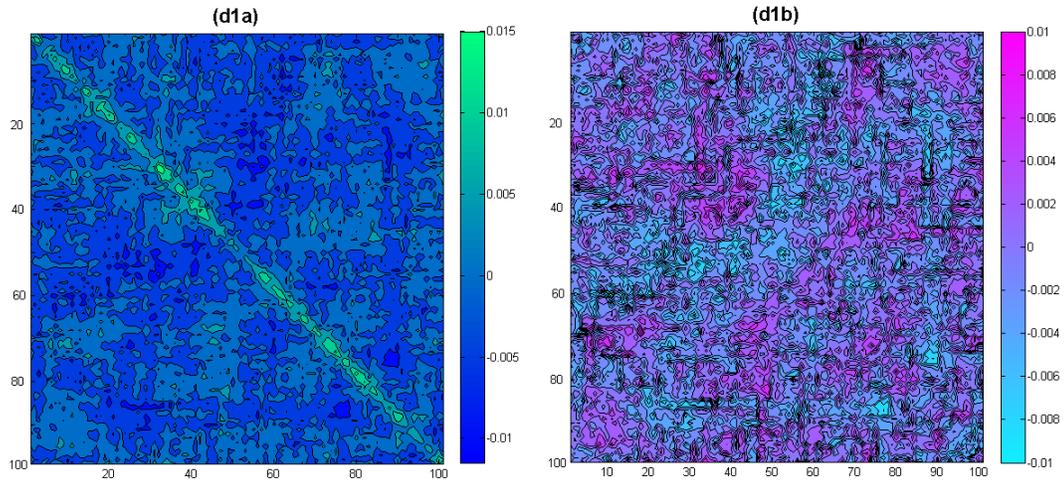
Figure 6.1: Differences in the evaluations of the left-hand side and right-hand side of;
(d1a) diagnostic 1 (2.34) that assumes a perfect model but calculated with an imperfect model,
(d1b) diagnostic 1 (6.8) that accounts for the error in the model.
Note the change of scale on the colour bar axis.
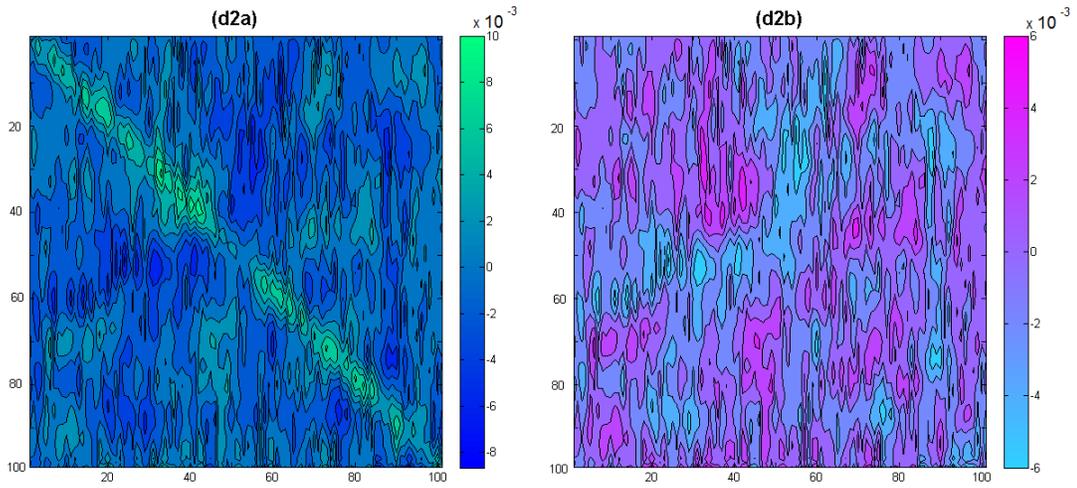
Figure 6.2: Differences in the evaluations of the left-hand side and right-hand side of; **(d2a)** diagnostic 2 (2.35) that assumes a perfect model but calculated with imperfect model, **(d2b)** diagnostic 2 (6.10) that accounts for the error in the model.
Note the change of scale on the colour bar axis.
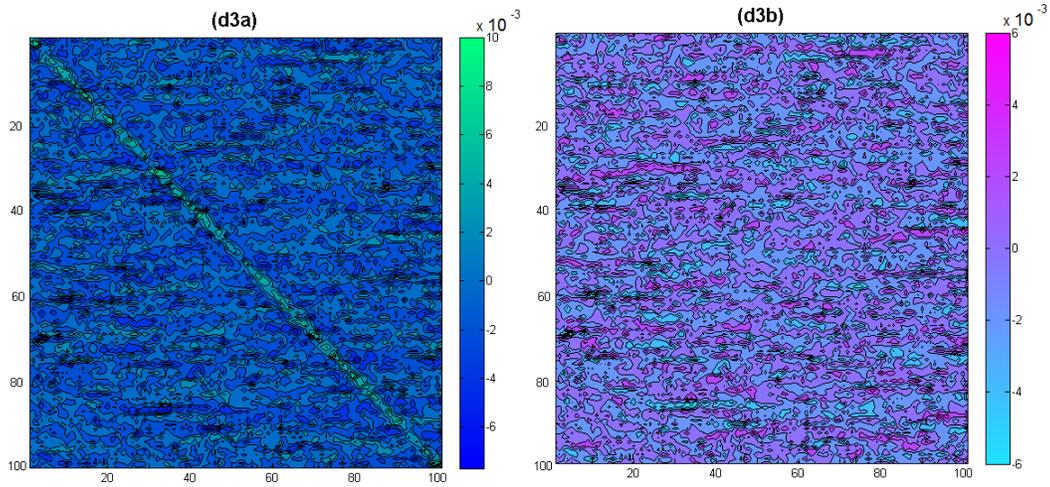
Figure 6.3: Differences in the evaluations of the left-hand side and right-hand side of; **(d3a)** diagnostic 3 (2.36) that assumes a perfect model but calculated with an imperfect model, **(d3b)** diagnostic 3 (6.12) that accounts for the error in the model.
Note the change of scale on the colour bar axis.

102
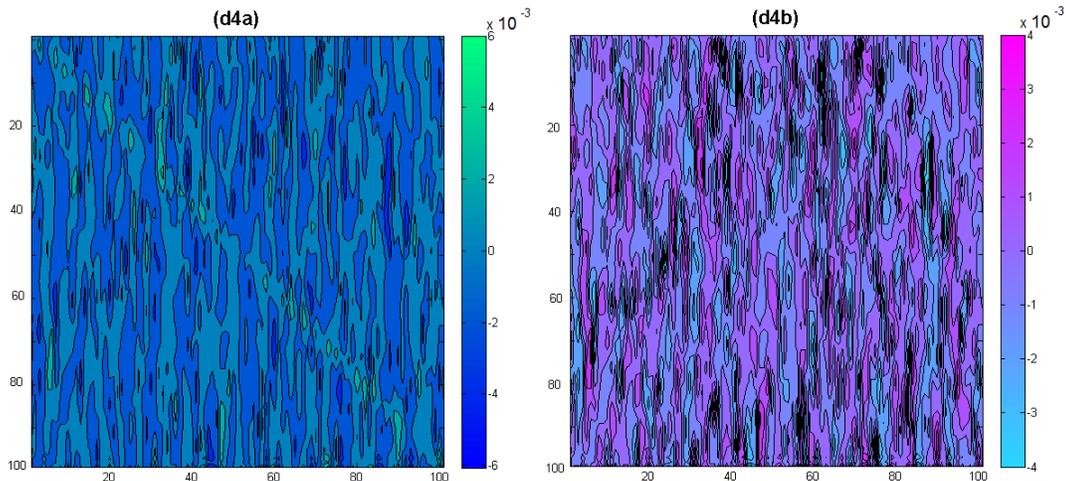
Figure 6.4: Differences in the evaluations of the left-hand side and right-hand side of;
(d4a) diagnostic 4 (2.37) that assumes a perfect model but calculated with an imperfect model
(d4b) diagnostic 4 (6.13) that accounts for the error in the model.
Note the change of scale on the colour bar axis.

6.1.2 Refining an estimated model error covariance matrix

For an invertible observation operator $\mathbf{H}$, specifically an observation operator with full rank, an estimate for the model error covariance matrix $\mathbf{Q}$ can be calculated by simple rearrangement of equation (6.14). However, often observation operators are of a complex nature and the inverse may not be available or even exist, therefore we next investigate how we can refine an estimated model error covariance matrix without requiring the computation of the inverse observation operator.

We define $\widetilde{\mathbf{Q}}$ to be the estimate of the true model error covariance matrix $\mathbf{Q}$. When $\widetilde{\mathbf{Q}}$ is used within the computation of the right-hand side of (6.8) and subsequently the result is subtracted from the left-hand side of (6.8) we obtain,

$$E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T] - \left(\mathbf{R} + \mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T\right),$$

$$\approx \mathbf{R} + \mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{H}\mathbf{Q}\mathbf{H}^T - \left(\mathbf{R} + \mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T\right),$$

$$= \mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T. \tag{6.15}$$

When the estimate of the model error covariance matrix is equal to the true model error covariance matrix, $\widetilde{\mathbf{Q}} = \mathbf{Q}$, the term $\mathbf{H}(\mathbf{Q} - \widetilde{\mathbf{Q}})\mathbf{H}^T$ (6.15) is the null matrix. In this case we can conclude the model error covariance matrix is consistent with both the background error and observation error covariance matrices in observation space. We next investigate how to improve the estimation $\widetilde{\mathbf{Q}}$ when the resulting calculation (6.15) is not a matrix of zeros.

Consider the case where the correlation structure of the model error covariance matrix is accurately known, but the estimate of the variances and covariances differ by the constant factor a, resulting in,

$$\widetilde{\mathbf{Q}} = a\mathbf{Q}, \tag{6.16}$$

where a is a positive scalar $a > 0$. Evaluating the term (6.15) gives,

$$\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T = (1 - a)\mathbf{H}\mathbf{Q}\mathbf{H}^T. \tag{6.17}$$

Therefore, we can conclude if the diagonal entries $(\mathbf{H}(\mathbf{Q} - \widetilde{\mathbf{Q}})\mathbf{H}^T)_{i,i} < 0$ then $a > 1$, so the model error variances are overestimated and if the diagonal entries $(\mathbf{H}(\mathbf{Q} - \widetilde{\mathbf{Q}})\mathbf{H}^T)_{i,i} > 0$ then $a < 1$, so the model error variances are underestimated.

This information can be used to amend the value of a to obtain an improved estimate of the model error covariance matrix $\widetilde{\mathbf{Q}}$.

We demonstrate use of this result with the experimental set up as described in Section 6.1.1, but we only observe every fifth spatial variable with a total of 20 observations. This leads to use of an uninvertible observation operator $\mathbf{H}$. To recap; the structure of the model error covariance matrix $\mathbf{Q}$ is defined with use of the SOAR function with correlation length scale $\mathcal{L} = 0.1$ and variance $\sigma_q{}^2 = 0.01$. We increase the sample size to $10,000$ innovation vectors, where each innovation vector is produced using an independently calculated background vector, model error vector and observation vector as described in Section 6.1.1. We firstly underestimate the model error covariance matrix by setting $a = \frac{1}{2}$ in equation (6.16). The central row (11th) of the estimated model error covariance matrix in observation space $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$ is shown in Figure 6.5. The resulting calculation, using the sample data, of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15) with $\widetilde{\mathbf{Q}} = \frac{1}{2}\mathbf{Q}$ is shown in Figure 6.6 (left). Note that this matrix is of size 20×20 elements as it is in observation space and we are only observing 20 spatial points. This estimated matrix $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15) contains positive values along the diagonal which correctly implies that $a < 1$. The fact that the estimated model error covariance matrix is of the form (6.16), the rows in $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ should be multiples of the rows in $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$, we next examine this property. The other plot shown in Figure 6.6 (right) shows the estimated entries of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15) along the central row (11th). This row is comparable to the plot of the central row of $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$ in Figure 6.5, where the main structure is maintained, but the presence of sample error prevents the plots from being exact multiples of each other.
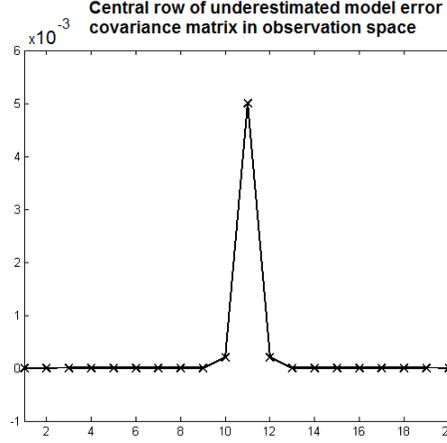
Figure 6.5: The Figure shows the central row (11th) of the estimated model error covariance matrix in observation space $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$, where $\widetilde{\mathbf{Q}}$ is underestimated with $\widetilde{\mathbf{Q}} = \frac{1}{2}\mathbf{Q}$.
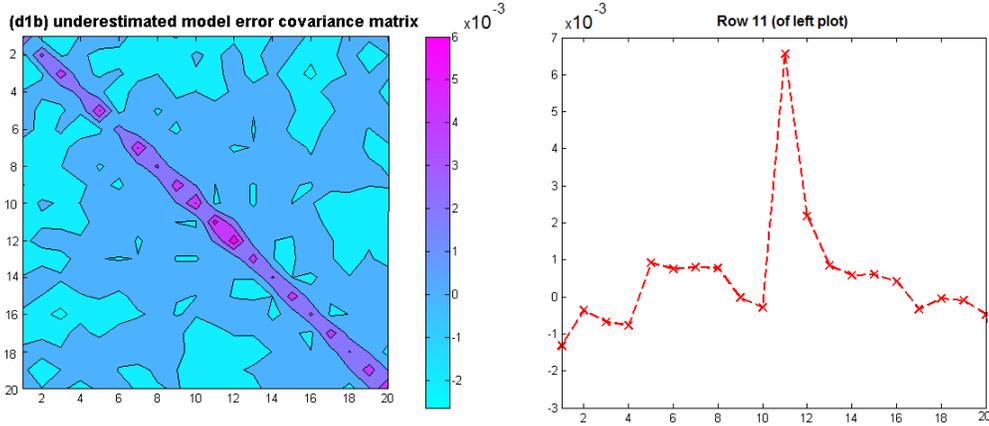
Figure 6.6: The resulting estimation of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15) using $\widetilde{\mathbf{Q}} = \frac{1}{2}\mathbf{Q}$ (left). Note that this matrix is of size 20×20 elements as it is in observation space and we are only observing 20 spatial points. The plot on the right shows the central row (11th) of the estimated entries of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15).

We secondly overestimate the model error covariance matrix with $a = \frac{3}{2}$ in equation (6.16). The central row (11th) of the estimated model error covariance matrix in observation space $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$ is shown in Figure 6.7. The resulting calculation, with the sample data, of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15) with $\widetilde{\mathbf{Q}} = \frac{3}{2}\mathbf{Q}$ is shown in Figure

106

6.8 (left), which contains negative values along the diagonal, this correctly implies that $a > 1$. The fact that the estimated model error covariance matrix is of the form (6.16), the rows in $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ should be multiples of the rows in $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$, we next examine this property. The other plot shown in Figure 6.8 (right) shows the estimated entries of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15) along the central row (11th). This row is comparable to the plot of the central row of $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$ in Figure 6.7, where the main structure is maintained, but the presence of sample error prevents the plots from being exact multiples of each other. Note in this case the multiple factor would be negative due to the overestimation of $\widetilde{\mathbf{Q}}$.
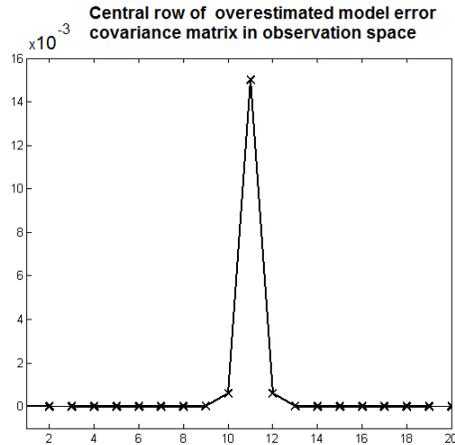
Figure 6.7: The central row (11th) of the estimated model error covariance matrix in observation space $\mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T$, where $\widetilde{\mathbf{Q}}$ is overestimated with $\widetilde{\mathbf{Q}} = \frac{3}{2}\mathbf{Q}$.
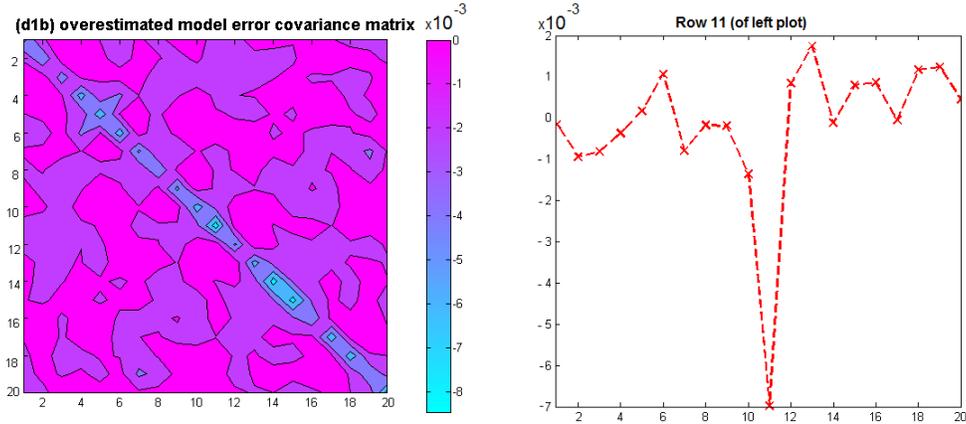
Figure 6.8: The resulting estimation of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15) using $\widetilde{\mathbf{Q}} = \frac{3}{2}\mathbf{Q}$ (left). Note that this matrix is of size 20×20 elements as it is in observation space and we are only observing 20 spatial points. The plot on the right shows the central row (11th) of the estimated entries of the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15).

We have shown that this method can be used to imply whether the variances are under or over estimated, when $\widetilde{\mathbf{Q}}$ is of the form described in equation (6.16). However, even with a simple linear model and with a relatively large sample size of $10,000$ compared to the 400 entries estimated in the term $\mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T$ (6.15), the calculations are affected by sample error and therefore this method should be used with caution.

If the correlation structure of a model error covariance matrix is not known, which is often likely to be the case, it is not possible to refine an estimated model error covariance matrix with use of the diagnostic tools, except when the inverse of the linearised observation operator is available. This motivates work in the next chapter to account for model error in 4DVar without the requirement for an explicit model error covariance matrix.

6.1.3 Inaccurately specified background error, observation error and model error statistics

So far we have considered that both the true background error covariance matrix $\mathbf{B}$ and the true observation error covariance matrix $\mathbf{R}$ are known. However if only estimates $\widetilde{\mathbf{B}}$ and $\widetilde{\mathbf{R}}$ respectively, that contain inaccurate error statistics are available, we need to consider the impact this has on the diagnostic equations. Any diagnostic equation that takes the expectation of differences in observation space involving the analysis is no longer valid. This is because the gain matrix in the analysis, equation (6.3), now contains the estimated matrices $\widetilde{\mathbf{B}}$ and $\widetilde{\mathbf{R}}$ as opposed to the true matrices $\mathbf{B}$ and $\mathbf{R}$. Therefore, when the background error and observation error statistics have been incorrectly specified, the diagnostic equations we derived to account for random error (6.10), (6.12) and (6.13) are no longer valid.

The methods we have developed to refine an estimated model error covariance matrix have included the use of equations (6.14) and (6.15). These equations do not involve the model state analysis. However, caution is required with inaccurately specified background error and observation error covariance matrices, as equation (6.14) changes as follows,

$$
\begin{aligned}
&E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T] - \left(\widetilde{\mathbf{R}} + \mathbf{HM}^e\widetilde{\mathbf{B}}\mathbf{M}^{eT}\mathbf{H}^T \right) \\
\approx\ & \mathbf{R} + \mathbf{HM}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{HQH}^T - \left(\widetilde{\mathbf{R}} + \mathbf{HM}^e\widetilde{\mathbf{B}}\mathbf{M}^{eT}\mathbf{H}^T \right), \\
=\ & \mathbf{R} - \widetilde{\mathbf{R}} + \mathbf{HM}^e(\mathbf{B} - \widetilde{\mathbf{B}})\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{HQH}^T, \qquad (6.18)
\end{aligned}
$$

and equation (6.15) becomes,

109

$$E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T] - \left(\widetilde{\mathbf{R}} + \mathbf{HM}^e\widetilde{\mathbf{B}}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T\right),$$

$$\approx \ \mathbf{R} + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{HQH}^T - \left(\widetilde{\mathbf{R}} + \mathbf{HM}^e\widetilde{\mathbf{B}}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{H}\widetilde{\mathbf{Q}}\mathbf{H}^T\right),$$

$$= \ \mathbf{R} - \widetilde{\mathbf{R}} + \mathbf{HM}^e(\mathbf{B} - \widetilde{\mathbf{B}})\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{H}\left(\mathbf{Q} - \widetilde{\mathbf{Q}}\right)\mathbf{H}^T. \tag{6.19}$$

Therefore equation (6.19) is no longer providing the difference between the true and estimated model error covariance matrices in observation space, due to the inaccuracies in the background error and observation error statistics. The method to refine an estimated model error covariance matrix where the correlation structure is known but the variances are incorrectly specified, as described in Section 6.1.2, is only useful if both $\mathbf{B}$ and $\mathbf{R}$ are correctly known.

We next discuss potential methods to obtain a sample of innovation vectors $\mathbf{d}_b^{o*}$ for the operational use of the diagnostic equation $E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T]$ (6.8).

6.1.4 Potential methods to produce a sample of innovation vectors operationally

Operational centres, such at the Met Office and ECMWF, use ensemble prediction systems that represent random error in the model forecast using stochastic physics [105] [20] [23]. When outlining origins of model error in Section 3.1.2, we discussed that the lack of model resolution leads to atmosphere and ocean models using parameterisations of physical processes that are on scales too small to be directly resolved by the model. These physical processes can be parameterised inadequately or even be absent entirely. Stochastic physics endeavours to account for this by introducing slightly different realizations of the effect of these subgrid-scale physical processes [23]. The effect this has on the ensemble forecast is to increase the spread of the ensemble

forecast members to account for the presence of random error in the model [105]. With this knowledge we present two suggestions to obtain a sample of innovation vectors $\mathbf{d}_b^{o*}$ operationally:

1. The first suggestion we make is to produce a sample of innovation vectors $\mathbf{d}_b^{o*}$ where perturbed background vectors are at time t_{i-1} and perturbed observation vectors are at time t_i. An ensemble prediction system that represents random error in the model forecast, should be used to evolve the sample of background vectors from time t_{i-1} to time t_i. A pictorial representation of this method is shown in Figure 6.9 for a sample size of 4. The model error covariance matrix $\mathbf{Q}$ corresponds to the model error vector $\boldsymbol{\eta}_i$ at time t_i only. This would be a consistent way of calculating the innovation vectors with the theory developed in this thesis if the stochastic physics used represented the total effect of all the random error present in the model. There may not be a sufficient number of observations available at a single time to calculate a sample of innovation vectors $\mathbf{d}_b^{o*}$. Therefore, we also suggest the following method.

2. If the model error covariance matrix is assumed to be constant over a short time period, for example, one hour. Then the sample of observations are not required to be at one time only, but instead within the specified period of time. A pictorial representation of this method is shown in Figure 6.10 for a sample size of 4. The innovation vectors $\mathbf{d}_b^{o*}$ can then be collated within the specified time period to produce a sample. Again, each background vector should be at the time-step prior to the observation vector with the ensemble prediction system used to evolve the background vector to the observation time. A sample of background vectors at these multiple times are assumed available from a previous ensemble forecast. In Section 3.2 we outlined the use of this method with hourly bins to produce operational innovation sample data (in areas where

frequent observations were available) by ECMWF [2]. Therefore, we know that this is an applicable method.

However, there is the consideration that should be noted that there are multiple other possible sources of random model error, as described in Section 3.1.2, that this stochastic physics method used to run an ensemble of forecasts may not account for. Investigation into how to best obtain a sample of innovation vectors $\mathbf{d}_b^{o*}$ operationally for use of the diagnostic equation $E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T]$ (6.8) is an area for further work, as discussed in Section 9.2, of this thesis.
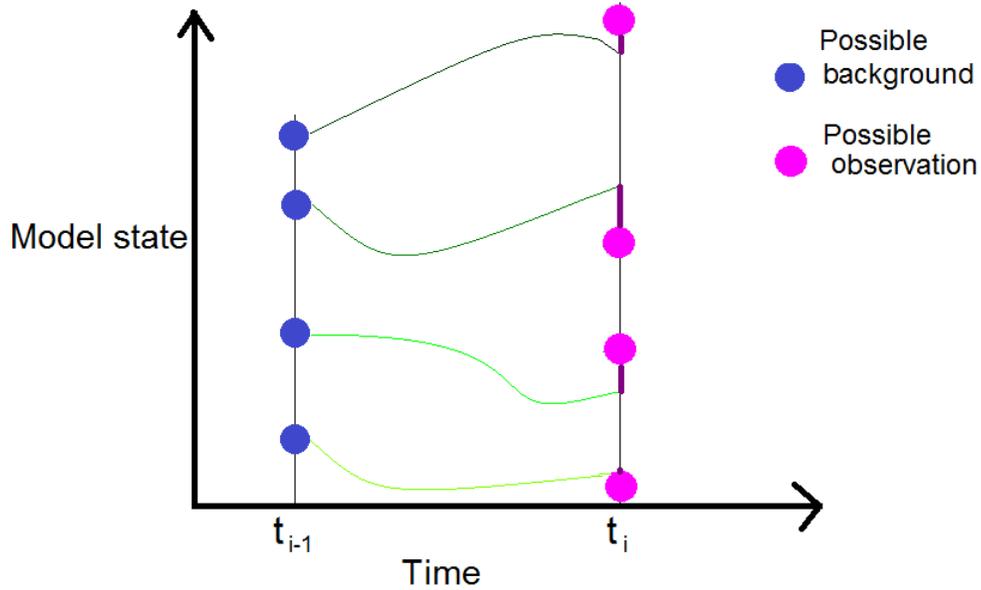
Figure 6.9: Pictorial representation of the first suggestion we make is to produce a sample of innovation vectors $\mathbf{d}_b^{o*}$. The four green lines represent different realizations of the model, therefore accounting for model error. The purple lines represent the innovation vectors.

We propose that it should be feasible to calculate the diagonal elements of the right-hand side of the the diagnostic tool $E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T]$ (6.8) operationally. This would involve use of the randomization technique to estimate the diagonal elements of $\mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T$ [2], as conducted previously in experiments at ECMWF using the
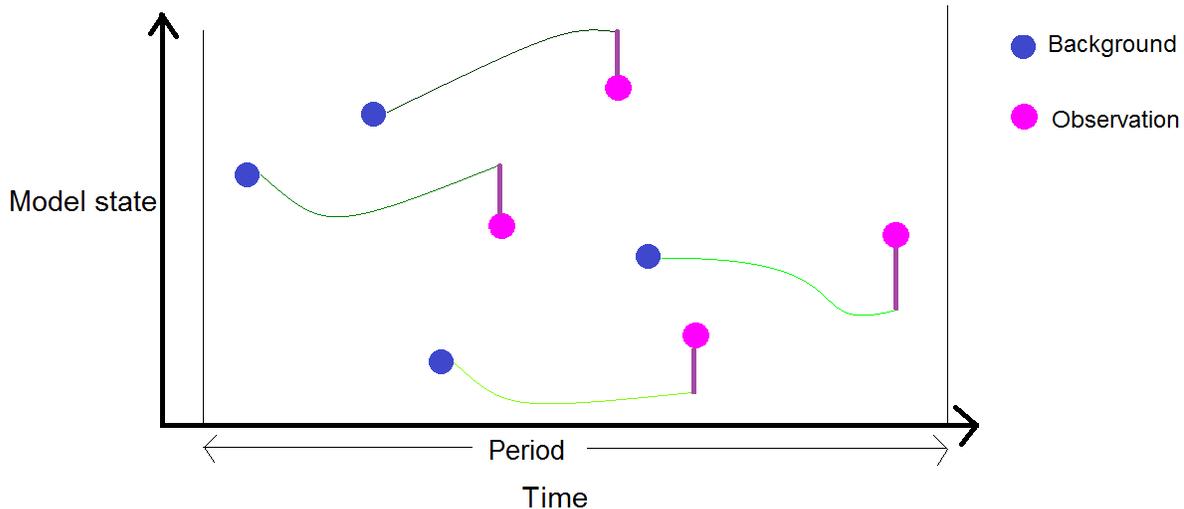
Figure 6.10: Pictorial representation of the second suggestion we make is to produce a sample of innovation vectors $\mathbf{d}_b^{o*}$. The four green lines represent different realizations of the model, therefore accounting for model error. The innovation vectors are computed at multiple times in a selected time period, for example one hour. The purple lines represent the innovation vectors.

atmospheric forecasting model, as described in Section 3.2. We have focused our attention on the methods to obtain a sample of innovation vectors $\mathbf{d}_b^{o*}$ for the operational use of the diagnostic equation $E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T]$ (6.8) as the other three diagnostic equations (6.10), (6.12) and (6.13) require samples of analysis vectors. Therefore, we advise that future work should first concentrate on how to operationally compute the diagnostic equation $E[\mathbf{d}_b^{o*}(\mathbf{d}_b^{o*})^T]$ (6.8), then this methodology can be extended and used with the diagnostic equations that require samples of analysis vectors (6.10), (6.12) and (6.13).

Once a model error covariance matrix has passed the diagnostic quality checks, it can be used for multiple purposes. These include: firstly to perturb members in an ensemble forecast, secondly to identify problem areas in the model and subsequently try to improve the model equations, or thirdly the model error covariance matrix can then be used in the weak constraint formulation of 4DVar. We next derive consistency diagnostics for the latter case, when weak constraint 4DVar is performed.

113

6.2 Weak constraint 4DVar diagnostics

We described a weak constraint formulation of 4DVar in Section 2.2.4 of this thesis, which estimates both the initial model state and the model error vectors at each time-step throughout an assimilation window. We now wish to derive diagnostic tools for this weak constraint 4DVar formulation. We assume a background model state $\mathbf{x}^b$ is available with the corresponding background error covariance matrix $\mathbf{B}$. We consider a vector of observations $\mathbf{y}$, of the form (2.4), with a corresponding observation error covariance matrix $\mathbf{R}$. The nonlinear observation operator $\mathcal{H}$ takes the model state from state space to observation space. An erroneous model, of the form (6.1), is used to take the model state from the background time t_0 to observation time t_1. The random error in the model at time t_1 is the vector $\boldsymbol{\eta}$ with the corresponding model error covariance matrix $\mathbf{Q}$. Using incremental weak constraint 4DVar [118] we can obtain explicit simultaneous equations for the analysis of the model state increment and the model error increment, applying the tangent linear hypothesis to both the model equations and observation operator. The model state analysis for the one time-step case after one 'outer loop' is as follows [118],

$$\mathbf{x}_0^a \;=\; \mathbf{x}^b + \mathbf{K}\mathbf{d}_b^{o\eta}, \tag{6.20}$$

where the gain matrix $\mathbf{K} = \mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T(\mathbf{H}\mathbf{M}^e\mathbf{B}\mathbf{M}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}$ and the innovation vector,

$$\mathbf{d}_b^{o\eta} \;=\; \mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b) + \boldsymbol{\eta}^a), \tag{6.21}$$

114

involves the analysis of the model error $\boldsymbol{\eta}^a$. The corresponding analysis of the model error is given as follows [118],

$$\boldsymbol{\eta}^a = \mathbf{K}^\eta \mathbf{d}_b^{ow}, \tag{6.22}$$

where the gain matrix $\mathbf{K}^\eta = \mathbf{Q}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{Q}\mathbf{H}^T)^{-1}$ and the innovation vector,

$$\mathbf{d}_b^{ow} = \mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^a{}_0)) \tag{6.23}$$

involves the analysis of the model state $\mathbf{x}^a{}_0$ (6.20). In order to obtain an accurate analysis of both the model state and model error, the values specified in the model error covariance matrix $\mathbf{Q}$ are required to be accurate. We next develop diagnostic tools for use as quality checks for the specification of an estimated model error covariance matrix when using the weak constraint formulation of 4DVar.

We take the statistical expectation of the product of innovation vectors (6.21) as follows,

$$
\begin{aligned}
E[\mathbf{d}_b^{o\eta}(\mathbf{d}_b^{o\eta})^T] &= E[(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b) + \boldsymbol{\eta}^a))(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)) + \boldsymbol{\eta}^a)^T], \\
&\approx E[(\epsilon_{ob} - \mathbf{H}\mathbf{M}^e\epsilon_b - \mathbf{H}\epsilon_\eta)((\epsilon_{ob} - \mathbf{H}\mathbf{M}^e\epsilon_b - \mathbf{H}\epsilon_\eta)^T], \tag{6.24}
\end{aligned}
$$

where $\epsilon_\eta = \boldsymbol{\eta}^a - \boldsymbol{\eta}$ is the error in the analysis of the model error vector, with true model error vector $\boldsymbol{\eta}$. The derivation of (6.24) assumes the tangent linear hypothesis holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}^e$. Note that the equation (6.24) is exactly equal when both the observation operator and

model equations are of a linear nature. For the expectation (6.24) to be evaluated, we require the statistics of ϵ_η. We firstly rearrange the equation for the model error analysis (6.22) so that it is independent of the model state analysis (6.20),

$$
\begin{aligned}
\boldsymbol{\eta}^a &= \mathbf{K}^\eta \mathbf{d}_b^{ow} = \mathbf{K}^\eta (\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^a{}_0))), \\
&= \mathbf{K}^\eta (\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b + \mathbf{K}\mathbf{d}_b^{o\eta}))), \\
&\approx \mathbf{K}^\eta (\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)) - \mathbf{HM}^e\mathbf{K}\mathbf{d}_b^{o\eta}), \\
&= \mathbf{K}^\eta (\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b))) - \mathbf{K}^\eta \mathbf{HM}^e\mathbf{K}(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b) + \boldsymbol{\eta}^a)), \\
&\approx \mathbf{K}^\eta (\mathbf{I} - \mathbf{HM}^e\mathbf{K})(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b))) + \mathbf{K}^\eta \mathbf{HM}^e\mathbf{KH}\boldsymbol{\eta}^a, \quad (6.25)
\end{aligned}
$$

assuming the tangent linear hypothesis holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}^e$ for small perturbations around the background model state. With simple rearrangement of equation (6.25), we present the analysis of the model error as follows,

$$
\boldsymbol{\eta}^a \approx \mathbf{L}(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b))). \quad (6.26)
$$

where $\mathbf{L} = (\mathbf{I} - \mathbf{K}^\eta \mathbf{HM}^e\mathbf{KH})^{-1}\mathbf{K}^\eta(\mathbf{I} - \mathbf{HM}^e\mathbf{K})$. We next take the expectation of the error in the model error analysis as follows,

$$
\begin{aligned}
E[\epsilon_\eta] &= \ <\boldsymbol{\eta}^a - \boldsymbol{\eta}>=<\boldsymbol{\eta}^a> \\
&\approx \ <\mathbf{L}(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)))>, \\
&\approx \ \mathbf{L} < \epsilon_{ob} - \mathbf{HM}^e\epsilon_b + \mathbf{H}\boldsymbol{\eta} > \\
&= \ \mathbf{0}, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (6.27)
\end{aligned}
$$

assuming a zero mean for the errors in the background, errors in the observations and random model error. Next we calculate the covariance of the errors in the model error analysis,

$$
\begin{aligned}
\mathbf{D} &= \ E[\epsilon_\eta(\epsilon_\eta)^T], \\
&\approx \ E[(\mathbf{L}(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b))) - \boldsymbol{\eta})(\mathbf{L}(\mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b))) - \boldsymbol{\eta})^T], \\
&\approx \ E[(\mathbf{L}(\epsilon_{ob} - \mathbf{HM}^e\epsilon_b + \mathbf{H}\boldsymbol{\eta}) - \boldsymbol{\eta})(\mathbf{L}(\epsilon_{ob} - \mathbf{HM}^e\epsilon_b + \mathbf{H}\boldsymbol{\eta}) - \boldsymbol{\eta})^T], \\
&= \ E[(\mathbf{L}\epsilon_{ob} - \mathbf{LHM}^e\epsilon_b + (\mathbf{LH} - \mathbf{I})\boldsymbol{\eta})(\mathbf{L}\epsilon_{ob} - \mathbf{LHM}^e\epsilon_b + (\mathbf{LH} - \mathbf{I})\boldsymbol{\eta})^T], \\
&= \ \mathbf{LRL}^T + \mathbf{LHM}^e\mathbf{BM}^{eT}\mathbf{H}^T\mathbf{L}^T + (\mathbf{LH} - \mathbf{I})\mathbf{Q}(\mathbf{LH} - \mathbf{I})^T, \\
&= \ \mathbf{L}(\mathbf{R} + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T)\mathbf{L}^T + (\mathbf{LH} - \mathbf{I})\mathbf{Q}(\mathbf{LH} - \mathbf{I})^T, \quad (6.28)
\end{aligned}
$$

assuming the background error, observation error and model error are all uncorrelated with each other. Having derived the error statistics of the term ϵ_η, we next continue to derive the consistency diagnostic (6.24) as follows,

$$
\begin{aligned}
E[\mathbf{d}_b^{o\eta}(\mathbf{d}_b^{o\eta})^T] \;\approx\;& E[(\epsilon_{ob} - \mathbf{HM}^e\epsilon_b - \mathbf{H}\epsilon_\eta)((\epsilon_{ob} - \mathbf{HM}^e\epsilon_b - \mathbf{H}\epsilon_\eta)^T], \\
=\;& \mathbf{R} + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{HDH}^T \\
&- E[\epsilon_{ob}(\mathbf{H}\epsilon_\eta)^T] + E[\mathbf{HM}^e\epsilon_b(\mathbf{H}\epsilon_\eta)^T] \\
&- E[\mathbf{H}\epsilon_\eta(\epsilon_{ob})^T] + E[\mathbf{H}\epsilon_\eta(\mathbf{HM}^e\epsilon_b)^T], \\
=\;& \mathbf{R} + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{HDH}^T \\
&- \mathbf{RL}^T\mathbf{H}^T - \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\mathbf{L}^T\mathbf{H}^T \\
&- \mathbf{HLR} - \mathbf{HLHM}^e\mathbf{BM}^{eT}\mathbf{H}^T,
\end{aligned}
\tag{6.29}
$$

using $\epsilon_\eta \approx \mathbf{L}\epsilon_{ob} - \mathbf{LHM}^e\epsilon_b + (\mathbf{LH} - \mathbf{I})\boldsymbol{\eta}$ and where the covariance matrix $\mathbf{D}$ is as defined by equation (6.28). When the erroneous model operator $\mathcal{M}^e$ is used in strong constraint 4DVar, the corresponding diagnostic equation is defined by (6.8). Comparing this strong constraint 4DVar diagnostic equation (6.8) with the weak constraint 4DVar diagnostic equation we have just derived (6.29) we observe that both diagnostic equations share the same first two terms, however the third term in (6.8) is $\mathbf{HQH}^T$, compared to the third term in (6.29) of $\mathbf{HDH}^T$, where $\mathbf{D}$ is the covariance of the errors in the model error analysis. This is because the model error is not accounted for in the strong constraint 4DVar data assimilation process, but the model error is estimated in the weak constraint formulation of 4DVar. The weak constraint 4DVar diagnostic equation (6.29) also has an additional four terms, when compared to the strong constraint 4DVar diagnostic equation (6.8), which are due to the correlations the error in the model error analysis has with the error in the background model state and error in the observation vector. This equation (6.29) provides a consistency check for the model error covariance matrix with both the background error and observation error covariance matrices in observation space.

This diagnostic equation (6.29) involves the analysis of the model error and so also implicitly depends on the analysis of the model state, unlike the corresponding strong constraint 4DVar diagnostic equation (6.8).

We aim to derive the corresponding four diagnostic tools for weak constraint 4DVar, to those presented in the last section for strong constraint 4DVar (6.8), (6.10), (6.12) and (6.13). Therefore, next we define the following differences in observation space,

$$
\begin{aligned}
\mathbf{d}_b^{a\eta} &= \mathcal{H}(\mathcal{M}^e(\mathbf{x}^a{}_0)) - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^b)) \approx \mathbf{HM}^e\mathbf{K}\mathbf{d}_b^{o\eta}, && (6.30) \\
\mathbf{d}_a^{o\eta} &= \mathbf{y}_1 - \mathcal{H}(\mathcal{M}^e(\mathbf{x}^a{}_0) + \boldsymbol{\eta}^a) \approx (\mathbf{I} - \mathbf{HM}^e\mathbf{K})\mathbf{d}_b^{o\eta}, && (6.31)
\end{aligned}
$$

where the analysis of the model state is as described by equation (6.20). The approximations in equations (6.30) and (6.31) assume the tangent linear hypothesis holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}^e$ for small perturbations around the background model state. We derive the second of our weak constraint consistency diagnostics by taking the following statistical expectation,

$$
\begin{aligned}
E[\mathbf{d}_b^{a\eta}(\mathbf{d}_b^{o\eta})^T] \;\approx\;& \mathbf{HM}^e\mathbf{K}E[\mathbf{d}_b^{o\eta}(\mathbf{d}_b^{o\eta})^T], & (6.32)\\[4pt]
\approx\;& \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\\[2pt]
& \times(\mathbf{R} + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{HDH}^T\\[2pt]
-\;& \mathbf{RL}^T\mathbf{H}^T - \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\mathbf{L}^T\mathbf{H}^T\\[2pt]
-\;& \mathbf{HLR} - \mathbf{HLHM}^e\mathbf{BM}^{eT}\mathbf{H}^T),\\[2pt]
=\;& \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\\[2pt]
+\;& \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{HDH}^T\\[2pt]
-\;& \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}\\[2pt]
& \times(\mathbf{RL}^T\mathbf{H}^T + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\mathbf{L}^T\mathbf{H}^T\\[2pt]
& +\mathbf{HLR} + \mathbf{HLHM}^e\mathbf{BM}^{eT}\mathbf{H}^T), & (6.33)
\end{aligned}
$$

using (6.29). In Section 6.1 we derived the corresponding diagnostic equation (6.10) for when the erroneous model operator $\mathcal{M}^e$ is used within strong constraint 4DVar. Comparing (6.33) with (6.10) we identify the presence of the covariance of the errors in the model error analysis $\mathbf{D}$ in the place of the model error covariance matrix $\mathbf{Q}$. This is because in this formulation of weak constraint 4DVar the model error is estimated with error covariance $\mathbf{D}$, as opposed to when using strong constraint 4DVar where the model error, with covariance matrix $\mathbf{Q}$, is not estimated. Further to this are another four terms present in the weak constraint 4DVar diagnostic (6.33), as opposed to the strong constraint 4DVar diagnostic (6.10), due to the fact the model error analysis is dependent on both the error statistics of the background and the error statistics of the observations. We next derive the third of the weak constraint 4DVar diagnostics as follows,

$$
\begin{aligned}
E[\mathbf{d}_a^{o\eta}(\mathbf{d}_b^{o\eta})^T] \;\approx\;& (\mathbf{I} - \mathbf{HM}^e\mathbf{K})E[\mathbf{d}_b^{o\eta}(\mathbf{d}_b^{o\eta})^T], && (6.34) \\
\approx\;& (\mathbf{I} - \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R})^{-1}) \\
& \times (\mathbf{R} + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{HDH}^T \\
& -\;\; \mathbf{RL}^T\mathbf{H}^T - \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\mathbf{L}^T\mathbf{H}^T \\
& -\;\; \mathbf{HLR} - \mathbf{HLHM}^e\mathbf{BM}^{eT}\mathbf{H}^T), && (6.35)
\end{aligned}
$$

using (6.29). Equation (6.34) enables us to note that the summation of this diagnostic $E[\mathbf{d}_a^{o\eta}(\mathbf{d}_b^{o\eta})^T]$ and the second of the weak constraint diagnostics $E[\mathbf{d}_b^{a\eta}(\mathbf{d}_b^{o\eta})^T]$ (6.33) add to make the first of the weak constraint diagnostics $E[\mathbf{d}_b^{o\eta}(\mathbf{d}_b^{o\eta})^T]$ (6.29). This property is shared with the strong constraint 4DVar diagnostics as can be seen in Section 6.1. We finally derive the fourth weak constraint 4DVar diagnostic by taking the following statistical expectation of the differences in observation space,

$$
\begin{aligned}
E[\mathbf{d}_b^{a\eta}(\mathbf{d}_a^{o\eta})^T] \;\approx\;& \mathbf{HM}^e\mathbf{K}E[(\mathbf{d}_b^{o\eta})(\mathbf{d}_b^{o\eta})^T](\mathbf{I}-\mathbf{HM}^e\mathbf{K})^T, \\
\approx\;& \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{R})^{-1} \\
& \times(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{HDH}^T \\
-\;& \mathbf{RL}^T\mathbf{H}^T-\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\mathbf{L}^T\mathbf{H}^T \\
-\;& \mathbf{HLR}-\mathbf{HLHM}^e\mathbf{BM}^{eT}\mathbf{H}^T) \\
& \times(\mathbf{I}-(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T), \\
=\;& \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{R})^{-1} \\
& \times(\mathbf{R}+\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{HDH}^T \\
-\;& \mathbf{RL}^T\mathbf{H}^T-\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T\mathbf{L}^T\mathbf{H}^T \\
-\;& \mathbf{HLR}-\mathbf{HLHM}^e\mathbf{BM}^{eT}\mathbf{H}^T) \\
& \times(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{R},
\end{aligned}
\tag{6.36}
$$

using (6.29). In Section 6.1 we derived the corresponding diagnostic equation (6.13) for the case where the erroneous model operator $\mathcal{M}^e$ is used in strong constraint 4DVar. Comparing (6.36) with (6.13) we identify the presence of $\mathbf{D}$ in the place of $\mathbf{Q}$, which is because the model error is estimated in this formulation of weak constraint 4DVar, but is not estimated in strong constraint 4DVar. Further to this there are four extra terms present in the weak constraint 4DVar diagnostic (6.33), as opposed to the strong constraint 4DVar diagnostic (6.13), that are present due to the fact the error in the analysis of the model error is correlated with both the error in the background model state and error in the observations.

Obviously, as with the strong constraint 4DVar diagnostics that account for model error, the weak constraint 4DVar diagnostics (6.29), (6.33), (6.35) and (6.36) only hold

when $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{Q}$ are specified accurately. We next detail when we would advise use of the weak constraint 4DVar diagnostics and demonstrate use of these diagnostics as quality checks for an estimated model error covariance matrix with a simple erroneous model.

6.2.1 Verifying an estimated model error covariance matrix

As we used the same erroneous model (6.1) over a single time-step in the derivation of both the strong and weak constraint diagnostic equations, the model error covariance matrix $\mathbf{Q}$ present in the strong constraint 4DVar diagnostics (6.8), (6.10), (6.12) and (6.13) is of course the same $\mathbf{Q}$ present in the weak constraint 4DVar diagnostics (6.29), (6.33), (6.35) and (6.36). The strong constraint 4DVar diagnostics are also easier to compute than the weak constraint 4DVar diagnostics. Therefore, we need to explain when we would advise the weak constraint 4DVar diagnostics to be used. When the data assimilation method in operation for a particular system is the weak constraint formulation of 4DVar, the weak constraint 4DVar diagnostic equations (6.29), (6.33), (6.35) and (6.36) not only provide quality checks for the consistency of an estimated model error covariance matrix $\widetilde{\mathbf{Q}}$ with $\mathbf{B}$ and $\mathbf{R}$ (in observation space), but also quality checks for the performance of the weak constraint 4DVar scheme. The innovation vectors $\mathbf{d}_b^{o\eta}$ (6.21), $\mathbf{d}_b^{a\eta}$ (6.30) and $\mathbf{d}_a^{o\eta}$ (6.31) are used in the weak constraint 4DVar diagnostic equations (6.29), (6.33), (6.35) and (6.36) and are dependent on both the analysis of the model state $\mathbf{x}^a{}_0$ and the analysis of the model error $\boldsymbol{\eta}^a$. Therefore, the weak constraint 4DVar diagnostics provide quality checks on whether the particular implementation of the weak constraint 4DVar method is successfully minimising the cost function to obtain the optimal estimate of the model state analysis and model error analysis. The diagnostics also provide quality checks on other things like the appropriateness of the tangent linear hypothesis, and the 'Gaussian-ness' of

the data. For example, the minimisation of the weak constraint 4DVar cost function is performed using a minimisation algorithm which requires specification of a stopping criteria to determine how many iterations the minimisation algorithm will perform. If this stopping criteria does not allow enough iterations to be performed for the minimum of the weak constraint cost function to be reached, the optimal analysis of both the model state and model error will not be obtained. In this case, even with accurately specified error covariance matrices $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{Q}$, the weak constraint 4DVar diagnostics will not hold and this indicates that the weak constraint 4DVar implementation is not optimal.

When the 4DVar formulation in operation is the weak formulation, the analysis from this formulation cannot be used in the strong constraint 4DVar diagnostics. Therefore, if the diagnostic required to be calculated involves the analysis, this is another reason why the weak constraint 4DVar diagnostics are of use.

We now demonstrate the use of the weak constraint 4DVar diagnostics (6.29), (6.33), (6.35) and (6.36) to provide quality checks for the specification of an estimated model error covariance matrix $\widetilde{\mathbf{Q}}$. We use the experimental set up as defined at the start of Section 6.1.1 where our erroneous model (6.1) is the time-stepping solution of the linear advection equation over the spatial domain $x \in [0, 10)$, with the spatial step $\Delta x = 0.1$ and time-step $\Delta t = 0.1$, as described in Section 5.1.1. The true initial conditions are defined, as in Section 5.1, with the exponential function f. We run the model over one time-step from time t_0 to time t_1. The true model state at time t_1 differs from the erroneous model state by random error $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, as defined by equation (6.2). The model error covariance matrix $\mathbf{Q}$ is defined with use of the SOAR function, with correlation length scale $\mathcal{L} = 0.1$ and variance $\sigma_q{}^2 = 0.01$. We also use the SOAR function, with correlation length scale $\mathcal{L} = 0.4$, to assign the correlations in the background error covariance matrix $\mathbf{B}$ and set the variance $\sigma_b{}^2 = 0.04$. We assume we have direct observations $\mathbf{y}$ (linear observation operator $\mathbf{H} = \mathbf{I}$) at time

124

t_1 of all spatial points, with diagonal observation error covariance matrix $\mathbf{R} = \sigma_{ob}{}^2\mathbf{I}$ where $\sigma_{ob}{}^2 = 0.04$.

To evaluate the left-hand side of the weak constraint 4DVar consistency diagnostics (6.29), (6.33), (6.35) and (6.36), a sample of innovation vectors $\mathbf{d}_b^{o\eta}$, $\mathbf{d}_b^{a\eta}$ and $\mathbf{d}_a^{o\eta}$ are required. We repeat the following steps $1,000$ times to produce a sample size of $1,000$ background vectors, model error vectors, observation vectors and weak constraint 4DVar model state analysis and model error analysis vectors:

1. Produce a background vector $\mathbf{x}^b$ by adding noise to the true initial state $\mathbf{x}^t{}_0$ using the statistics specified in $\mathbf{B}$.

2. Produce a random model error vector $\boldsymbol{\eta}$ using the statistics specified in the model error covariance matrix $\mathbf{Q}$.

3. Produce a vector of observations $\mathbf{y}$ by firstly evaluating the true model state vector $\mathbf{x}^t{}_1$ with equation (6.2) and subsequently adding noise in proportion to the error statistics specified in $\mathbf{R}$.

4. Conduct weak constraint 4DVar to produce a corresponding model state analysis and model error analysis.

We evaluate the left-hand side (LHS) of the weak constraint 4DVar diagnostic equations (6.29), (6.33), (6.35) and (6.36) using this sample data and then subsequently compute and subtract the right-hand side (RHS) of the diagnostic equations (6.29), (6.33), (6.35) and (6.36) respectively. The RHS weak constraint 4DVar diagnostic calculations are conducted with the correct error covariance matrices $\mathbf{B}$, $\mathbf{R}$ and $\mathbf{Q}$ as just defined. Therefore our estimated model error covariance matrix $\widetilde{\mathbf{Q}}$ is equal to the true model error covariance matrix $\mathbf{Q}$. The LHS expectations have captured the same structure as the calculations on the RHS, leaving the structure of the difference between the LHS and RHS calculations to be randomly distributed (not shown).

Diagnostic equation	RMSE
$E[\mathbf{d}_b^{o\eta}(\mathbf{d}_b^{o\eta})^T]$	0.0023
$E[\mathbf{d}_b^{a\eta}(\mathbf{d}_b^{o\eta})^T]$	0.0015
$E[\mathbf{d}_a^{o\eta}(\mathbf{d}_b^{o\eta})^T]$	0.0015
$E[\mathbf{d}_b^{a\eta}(\mathbf{d}_a^{o\eta})^T]$	0.0010

Table 6.1: The RMSE of the matrix elements between the left-hand side and right-hand side of each of the four weak constraint 4DVar diagnostics.

Therefore, we can state that the weak constraint 4DVar diagnostics imply the model error covariance matrix is consistent with both the background error covariance matrix and observation error covariance matrix in observation space. We show the resulting RMSE of the matrix elements between the LHS and RHS of each of the four weak constraint 4DVar diagnostics in Table 6.1. These RMSEs are of two orders of magnitude less than the specified error standard deviations in $\mathbf{B}$, $\mathbf{R}_i$ and $\mathbf{Q}_i$ and are non-zero due to sample error. Note that the larger the sample size, the smaller sample error. Here we are using a relatively small sample size of $1,000$ in comparison to the $10,000$ entries we are estimating.

We have shown how the weak constraint 4DVar diagnostics can be used to verify the consistency of an estimated model error covariance matrix $\widetilde{\mathbf{Q}}$ with both background error and observation error covariance matrices in observation space. Further work, described in Section 9.2, includes investigation into whether the right-hand side of the weak constraint 4DVar diagnostics equations (6.29), (6.33), (6.35), (6.36) can be simplified and also includes investigation into whether the weak constraint 4DVar diagnostics could be used to help refine an estimated model error covariance matrix.

6.3   Summary

In this chapter we have developed diagnostic tools that can be used as quality checks to verify the consistency of an estimated model error covariance matrix with both background error and observation error covariance matrices in observation space. We firstly derived strong constraint 4DVar diagnostic equations that account for random error present in a model. We subsequently described how strong constraint 4DVar diagnostic tools have the potential to be used to refine an estimated model error covariance matrix if it is of a certain form. Finally, we developed diagnostic equations specifically for the weak constraint formulation of 4DVar. We next investigate how the strong constraint 4DVar cost function can be amended to account for error in a model, with the aim of improving the accuracy of the model state analysis, without the need to explicitly specify the model error statistics.

Chapter 7

Improving analysis accuracy

In this chapter we consider the use of erroneous models in strong constraint 4DVar. In Section 7.1 we derive an expression for a covariance matrix which includes both observation error statistics and model error statistics and define this as the 'combined error' covariance matrix. The expression for this combined error covariance matrix includes the specification of model error covariance matrices, which are often unknown. We develop a method to estimate the combined error covariance matrix in Section 7.2, which does not require explicit specification of model error statistics. Subsequently in Section 7.3 we show when this combined error covariance matrix replaces the observation error covariance matrix in the strong constraint 4DVar cost function, a statistically better estimate of the initial state is obtained. Finally, in Sections 7.4 and 7.5 we demonstrate, with use of idealized numerical models, how use of estimated combined error covariance matrices to replace observation error covariance matrices in the strong constraint 4DVar cost function can improve the analysis accuracy.

7.1 An alternative 4DVar approach

The theory derived in this chapter is formulated with models of a linear nature. We use linear model matrices in this chapter to clearly present the methodology and to clearly understand the implications on the results. This theory will be valid for the inner loop of incremental 4DVar. Further to this, in Section 7.5 we show how methods developed in this chapter with linear model matrices can be successfully applied with erroneous models of a nonlinear nature.

The linear strong constraint 4DVar cost function (2.19) can be split into the sum of the two components $\mathcal{J}_b$ and $\mathcal{J}_{ob}$. The term $\mathcal{J}_b$ involves the comparison of the initial model state with the background and hence does not include any model evolution. Whereas the term $\mathcal{J}_{ob}$ involves the comparison of observations with the model evolved state. In Section 4.4 we discussed methods developed to account for the improper comparison between observations and the model state in $\mathcal{J}_{ob}$ due to representativity error [56] [79]. Our work aims to account for the effect model error has on the comparison between observations and the model evolved initial state in $\mathcal{J}_{ob}$ and hence account for the effect model error has on the analysis.

Let us consider the situation where the perfect model dynamics are unknown and linear erroneous model matrices $\mathbf{M}^e_{\{i-1\}\to i}$ are used to describe the model dynamics from time t_{i-1} to time t_i. The true model state is acquired as follows,

$$\mathbf{x}^t_i = \mathbf{M}^e_{\{i-1\}\to i}\mathbf{x}^t_{i-1} + \boldsymbol{\eta}_i \qquad i = 1, 2, ... \tag{7.1}$$

where the additive model error vector $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$. Note this is as previously described with equation (2.24) in Section 2.2.4, but here we consider dynamics purely of a linear nature. The linearity of the model allows us to define the operator $\hat{\mathbf{H}}^*$ as

follows,

$$\hat{\mathbf{H}}^* \in \mathbb{R}^{(N+1) \times m} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \mathbf{M}^e{}_{0 \to 1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{H}_N \mathbf{M}^e{}_{0 \to N} \end{pmatrix}.$$

When erroneous model matrices of the form $\mathbf{M}^e{}_{\{i-1\} \to i}$ are used to be evolve the model state, the operator $\hat{\mathbf{H}}$ which contains the perfect model dynamics is replaced with $\hat{\mathbf{H}}^*$ containing the erroneous model matrices in the 4DVar cost function (2.19).

Let us define $\hat{\boldsymbol{\epsilon}}_{ob} = (\boldsymbol{\epsilon}_{ob0}{}^T, \boldsymbol{\epsilon}_{ob1}{}^T, \cdots, \boldsymbol{\epsilon}_{obN}{}^T)^T$ which contains the observation error vectors $\boldsymbol{\epsilon}_{obi}$ at each of the $N+1$ observation times t_i. Let us also define the '$\mathcal{J}_{ob}$ innovation' as the difference between the vector of observations $\hat{\mathbf{y}}$ and the true initial model state $\mathbf{x}^t{}_0$ evolved with a model to the respective observation times t_i and subsequently mapped to observation space. Therefore, with perfect linear model matrices the '$\mathcal{J}_{ob}$ innovation' is $\hat{\mathbf{y}} - \hat{\mathbf{H}} \mathbf{x}^t{}_0$, whereas with erroneous linear model matrices the '$\mathcal{J}_{ob}$ innovation' is $\hat{\mathbf{y}} - \hat{\mathbf{H}}^* \mathbf{x}^t{}_0$. The error to be accounted for in the comparison of the observations with the model evolved initial state in the strong constraint 4DVar cost function (2.19) is evaluated by computing the error in the '$\mathcal{J}_{ob}$ innovation'. When a perfect model is used, the error in the '$\mathcal{J}_{ob}$ innovation' is simply the error in the observations,

$$\hat{\boldsymbol{\epsilon}}_{ob} = \hat{\mathbf{y}} - \hat{\mathbf{H}} \mathbf{x}^t{}_0, \tag{7.2}$$

where $\hat{\epsilon}_{ob} \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{R}})$. The operator $\hat{\mathbf{H}}$ takes the initial model state to the respective observation time using the perfect linear model matrices $\mathbf{M}_{\{i-1\} \to i}$ and subsequently

maps to observation space, as previously defined in Section 2.2.2.

Let us recap the objective of strong constraint 4DVar, which is to best estimate the true model state initial conditions $\mathbf{x}^t{}_0$. With perfect model dynamics, for the expectation of the minimum of the cost function (2.19) to be at the true initial state, the specification of both the background error covariance matrix $\mathbf{B}$ and observation error covariance matrix $\hat{\mathbf{R}}$ are required to accurately represent the errors in the background model state and the '$\mathcal{J}_{ob}$ innovation' respectively. With an erroneous model, the error in the '$\mathcal{J}_{ob}$ innovation' is no longer $\hat{\boldsymbol{\epsilon}}_{ob}$ (7.2) and we redefine as follows,

$$\hat{\boldsymbol{\epsilon}}^*_{ob} = \hat{\mathbf{y}} - \hat{\mathbf{H}}^* \mathbf{x}^t{}_0. \tag{7.3}$$

We next seek to obtain the statistics of this error term $\hat{\boldsymbol{\epsilon}}^*_{ob}$.

7.1.1 Combined model error and observation error statistics

Let us subtract the equation defining the observation error $\hat{\boldsymbol{\epsilon}}_{ob}$ (7.2) from the error term $\hat{\boldsymbol{\epsilon}}^*_{ob}$ (7.3) and rearrange,

$$\hat{\boldsymbol{\epsilon}}^*_{ob} \;=\; \hat{\boldsymbol{\epsilon}}_{ob} + (\hat{\mathbf{H}} - \hat{\mathbf{H}}^*)\mathbf{x}^t{}_0. \tag{7.4}$$

Simple rearrangement of equation (7.1) allows us to deduce an expression for the difference between the true state at time t_i and the true initial conditions evolved with the erroneous model matrices to time t_i ,

$$\mathbf{x}^t{}_i - \mathbf{M}^e{}_{0 \to i} \mathbf{x}^t_0 \quad = \quad \sum_{j=1}^{i} \mathbf{M}^e{}_{j \to i} \boldsymbol{\eta}_j. \tag{7.5}$$

Using equations (7.4) and (7.5) the error vector $\boldsymbol{\epsilon}^*_{obi}$ at each time t_i can now be evaluated as,

$$
\begin{aligned}
\boldsymbol{\epsilon}^*_{obi} \quad &= \quad \boldsymbol{\epsilon}_{obi} + \mathbf{H}_i (\mathbf{x}_i{}^t - \mathbf{M}^e{}_{0 \to i} \mathbf{x}^t{}_0), \\
&= \quad \boldsymbol{\epsilon}_{obi} + \mathbf{H}_i \sum_{j=1}^{i} \mathbf{M}^e{}_{j \to i} \boldsymbol{\eta}_j.
\end{aligned} \tag{7.6}
$$

Equation (7.6) explicitly states that the error in the difference between the observations and the model state initial conditions evolved with an erroneous model, is a combination of observation error and model error.

We proceed by deriving the mean of this combined error term $\boldsymbol{\epsilon}^*_{obi}$ (7.6) by taking the statistical expectation,

$$
\begin{aligned}
E[\boldsymbol{\epsilon}^*_{obi}] \quad &= \quad E[\boldsymbol{\epsilon}_{obi} + \mathbf{H}_i \sum_{j=1}^{i} \mathbf{M}^e{}_{j \to i} \boldsymbol{\eta}_j], \\
&= \quad E[\boldsymbol{\epsilon}_{obi}] + \mathbf{H}_i \sum_{j=1}^{i} \mathbf{M}^e{}_{j \to i} E[\boldsymbol{\eta}_j] = \mathbf{0},
\end{aligned} \tag{7.7}
$$

using the assumptions that both the observation errors and model errors are distributed with a zero mean. Next we evaluate the covariance of the combined error terms at observation times t_i and t_k respectively by taking the following statistical expectation,

$$
\begin{aligned}
\mathbf{R}^*{}_{(i,k)} &= E[\boldsymbol{\epsilon}^*_{obi}(\boldsymbol{\epsilon}^*_{obk})^T], \\[2mm]
&= E[(\boldsymbol{\epsilon_{obi}} + \mathbf{H}_i \sum_{j=1}^{i} \mathbf{M}^e{}_{j\to i}\boldsymbol{\eta}_j)(\boldsymbol{\epsilon_{obk}} + \mathbf{H}_k \sum_{j=1}^{k} \mathbf{M}^e{}_{j\to k}\boldsymbol{\eta}_j)^T], \\[2mm]
&= E[\boldsymbol{\epsilon_{obi}}(\boldsymbol{\epsilon_{obk}})^T] + \mathbf{H}_i E[\sum_{j=1}^{i} \mathbf{M}^e{}_{j\to i}\boldsymbol{\eta}_j (\sum_{j=1}^{k} \mathbf{M}^e{}_{j\to k}\boldsymbol{\eta}_j)^T]\mathbf{H}_k{}^T, \qquad (7.8)
\end{aligned}
$$

using equation (7.6) and assuming that the model errors are of a random nature and are uncorrelated to the errors in the observations. Evaluating the expectations in equation (7.8) leads to,

$$
\mathbf{R}^*{}_{(i,k)} =
\begin{cases}
\mathbf{R}_0 & \text{for i=k=0,} \\[4mm]
\mathbf{R}_i + \mathbf{H}_i \left[\displaystyle\sum_{j=1}^{min(i,k)} \mathbf{M}^e{}_{j\to i}\mathbf{Q}_j\mathbf{M}^e{}_{j\to k}{}^T \right] \mathbf{H}_k{}^T & \text{for i=k} \neq 0, \\[4mm]
\mathbf{H}_i \left[\displaystyle\sum_{j=1}^{min(i,k)} \mathbf{M}^e{}_{j\to i}\mathbf{Q}_j\mathbf{M}^e{}_{j\to k}{}^T \right] \mathbf{H}_k{}^T & \text{otherwise,}
\end{cases}
\qquad (7.9)
$$

where the model error vectors $\boldsymbol{\eta}_i$ are assumed to be of a random nature. We note that no model evolution is required in the comparison of observations and the model state at time t_0, hence the observation error covariance matrix $\mathbf{R}_0$ fully describes the error statistics in this comparison at time t_0. We now investigate the composition of the combined error covariance terms (7.9) and discuss the structure of the full combined error covariance matrix.

Let $\mathbf{Q}^*{}_{(i,k)}$ be the terms in $\mathbf{R}^*{}_{(i,k)}$ (7.9) that are present due to the errors in the model,

$$\mathbf{Q}^*{}_{(i,k)} = \mathbf{H}_i \left[\sum_{j=1}^{min(i,k)} \mathbf{M}^e{}_{j \to i} \mathbf{Q}_j \mathbf{M}^e{}_{j \to k}{}^T \right] \mathbf{H}_k{}^T. \tag{7.10}$$

We can then present the full combined model error and observation error covariance matrix $\hat{\mathbf{R}}^*$, containing the covariance sub matrices $\mathbf{R}^*{}_{(i,k)}$ (7.9) as follows,

$$\hat{\mathbf{R}}^* = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{R}_1 + \mathbf{Q}^*{}_{(1,1)} & \mathbf{Q}^*{}_{(1,2)} & \cdots & \mathbf{Q}^*{}_{(1,N)} \\ \vdots & \mathbf{Q}^*{}_{(2,1)} & \mathbf{R}_2 + \mathbf{Q}^*{}_{(2,2)} & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & \vdots \\ 0 & \mathbf{Q}^*{}_{(N,1)} & \cdots & \cdots & \mathbf{R}_N + \mathbf{Q}^*{}_{(N,N)} \end{pmatrix}. \tag{7.11}$$

To recap, when a perfect model is used in the strong constraint 4DVar cost function (2.19), the error to be accounted for in the '$\mathcal{J}_{ob}$ innovation' is $\hat{\epsilon}_{ob} \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{R}})$. Whereas when an erroneous model is used in the strong constraint 4DVar cost function (2.19), the error to be accounted for in the '$\mathcal{J}_{ob}$ innovation' is $\hat{\epsilon}^*{}_{ob} \sim \mathcal{N}(\mathbf{0}, \hat{\mathbf{R}}^*)$, with $\hat{\mathbf{R}}^*$ given by (7.11).

The effect the model error has on the covariance matrix for the error in the '$\mathcal{J}_{ob}$ innovation' is:

1. An additional term $\mathbf{Q}^*{}_{(i,i)}$, as described by equation (7.10) with $i = k$, on the block diagonal sub matrices $\mathbf{R}^*{}_{(i,i)}$. From (7.10) we see that this is an accumulation of the model error over the assimilation time window.

2. The formation of off diagonal block covariance sub matrices described by $\mathbf{Q}^*{}_{(i,k)}$, as described by equation (7.10) with $i \neq k$. This is the presence of time correlations caused by the error in the model.

We now explain the presence of the additional term $\mathbf{Q}^*_{(i,i)}$ on the block diagonal sub matrices $\mathbf{R}^*_{(i,i)}$ described by point 1. When an erroneous model is used within strong constraint 4DVar, observations $\mathbf{y}_i$ are compared to the model state $\mathbf{M}^e_{i-1 \to i} \mathbf{M}^e_{i-2 \to i-1} \ldots \mathbf{M}^e_{1 \to 2} \mathbf{M}^e_{0 \to 1} \mathbf{x}_0$ (mapped to observation space). With each further time-step in the assimilation window, a further evolution of the model state initial conditions is required, which requires multiplication with yet another erroneous model matrix. Each erroneous model matrix has a corresponding vector of model error and a model error covariance matrix. The uncertainty at each model time-step is accumulated. This uncertainty in the comparison of the observations and model state (in observation space) is represented by $\mathbf{Q}^*_{(i,i)}$.

We now detail the presence of off diagonal block covariance sub matrices $\mathbf{Q}^*_{(i,k)}$ (7.10), where $i \neq k$, as stated in point 2. When observations $\mathbf{y}_i$ and $\mathbf{y}_k$ are available at distinct times t_i and t_k in an assimilation window, these are compared to the initial model state evolved to times t_i and t_k respectively. These comparisons at the distinct times t_i and t_k, have correlations due to the errors in the model. Although the model error vectors $\boldsymbol{\eta}_i$ as defined in (7.1) are independent at each time-step, they affect the evolution of the model state at the subsequent times. For example, the values of $\mathbf{x}_2$ and $\mathbf{x}_3$ have both been affected by the error in the model $\boldsymbol{\eta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_1)$ and $\boldsymbol{\eta}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_2)$ at times t_1 and t_2. The summation in $\mathbf{Q}^*_{(i,k)}$ is taken up to the minimum of the two times t_i and t_k. The reason for this is that the values of model states at times t_i and t_k are both dependent on the same model error vectors up to this time $(\min(t_i, t_k))$.

We propose that the use of $\hat{\mathbf{R}}^*$ (7.11), as opposed to $\hat{\mathbf{R}}$, in the cost function (2.19), will statistically improve the analysis accuracy when random error is present in the model. We next show how use of this derived combined error covariance matrix $\hat{\mathbf{R}}^*$ (7.11) in the strong constraint 4DVar cost function (2.19) allows the strong constraint 4DVar diagnostics (2.34)-(2.37) to hold, even in the presence of model error.

7.1.2 Consistency with diagnostics

We propose the following theorem.

Theorem 7.1.1. *When an erroneous linear model matrix $\mathbf{M}^e$, with corresponding additive model error vector $\boldsymbol{\eta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, is used within the strong constraint 4DVar cost function with observations $\mathbf{y}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ present at time t_1, the use of $\mathbf{R}^* = \mathbf{R} + \mathbf{HQH}^T$ to replace $\mathbf{R}$, where $\mathbf{H}$ is a linear observation operator, ensures that the strong constraint 4DVar diagnostics (2.34)-(2.37) are upheld.*

Proof. We consider a background model state $\mathbf{x}^b$ and corresponding background error covariance matrix $\mathbf{B}$ of the form (2.5). We assume the background errors, observation errors and model errors are all uncorrelated. This situation is as described in Section 2.3.1, however with an erroneous model matrix $\mathbf{M}^e$ replacing the perfect model matrix $\mathbf{M}$. Using the formulae we derived earlier in this chapter (7.11) we define the combined model error and observation error covariance matrix $\mathbf{R}^* = \mathbf{R} + \mathbf{HQH}^T$. We replace $\mathbf{R}$ in the strong constraint 4DVar cost function (2.19) with $\mathbf{R}^*$ and minimise with respect to the initial state to obtain the following analysis,

$$\mathbf{x}^a_0{}^* = \mathbf{x}^b + \mathbf{K}^* \mathbf{d}^o{}_b{}^*, \tag{7.12}$$

where the gain matrix $\mathbf{K}^* = \mathbf{BM}^{eT}\mathbf{H}^T(\mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T + \mathbf{R}^*)^{-1}$ and the innovation vector $\mathbf{d}^o{}_b{}^* = \mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^b$. Let us define the following differences in observation space,

$$\mathbf{d}_b^{a*} = \mathbf{HM}^e\mathbf{x}^a{}_0{}^* - \mathbf{HM}^e\mathbf{x}^b,$$

$$\mathbf{d}_a^{o*} = \mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^a{}_0{}^*. \tag{7.13}$$

The first of these compares the analysis with the background, both of which are evolved to time t_1 with the erroneous model and subsequently mapped to observation space. The second of these compares the vector of observations at time t_1 with the analysis evolved to time t_1 with the erroneous model and subsequently mapped to observation space.

We now derive the first of the four diagnostics by taking the statistical expectation of the product of innovation vectors,

$$
\begin{aligned}
E[\mathbf{d}^o{}_b{}^*(\mathbf{d}^o{}_b{}^*)^T] &= E[(\mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^b)(\mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^b)^T], \\
&= E[(\mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^t{}_0 + \mathbf{HM}^e\mathbf{x}^t{}_0 - \mathbf{HM}^e\mathbf{x}^b) \\
&\quad (\mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^t{}_0 + \mathbf{HM}^e\mathbf{x}^t{}_0 - \mathbf{HM}^e\mathbf{x}^b)^T], \\
&= E[(\boldsymbol{\epsilon}^*_{ob1} - \mathbf{HM}^e\epsilon_b)(\boldsymbol{\epsilon}^*_{ob1} - \mathbf{HM}^e\epsilon_b)^T], \\
&= \mathbf{R}^* + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T, \tag{7.14}
\end{aligned}
$$

using the definitions of the background error (2.5) and combined model error and observation error (7.3). This result (7.14) can also be obtained by substituting $\mathbf{R} + \mathbf{HQH}^T = \mathbf{R}^*$ into the diagnostic equation (6.8), as this diagnostic equation does not involve use of the analysis. Next the second of the four diagnostics is derived by taking the statistical expectation of the product of the differences $\mathbf{d}_b^{a*}$ and $\mathbf{d}^o{}_b{}^*$,

$$
\begin{aligned}
E[\mathbf{d}_b^{a*}(\mathbf{d}^o{}_b{}^*)^T] &= E[(\mathbf{HM}^e\mathbf{x}^a{}_0{}^* - \mathbf{HM}^e\mathbf{x}^b)(\mathbf{d}^o{}_b{}^*)^T], \\
&= E[(\mathbf{HM}^e(\mathbf{x}^b + \mathbf{K}^*\mathbf{d}^o{}_b{}^*) - \mathbf{HM}^e\mathbf{x}^b)(\mathbf{d}^o{}_b{}^*)^T], \\
&= \mathbf{HM}^e\mathbf{K}^* E[(\mathbf{d}^o{}_b{}^*)(\mathbf{d}^o{}_b{}^*)^T], \\
&= \mathbf{HM}^e\mathbf{K}^*(\mathbf{R}^* + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T), \\
&= \mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T(\mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T + \mathbf{R}^*)^{-1} \\
&\quad (\mathbf{R}^* + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T), \\
&= \mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T, \quad\quad\quad\quad\quad\quad\quad (7.15)
\end{aligned}
$$

using the analysis (7.12) and the previous diagnostic result (7.14). The third of the four diagnostics is derived by taking the statistical expectation of the product of the differences $\mathbf{d}^o{}_a{}^*$ and $\mathbf{d}^o{}_b{}^*$,

$$
\begin{aligned}
E[\mathbf{d}^o{}_a{}^*(\mathbf{d}^o{}_b{}^*)^T] &= E[(\mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^a{}_0)(\mathbf{d}^o{}_b{}^*)^T], \\
&= E[(\mathbf{y}_1 - \mathbf{HM}^e(\mathbf{x}^b + \mathbf{K}^*\mathbf{d}^o{}_b{}^*))(\mathbf{d}^o{}_b{}^*)^T], \\
&= E[(\mathbf{y}_1 - \mathbf{HM}^e\mathbf{x}^t{}_0 + \mathbf{HM}^e\mathbf{x}^t{}_0 - \mathbf{HM}^e\mathbf{x}^b - \mathbf{HM}^e\mathbf{K}^*\mathbf{d}^o{}_b{}^*)(\mathbf{d}^o{}_b{}^*)^T], \\
&= E[(\boldsymbol{\epsilon}^*_{ob1} - \mathbf{HM}^e\epsilon_b - \mathbf{HM}^e\mathbf{K}^*\mathbf{d}^o{}_b{}^*)(\mathbf{d}^o{}_b{}^*)^T], \\
&= E[\boldsymbol{\epsilon}^*_{ob1}(\boldsymbol{\epsilon}^*_{ob1} - \mathbf{HM}^e\epsilon_b)^T - \mathbf{HM}^e\epsilon_b(\boldsymbol{\epsilon}^*_{ob1} - \mathbf{HM}^e\epsilon_b)^T \\
&\quad - \mathbf{HM}^e\mathbf{K}^*\mathbf{d}^o{}_b{}^*(\mathbf{d}^o{}_b{}^*)^T], \\
&= \mathbf{R}^* + \mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T \\
&\quad - \mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T(\mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T + \mathbf{R}^*)^{-1} \\
&\quad (\mathbf{R}^* + \mathbf{HM}^e\mathbf{BM}^{eT}\mathbf{H}^T), \\
&= \mathbf{R}^*, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (7.16)
\end{aligned}
$$

using the analysis (7.12) and the diagnostic result (7.14). Lastly, the fourth diagnostic is derived by taking the statistical expectations of the product of the differences $\mathbf{d}^a{}_b{}^*$ and $\mathbf{d}^o{}_a{}^*$,

$$
\begin{aligned}
E[\mathbf{d}^a{}_b{}^*(\mathbf{d}^o{}_a{}^*)^T] &= E[(\mathbf{HM}^e\mathbf{x}^a{}_0{}^* - \mathbf{HM}^e\mathbf{x}^b)(\mathbf{d}^o{}_a{}^*)^T], \\
&= E[\mathbf{HM}^e\mathbf{K}^*\mathbf{d}^o{}_b{}^*(\mathbf{d}^o{}_a{}^*)^T], \\
&= \mathbf{HM}^e\mathbf{K}^* E[\mathbf{d}^o{}_b{}^*(\mathbf{d}^o{}_a{}^*)^T], \\
&= \mathbf{HM}^e\mathbf{K}^* \left(E[\mathbf{d}^o{}_a{}^*(\mathbf{d}^o{}_b{}^*)^T] \right)^T, \\
&= \mathbf{HM}^e\mathbf{K}^*(\mathbf{R}^*)^T, \\
&= \mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T(\mathbf{HM}^e\mathbf{B}(\mathbf{M}^e)^T\mathbf{H}^T + \mathbf{R}^*)^{-1}\mathbf{R}^*, \quad (7.17)
\end{aligned}
$$

using workings from the derivation of (7.15) and the diagnostic result (7.16). Note that the three diagnostic equations we just derived (7.15)-(7.17) differ from the corresponding diagnostic equations (6.10), (6.12) and (6.13) derived in Chapter 6 that account for the presence of random model error. This is because the diagnostic equations (6.10), (6.12) and (6.13) use the analysis produced from the unamended 4DVar cost function, whereas (7.15)-(7.17) use the analysis produced with $\mathbf{R}^*$ in the 4DVar cost function.

We have now proved that, when error is present in the model, use of $\mathbf{R}^* = \mathbf{R} + \mathbf{HQH}^T$, to replace $\mathbf{R}$ in the strong constraint 4DVar cost function, ensures all four of the strong constraint 4DVar diagnostics (2.34)-(2.37) are upheld. $\qquad \square$

For use of $\hat{\mathbf{R}}^*$ (7.11) in the strong constraint 4DVar cost function, specification of model error covariance matrices $\mathbf{Q}_i$ are required, at each time t_i in the assimilation window up to the last observation time. We described the difficulties that operational NWP centres have in specifying model error covariance matrices $\mathbf{Q}_i$ for atmosphere

139

and ocean models in Section 4.1. Therefore, we now aim to derive a method of evaluating the entries in the combined error covariance matrix $\hat{\mathbf{R}}^*$ without having to explicitly specify the model error covariance matrices $\mathbf{Q}_i$ themselves.

7.2 Estimation of the combined model error and observation error covariance matrix

In this section we develop a method to estimate the combined model error and observation error matrix $\hat{\mathbf{R}}^*$. This method does not require explicit specification of model error covariance matrices.

7.2.1 Diagnostic tools for estimation of the combined error covariance matrix

Let us derive the first of the four diagnostics (7.14) for the case where observations are available at multiple times through an assimilation window. This derivation will use erroneous model matrices of the form $\mathbf{M}^e_{\{i-1\}\to i}$, as described in (7.1), in the strong constraint 4DVar cost function (2.19).

The definitions of observation error (2.4) and background error (2.5) along with a simple rearrangement of equation (7.1) enables the innovation involving the erroneous model, at each time t_i, to be defined as follows,

$$(\mathbf{d}^{o*}_b)_i = \mathbf{y}_i - \mathbf{H}_i\mathbf{M}^e_{0\to i}\mathbf{x}^b = \boldsymbol{\epsilon}_{obi} - \mathbf{H}_i\mathbf{M}^e_{0\to i}\boldsymbol{\epsilon}_b + \mathbf{H}_i\sum_{j=1}^{i}\mathbf{M}^e_{j\to i}\boldsymbol{\eta}_j. \qquad (7.18)$$

We calculate the expectation of the following innovation product,

$$
\begin{aligned}
E[(\mathbf{d}_b^{o*})_i(\mathbf{d}_b^{o*})_k{}^T] \;\; = \;\; & < (\mathbf{y}_i - \mathbf{H}_i\mathbf{M}^e{}_{0\to i}\mathbf{x}^b)(\mathbf{y}_k - \mathbf{H}_k\mathbf{M}^e{}_{0\to k}\mathbf{x}^b)^T >, \\
= \;\; & < (\epsilon_{obi} - \mathbf{H}_i\mathbf{M}^e{}_{0\to i}\epsilon_b + \mathbf{H}_i\sum_{j=1}^{i}\mathbf{M}^e{}_{j\to i}\boldsymbol{\eta}_j) \\
& (\epsilon_{obk} - \mathbf{H}_k\mathbf{M}^e{}_{0\to k}\epsilon_b + \mathbf{H}_k\sum_{j=1}^{k}\mathbf{M}^e{}_{j\to k}\boldsymbol{\eta}_j)^T >,
\end{aligned}
$$
$$(7.19)$$

using (7.18). The definition of $\mathbf{R}^*{}_{(i,k)}$ as defined in (7.8) allows us to deduce,

$$
E[(\mathbf{d}_b^{o*})_i(\mathbf{d}_b^{o*})_k{}^T] = \mathbf{R}^*{}_{(i,k)} + \mathbf{H}_i\mathbf{M}^e{}_{0\to i}\mathbf{B}\mathbf{M}^e{}_{0\to k}{}^T\mathbf{H}_k{}^T. \tag{7.20}
$$

Assembling the entries at all observation times into the vector $\hat{\mathbf{d}}_b^{o*}$ we have,

$$
E[\hat{\mathbf{d}}_b^{o*}(\hat{\mathbf{d}}_b^{o*})^T] \;\; = \;\; \hat{\mathbf{R}}^* + \hat{\mathbf{H}}^*\mathbf{B}\hat{\mathbf{H}}^*{}^T, \tag{7.21}
$$

where $\hat{\mathbf{R}}^*$ is defined by equation (7.11). With simple rearrangement of equation (7.21), we derive an expression for the estimation of the combined error matrix,

$$
\widetilde{\mathbf{R}}^* \;\; = \;\; E[\hat{\mathbf{d}}_b^{o*}(\hat{\mathbf{d}}_b^{o*})^T] - \hat{\mathbf{H}}^*\mathbf{B}\hat{\mathbf{H}}^*{}^T. \tag{7.22}
$$

Let us consider operational NWP centres, modelling the conditions of the Earth's atmosphere and ocean, where there is only one true state at each point in time. In order

to use equation (7.22) to estimate the combined statistics for an operational NWP system, a sample of background model states, model error vectors and observations are required. In Section 6.1.4 we outlined potential methods for obtaining samples of innovations operationally and stated that investigating the best method to do so is an area of suggested further work, as discussed in Section 9.2 of this thesis. In Section 3.2 we described how the randomization technique can be used to estimate the matrix $\mathbf{HMBM}^T\mathbf{H}^T$. This method allows the diagonal elements of $\mathbf{HMBM}^T\mathbf{H}^T$ to be evaluated for operational use [2]. This same methodology can be extended and applied to evaluate the diagonal elements of $\hat{\mathbf{H}}^*\mathbf{B}\hat{\mathbf{H}}^{*T}$ required in (7.22). Therefore use of equation (7.22) is potentially operationally feasible in areas of high observational frequency, to estimate diagonal elements of the combined error matrix. Next we demonstrate the use of equation (7.22) to estimate combined error statistics with use of an erroneous numerical model.

7.2.2 Application of developed method to estimate combined error statistics

We take our erroneous linear model to be the time-stepping solution of the linear advection equation over the spatial domain $x \in [0, 10)$, as described in Section 5.1.1, with the spatial step $\Delta x = 0.1$ and time-step $\Delta t = 0.1$. The true model state at each time t_i, differs from the erroneous model state by random error $\boldsymbol{\eta}_i$, as defined by equation (7.1). We define the structure of the model error covariance matrix $\mathbf{Q}_i$ to be diagonal with variance $\sigma_q{}^2 = 0.01$ at each model time-step. We have a background model state $\mathbf{x}^b$ and corresponding background error covariance matrix $\mathbf{B}$. We use the SOAR function, as described in Section 5.3.1 with correlation length scale $\mathcal{L} = 0.4$, to assign the correlations in the background error covariance matrix $\mathbf{B}$ and set the corresponding background error scalar variance $\sigma_b{}^2 = 0.04$. We assume we have

direct observations $\mathbf{y}_i$ of all spatial points every two time-steps over an assimilation window length of eight time-steps, with linear observation operator $\mathbf{H} = \mathbf{I}$. The observation error covariance matrices are specified to be $\mathbf{R}_i = \sigma_{ob}{}^2\mathbf{I}$ (for $i = 2, 4, 6, 8$) with variance $\sigma_{ob}{}^2 = 0.04$.

We calculate and compare the combined model error and observation error statistics at each observation time, firstly explicitly with equation (7.11) and secondly with the sample approximation (7.22). Calculations of the block diagonal combined error covariance matrices $\mathbf{R}^*_{(i,i)} \in \mathbb{R}^{100 \times 100}$ are shown in Figure 7.1 (left), at observation times t_i $(i = 2, 4, 6, 8)$. These matrices maintain a diagonal structure at all observation times. At each subsequent observation time the variances in the combined error matrix increase, in this case $\sigma_{ob}{}^{*2} = 0.06$ at t_2, $\sigma_{ob}{}^{*2} = 0.08$ at t_4, $\sigma_{ob}{}^{*2} = 0.10$ at t_6 and $\sigma_{ob}{}^{*2} = 0.12$ at t_8. This represents the increase in uncertainty of the model trajectory throughout the assimilation window, due to the errors present in the model.

We next estimate the same block diagonal combined error matrices $\widetilde{\mathbf{R}}^*_{(i,i)}$ $(i = 2, 4, 6, 8)$ with equation (7.22). We therefore need to set up a numerical environment where the required sample innovation data is available. We take the true initial conditions of the model state to be the exponential function f centred around $x_0 = 5$ in the spatial domain, as defined in equation (5.2). We use equation (7.1) to produce the true model trajectories, where the model operator $\mathbf{M}^e_{\{i-1\}\to i}$ is taken to be the linear advection equation. We repeat the following steps $5,000$ times to produce a sample size of $5,000$ background values, model error vectors and observations:

1. Produce a background vector $\mathbf{x}^b$ by adding noise, consistent with the statistics prescribed in $\mathbf{B}$, to the true initial state $\mathbf{x}^t{}_0$.

2. Produce a random model error vector $\boldsymbol{\eta}_i$ for each time $t_1, ..., t_8$ using the statistics specified in the model error covariance matrices $\mathbf{Q}_i$.

3. Produce a vector of observations $\mathbf{y}_i$ at each observation time t_i $(i=2, 4, 6, 8)$ by

evaluating the true model state vector $\mathbf{x}^t{}_i$ with equation (7.1) and adding noise in proportion to the error statistics specified in $\mathbf{R}_i$.

This process gives us a sample size of 5000 innovation vectors $(\mathbf{d}_b^{o*})_i$ for the estimation of $\widetilde{\mathbf{R}}^*_{(i,i)} \in \mathbb{R}^{100 \times 100}$ at each observation time t_i $(i = 2, 4, 6, 8)$ in equation (7.22). The estimation is found to be very successful in representing the structure and values of the combined error covariance matrix. For example, the difference between the central row (51st) of $\mathbf{R}^*_{(i,i)}$ and $\widetilde{\mathbf{R}}^*_{(i,i)}$ for each time t_i $(i = 2, 4, 6, 8)$ can clearly be seen in the RHS of Figure 7.1. The overall RMSEs (Root Mean Square Error) between $\mathbf{R}^*_{(i,i)}$ and $\widetilde{\mathbf{R}}^*_{(i,i)}$ for each time t_i $(i = 2, 4, 6, 8)$ are shown in Table 7.1. These RMSEs are of two orders of magnitude less than the specified error standard deviations in $\mathbf{B}$, $\mathbf{R}_i$ and $\mathbf{Q}_i$ and are only non-zero due to sample error. Note that the larger the sample size, the smaller the sample error. Here we are using a relatively small sample size of $5,000$ in comparison to the $10,000$ entries we are estimating.
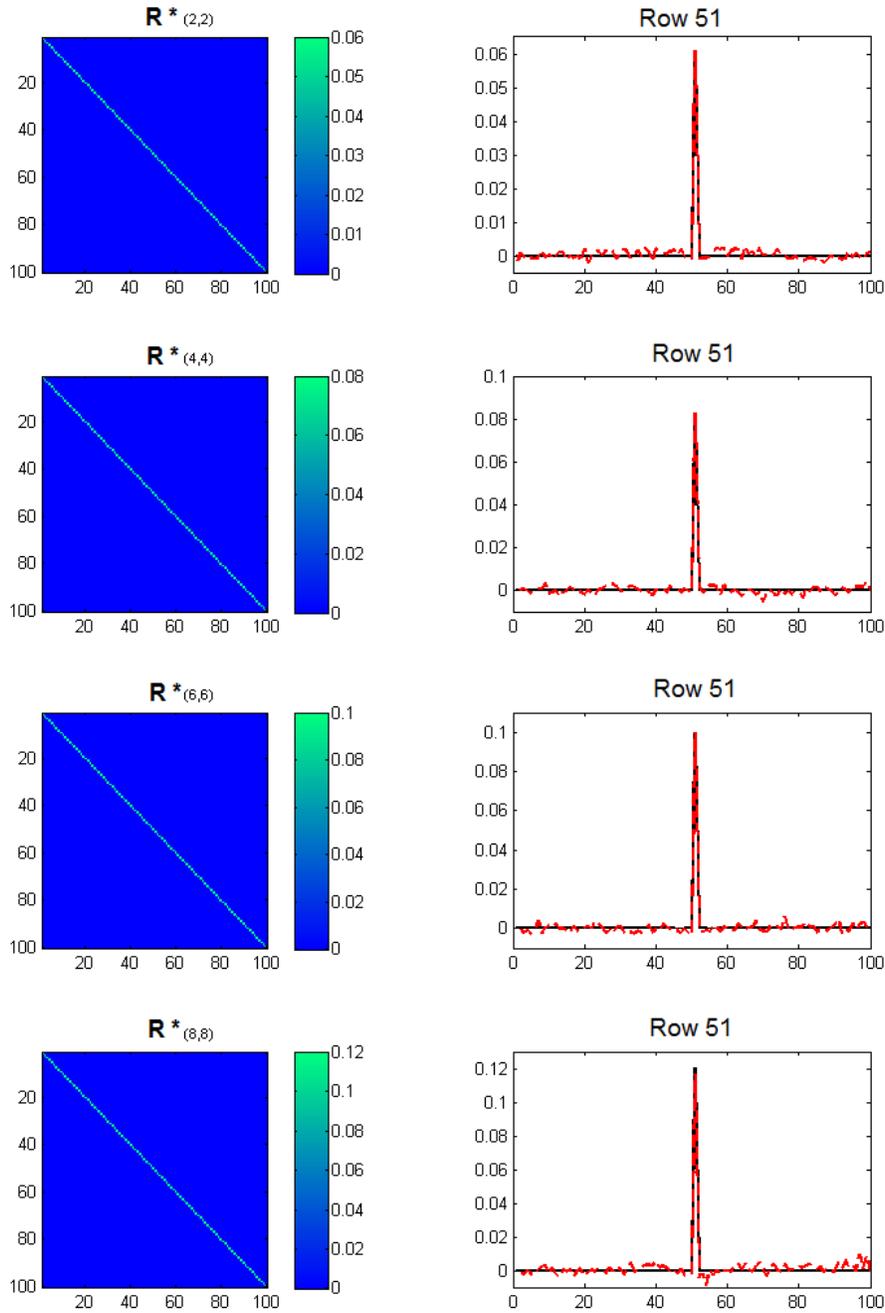
Figure 7.1: Combined model error and observation error covariance matrix $\mathbf{R}^*_{(i,i)} \in \mathbb{R}^{100 \times 100}$ at each time t_i (i=2, 4, 6, 8) calculated explicitly using equation (7.11) (left). The central row (51st) of $\mathbf{R}^*_{(i,i)}$ for each time t_i (i=2, 4, 6, 8) calculated explicitly using equation (7.11) is shown with a black line (right). The central row (51st) of $\widetilde{\mathbf{R}}^*_{(i,i)}$ for each time t_i ($i = 2, 4, 6, 8$) estimated from the sample (7.22) is shown with a red dotted (right).

Time	RMSE
t_2	0.0014
t_4	0.0017
t_6	0.0020
t_8	0.0023

Table 7.1: The RMSE of the matrix elements between $\mathbf{R}^*_{(i,i)}$ and $\widetilde{\mathbf{R}}^*_{(i,i)}$ for each time t_i $(i = 2, 4, 6, 8)$.

We now summarise this section. We have developed a method to estimate $\hat{\mathbf{R}}^*$ (7.11) using diagnostic tools, which we denote $\widetilde{\mathbf{R}}^*$ (7.22). In our numerical experiments we have used a 'toy' model with corresponding 'toy' observational data and model error vectors produced using the statistics stated in $\mathbf{R}_i$ and $\mathbf{Q}_i$ respectively. Once we have produced the sample of innovations required for the calculation of $\widetilde{\mathbf{R}}^*$ (7.22), the matrices $\mathbf{R}_i$ and $\mathbf{Q}_i$ were no longer used. We should mention that for operational use of this method, this estimation assumes accurate specification of the background error covariance matrix $\mathbf{B}$. However, neither the observation error covariance matrices $\mathbf{R}_i$ or model error covariance matrices $\mathbf{Q}_i$ are required in this estimation. Instead a sample of innovation vectors is required (refer back to Section 6.1.4 where potential methods to obtain this data operationally is discussed). We predict that use of combined model error and observation error statistics, in the place of the observation error covariance matrix in the strong constraint 4DVar cost function, will produce a more statistically accurate analysis. The next section will investigate this hypothesis.

7.3 Improvements to analysis accuracy

In this section we show the improvements to analysis accuracy that can be obtained through use of the combined error statistics.

7.3.1 Analysis error covariance matrix

The analysis when the erroneous model $\mathbf{x}_i = \mathbf{M}^e{}_{\{i-1\}\to i}\mathbf{x}_{i-1}$ is used within the strong constraint 4DVar cost function (2.19) and the error in the model is not accounted for is as follows,

$$\mathbf{x}^a{}_0 \;=\; \mathbf{x}^b + \hat{\mathbf{K}}\hat{\mathbf{d}}^o_b, \tag{7.23}$$

with the gain matrix $\hat{\mathbf{K}} = \mathbf{B}\hat{\mathbf{H}}^{*T}(\hat{\mathbf{H}}^*\mathbf{B}\hat{\mathbf{H}}^{*T} + \hat{\mathbf{R}})^{-1}$ and the innovation vector $\hat{\mathbf{d}}^o_b = \hat{\mathbf{y}} - \hat{\mathbf{H}}^*\mathbf{x}^b$. The corresponding error in the analysis is evaluated,

$$\boldsymbol{\epsilon}_a \;=\; \mathbf{x}^a{}_0 - \mathbf{x}^t{}_0 = \mathbf{x}^b + \hat{\mathbf{K}}\hat{\mathbf{d}}^o_b - \mathbf{x}^t{}_0 = \boldsymbol{\epsilon}_b + \hat{\mathbf{K}}\hat{\mathbf{d}}^o_b. \tag{7.24}$$

We take the statistical expectations of $\boldsymbol{\epsilon}_a$ to calculate the mean of the error in the analysis,

$$< \boldsymbol{\epsilon}_a > \;=\; < \boldsymbol{\epsilon}_b + \hat{\mathbf{K}}\hat{\mathbf{d}}^o_b > = < \boldsymbol{\epsilon}_b > + \hat{\mathbf{K}} < \hat{\mathbf{d}}^o_b > = \mathbf{0}, \tag{7.25}$$

using the assumptions that the background error, observation errors and model errors all have a zero mean. We next evaluate the analysis error covariance matrix,

$$
\begin{aligned}
\mathbf{A} \; &= \; <\boldsymbol{\epsilon}_a(\boldsymbol{\epsilon}_a)^T>, \\
&= \; <(\boldsymbol{\epsilon}_b + \hat{\mathbf{K}}\hat{\mathbf{d}}_b^o)(\boldsymbol{\epsilon}_b + \hat{\mathbf{K}}\hat{\mathbf{d}}_b^o)^T>, \\
&= \; \mathbf{B} + \mathbb{E}[\boldsymbol{\epsilon}_b(\hat{\mathbf{d}}_b^o)^T]\hat{\mathbf{K}}^T + \hat{\mathbf{K}}\mathbb{E}[\hat{\mathbf{d}}_b^o\boldsymbol{\epsilon_b}^T] + \hat{\mathbf{K}}\mathbb{E}[\hat{\mathbf{d}}_b^o(\hat{\mathbf{d}}_b^o)^T]\hat{\mathbf{K}}^T, \\
&= \; \mathbf{B} - \mathbf{B}\hat{\mathbf{H}}^{*^T}\hat{\mathbf{K}}^T - \hat{\mathbf{K}}\hat{\mathbf{H}}^{*}\mathbf{B} + \hat{\mathbf{K}}(\hat{\mathbf{R}}^{*} + \hat{\mathbf{H}}^{*}\mathbf{B}\hat{\mathbf{H}}^{*^T})\hat{\mathbf{K}}^T, \\
&= \; \mathbf{B} - \mathbf{B}\hat{\mathbf{H}}^{*^T}(\hat{\mathbf{R}} + \hat{\mathbf{H}}^{*}\mathbf{B}\hat{\mathbf{H}}^{*^T})^{-1}\hat{\mathbf{H}}^{*}\mathbf{B} \\
&+ \; \mathbf{B}\hat{\mathbf{H}}^{*^T}(\hat{\mathbf{R}} + \hat{\mathbf{H}}^{*}\mathbf{B}\hat{\mathbf{H}}^{*^T})^{-1}\hat{\mathbf{Q}}^{*}(\hat{\mathbf{R}} + \hat{\mathbf{H}}^{*}\mathbf{B}\hat{\mathbf{H}}^{*^T})^{-1}\hat{\mathbf{H}}^{*}\mathbf{B} \\
&= \; \mathbf{B} - \hat{\mathbf{K}}\hat{\mathbf{H}}^{*}\mathbf{B} + \hat{\mathbf{K}}\hat{\mathbf{Q}}^{*}\hat{\mathbf{K}}^T, \\
&= \; (\mathbf{I} - \hat{\mathbf{K}}\hat{\mathbf{H}}^{*})\mathbf{B} + \hat{\mathbf{K}}\hat{\mathbf{Q}}^{*}\hat{\mathbf{K}}^T, \quad\quad\quad\quad\quad\quad\quad (7.26)
\end{aligned}
$$

where $\hat{\mathbf{Q}}^{*} = \hat{\mathbf{R}}^{*} - \hat{\mathbf{R}}$, using the earlier result (7.21) and the assumption that the background errors, observation errors and model errors are uncorrelated.

The analysis when the erroneous model $\mathbf{x}_i = \mathbf{M}^e{}_{\{i-1\}\to i}\mathbf{x}_{i-1}$ is used within the strong constraint 4DVar cost function (2.19) and the error in the model is accounted for, by using $\hat{\mathbf{R}}^{*}$ as opposed to $\hat{\mathbf{R}}$ in the strong constraint 4DVar cost function, is as follows,

$$
\mathbf{x}^a{}_0{}^{*} \; = \; \mathbf{x}^b + \hat{\mathbf{K}}^{*}\hat{\mathbf{d}}_b^{o*}, \quad\quad\quad\quad\quad\quad\quad (7.27)
$$

with the gain matrix $\hat{\mathbf{K}}^{*} = \mathbf{B}\hat{\mathbf{H}}^{*^T}(\hat{\mathbf{H}}^{*}\mathbf{B}\hat{\mathbf{H}}^{*^T} + \hat{\mathbf{R}}^{*})^{-1}$ and the innovation vector $\hat{\mathbf{d}}_b^{o*} = \hat{\mathbf{y}} - \hat{\mathbf{H}}^{*}\mathbf{x}^b$. The corresponding error in the analysis is evaluated,

$$
\boldsymbol{\epsilon}_a{}^{*} \; = \; \mathbf{x}^a{}_0{}^{*} - \mathbf{x}^t{}_0 = \mathbf{x}^b + \hat{\mathbf{K}}^{*}\hat{\mathbf{d}}_b^{o*} - \mathbf{x}^t{}_0 = \boldsymbol{\epsilon}_b + \hat{\mathbf{K}}^{*}\hat{\mathbf{d}}_b^{o*}. \quad\quad\quad (7.28)
$$

We take the statistical expectation of $\boldsymbol{\epsilon}_a{}^*$ to calculate the mean of the error in the analysis,

$$< \boldsymbol{\epsilon}_a{}^* > \;\; = \;\; < \boldsymbol{\epsilon}_b + \hat{\mathbf{K}}^* \hat{\mathbf{d}}_b^{o*} > = < \boldsymbol{\epsilon}_b > + \hat{\mathbf{K}}^* < \hat{\mathbf{d}}_b^{o*} > = \mathbf{0}, \qquad (7.29)$$

using the assumption that the background errors, observation errors and model errors all have a zero mean. We next evaluate the analysis error covariance matrix,

$$
\begin{aligned}
\mathbf{A}^* \;\; &= \;\; < \boldsymbol{\epsilon}_a{}^*(\boldsymbol{\epsilon}_a{}^*)^T >, \\
&= \;\; < (\boldsymbol{\epsilon}_b + \hat{\mathbf{K}}^* \hat{\mathbf{d}}_b^{o*})(\boldsymbol{\epsilon}_b + \hat{\mathbf{K}}^* \hat{\mathbf{d}}_b^{o*})^T >, \\
&= \;\; \mathbf{B} + < \boldsymbol{\epsilon}_b(\hat{\mathbf{d}}_b^{o*})^T > \hat{\mathbf{K}}^{*T} + \hat{\mathbf{K}}^* < \hat{\mathbf{d}}_b^{o*} \boldsymbol{\epsilon}_b{}^T > + \hat{\mathbf{K}}^* < \hat{\mathbf{d}}_b^{o*}(\hat{\mathbf{d}}_b^{o*})^T > \hat{\mathbf{K}}^{*T}, \\
&= \;\; \mathbf{B} - \mathbf{B}\hat{\mathbf{H}}^{*T}\hat{\mathbf{K}}^{*T} - \hat{\mathbf{K}}^*\hat{\mathbf{H}}^*\mathbf{B} + \hat{\mathbf{K}}^*(\hat{\mathbf{R}}^* + \hat{\mathbf{H}}^*\mathbf{B}\hat{\mathbf{H}}^{*T})\hat{\mathbf{K}}^{*T}, \\
&= \;\; \mathbf{B} - \mathbf{B}\hat{\mathbf{H}}^{*T}(\hat{\mathbf{H}}^*\mathbf{B}\hat{\mathbf{H}}^{*T} + \hat{\mathbf{R}}^*)^{-1}\hat{\mathbf{H}}^*\mathbf{B} \\
&= \;\; \mathbf{B} - \hat{\mathbf{K}}^*\hat{\mathbf{H}}^*\mathbf{B}, \\
&= \;\; (\mathbf{I} - \hat{\mathbf{K}}^*\hat{\mathbf{H}}^*)\mathbf{B}, \qquad\qquad\qquad\qquad\qquad\qquad (7.30)
\end{aligned}
$$

using (7.21) and the assumption that the background errors, observation errors and model errors are uncorrelated.

The Best Linear Unbiased Estimate (BLUE) analysis is the solution to the non-erroneous linear 4DVar data assimilation problem with the minimum variance [65]. When no model error is present the Best Linear Unbiased Estimate (BLUE) of the true initial state $\mathbf{x}_0$ is known to have the analysis error covariance matrix $\mathbf{A} = (\mathbf{I} - \hat{\mathbf{K}}\hat{\mathbf{H}})\mathbf{B}$ with the optimal gain matrix $\hat{\mathbf{K}} = \mathbf{B}\hat{\mathbf{H}}^T(\hat{\mathbf{H}}\mathbf{B}\hat{\mathbf{H}}^T + \hat{\mathbf{R}})^{-1}$ [65] [77]. When model error is present, we account for model error in the comparison of the observations with the

model evolved state using $\hat{\mathbf{R}}^*$ as opposed to $\hat{\mathbf{R}}$. This replacement leads to the analysis error covariance matrix $\mathbf{A}^*$ (7.30) having the same form as the analysis error covariance matrix of the BLUE estimate. This implies that the analysis solution (7.27) is the optimal minimum variance estimate of the linear strong constraint 4DVar problem, when the model has error of a random nature at each time-step, and therefore implies that the analysis $\mathbf{x}^a{}_0{}^*$ (7.27) is more statistically accurate than the analysis $\mathbf{x}^a{}_0$ (7.23). Let us next compare the structure of $\mathbf{A}^*$ (7.30) with $\mathbf{A}$ (7.26).

An important point to be noted is that when model error is not accounted for, the analysis error covariance matrix $\mathbf{A}$ (7.26) is unbounded. As the model error increases, the diagonal of the matrix term $\hat{\mathbf{K}}\hat{\mathbf{Q}}^*\hat{\mathbf{K}}^T$ present in (7.26) increases and therefore the error variances of the analysis variables increase. The model errors could increase significantly leading to a less accurate analysis $\mathbf{x}^a{}_0$ (7.23) than the background state $\mathbf{x}^b$. Whereas the analysis error covariance matrix $\mathbf{A}^*$ (7.30) is bounded by the background error covariance matrix $\mathbf{B}$. As the model error increases, the analysis error covariance matrix $\mathbf{A}^*$ tends to the background error covariance matrix $\mathbf{B}$. Note that although $\mathbf{A}^*$ can tend to the background error covariance matrix $\mathbf{B}$, it cannot be equal $\mathbf{A}^* \neq \mathbf{B}$ for non-zero $\mathbf{B}$ and $\hat{\mathbf{R}}^*$.

7.3.2 Erroneous scalar model

We consider an erroneous linear scalar model over one time-step of the form $x_i = \beta_i^e x_{i-1}$, with the erroneous model constant $\beta_i^e \in \mathbb{R}$ describing the evolution of the model state from time t_{i-1} to time t_i. To obtain the true model state at time t_i,

$$x^t{}_i = \beta_i x^t{}_{i-1} = \beta_i^e x^t{}_{i-1} + \eta_i, \tag{7.31}$$

where $\beta_i \in \mathbb{R}$ is the true model constant and η_i is the random error normally distributed around a zero mean with variance σ_{qi}^2. We propose the following theorem.

Theorem 7.3.1. *Consider a direct observation y_i of the model state (where the observation operator $h_i = 1$) with error normally distributed around a zero mean with variance σ_{obi}^2. When an erroneous linear scalar model of the form $x_i = \beta_i^e x_{i-1}$, with corresponding additive model error η_i with zero mean and variance σ_{qi}^2, is used within the strong constraint 4DVar cost function, the use of $\sigma_{obi}^{*2} = \sigma_{obi}^2 + \sigma_{qi}^2$ to replace σ_{obi}^2 results in an analysis with a smaller error variance.*

Proof. We assume that we have a background estimate x^b of the model state at time t_{i-1} with error normally distributed around a zero mean with variance σ_b^2. We are only considering two times; the background time t_{i-1} and the observation time t_i, therefore to simplify notation we set $i = 1$. The analysis when the erroneous linear scalar model $x_1 = \beta^e x_0$ is used within the strong constraint 4DVar cost function (2.19) and the error in the model is not accounted for is as follows,

$$x^a{}_0 \;=\; x^b + k\,d^o{}_b, \tag{7.32}$$

where the gain constant $k = \frac{\sigma_b^2 \beta^e}{\sigma_b^2 \beta^{e2} + \sigma_{ob}^2}$ and the innovation vector $d^o{}_b = y_1 - \beta^e x^b$.

Using equation (7.11), we evaluate the combined error variance $\sigma_{ob}^{*2} = \sigma_{ob}^2 + \sigma_q^2$. The analysis when the erroneous linear scalar model $x_1 = \beta^e x_0$ is used within the strong constraint 4DVar cost function (2.19) and the error in the model is accounted for by replacing σ_{ob}^{*2} with σ_{ob}^2 is as follows,

$$x^a{}_0{}^* \;=\; x^b + k^* d^o{}_b{}^*, \tag{7.33}$$

where the gain constant $k^* = \frac{\sigma_b{}^2 \beta^e}{\sigma_b{}^2 \beta^{e2} + \sigma_{ob}{}^{*2}}$ and the innovation vector $d^o{}_b{}^* = y_1 - \beta^e x^b$.

We denote $\sigma_a{}^2$ to be the analysis error variance of $x^a{}_0$ (7.32) where,

$$\sigma_a{}^2 = \sigma_b{}^2 - \frac{\beta^{e2} \sigma_b{}^4}{\beta^{e2} \sigma_b{}^2 + \sigma_{ob}{}^2} + \frac{\beta^{e2} \sigma_b{}^4 \sigma_q{}^2}{(\beta^{e2} \sigma_b{}^2 + \sigma_{ob}{}^2)^2}, \tag{7.34}$$

using (7.26) and we denote $\sigma_a{}^{*2}$ to be the analysis error variance of $x^a{}_0{}^*$ (7.33) where,

$$\sigma_a{}^{*2} = \sigma_b{}^2 - \frac{\beta^{e2} \sigma_b{}^4}{\beta^{e2} \sigma_b{}^2 + \sigma_{ob}{}^2 + \sigma_q{}^2}, \tag{7.35}$$

using (7.30). To enable us to assess which analysis $x^a{}_0$ or $x^a{}_0{}^*$ is more statistically accurate we subtract $\sigma_a{}^{*2}$ (7.35) from $\sigma_a{}^2$ (7.34),

$$\sigma_a{}^2 - \sigma_a{}^{*2} = \frac{\beta^{e2} \sigma_q{}^4 \sigma_b{}^4}{(\beta^{e2} \sigma_b{}^2 + \sigma_{ob}{}^2)^2 (\beta^{e2} \sigma_b{}^2 + \sigma_{ob}{}^2 + \sigma_q{}^2)}, \tag{7.36}$$

$$\geq 0. \tag{7.37}$$

We have proven that the analysis $x^a{}_0{}^*$ is either statistically more accurate than $x^a{}_0$ or at worst of the same statistical accuracy. However, the only case where $x^a{}_0{}^*$ has the same error variance as $x^a{}_0$ is when $\sigma_q{}^2 = 0$ or $\sigma_b{}^2 = 0$ or when the erroneous model constant $\beta^e = 0$. Let us make the reasonable assumption that we have a non-zero model constant $\beta^e \neq 0$. In the strong constraint 4DVar formulation, we assume the background error variance is non-zero $\sigma_b{}^2 \neq 0$. The model we are considering $x_1 = \beta^e x_0$ is of an erroneous nature and therefore $\sigma_q{}^2 \neq 0$. Hence we can conclude the proof and state,

$$\sigma_a{}^2 - \sigma_a{}^{*2} \quad > \quad 0, \tag{7.38}$$

therefore the analysis $x^a{}_0{}^*$ is statistically more accurate than $x^a{}_0$. $\qquad\square$

We next investigate when the difference in analysis accuracy of $x^a{}_0{}^*$ compared to $x^a{}_0$ is most significant. Let us first denote $r = \frac{\sigma_b{}^2}{\sigma_{ob}{}^2}$. Then we can re-write equation (7.36) as follows,

$$\sigma_a{}^2 - \sigma_a{}^{*2} \quad = \quad \frac{\sigma_q{}^4 r^2 \beta^{e2}}{(\beta^{e2}\sigma_b{}^2 + \sigma_{ob}{}^2 + \sigma_q{}^2)(\beta^{e2}r + 1)^2}. \tag{7.39}$$

Equation (7.39) enables us to state that improvement in analysis accuracy of $x^a{}_0{}^*$ from $x^a{}_0$ is most significant when there is; large model error variance, large background error variance and small observation error variance. The size of the difference (7.39) is also dependent on the erroneous model constant β^e, which we will investigate the significance of later in this section. The reliance of the increase in accuracy $\sigma_a{}^2 - \sigma_a{}^{*2}$ on both the model error variance $\sigma_q{}^2$ and the size of the ratio r is demonstrated in Figure 7.2. The plot in Figure 7.2 uses a range of model error and background error variances, with erroneous model constant $\beta^e = 1$ and the observation error variance set $\sigma_{ob}{}^2 = 0.5$. Figure 7.2 clearly shows as the ratio r increases, the difference $\sigma_a{}^2 - \sigma_a{}^{*2}$ increases and when the model error variance $\sigma_q{}^2$ increases, the difference $\sigma_a{}^2 - \sigma_a{}^{*2}$ increases.

Next let us relate these results to use of a generic erroneous model in the strong constraint 4DVar cost function (2.19). When there is either an increase in background error variance or decrease in observation error variance, more weight is given to the $\mathcal{J}_{ob}$ term comparing the observation and model evolved state in observation space.

When model error is not accounted for in this term, the analysis is less accurate. The larger the model error, the more significant the statistical difference between the accuracy of the resulting analysis $x^a_0{}^*$ and x^a_0.
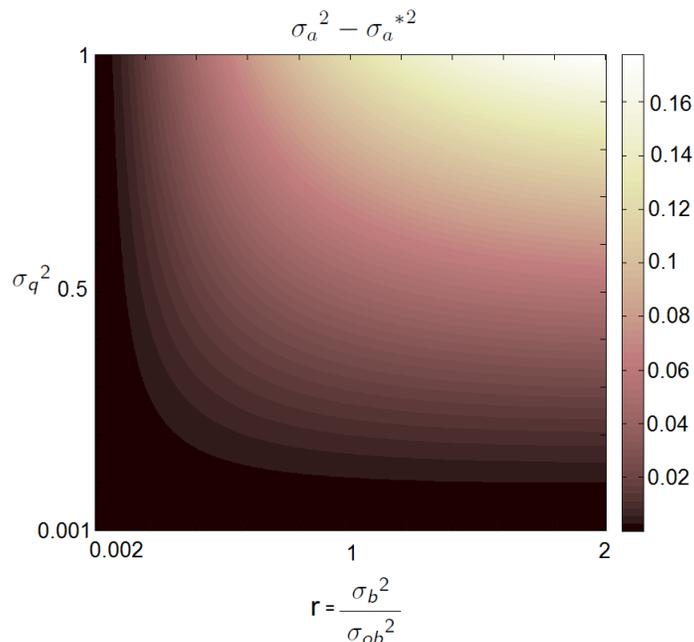
Figure 7.2: Difference in analysis error variance $\sigma_a{}^2 - \sigma_a{}^{*2}$ (7.39) for the case where the erroneous model is $x_1 = \beta^e x_0$ with $\beta^e = 1$ and $\sigma_{ob}{}^2 = 0.5$.

Let us refer back to the simple scalar example to further investigate properties of the difference in the analysis error variances $\sigma_a{}^2 - \sigma_a{}^{*2}$ (7.39). We next evaluate the rate of change of the analysis error covariances $\sigma_a{}^2$ and $\sigma_a{}^{*2}$ with respect to the model error variance. The analysis error variances differentiated with respect to the model error variance $\sigma_q{}^2$ are as follows,

$$\frac{\partial \sigma_a{}^2}{\partial \sigma_q{}^2} = k^2, \tag{7.40}$$

$$\frac{\partial \sigma_a{}^{*2}}{\partial \sigma_q{}^2} = k^{*2}, \tag{7.41}$$

154

using the equations (7.26) and (7.30), where $k = \frac{\sigma_b{}^2 \beta^e}{\sigma_b{}^2 \beta^{e2} + \sigma_{ob}{}^2}$ and $k^* = \frac{\sigma_b{}^2 \beta^e}{\sigma_b{}^2 \beta^{e2} + \sigma_{ob}{}^{*2}}$.
Therefore the rate of change of $\sigma_a{}^2$ with respect to model error variance is not a function of $\sigma_q{}^2$ and so it increases linearly as $\sigma_q{}^2$ increases, whereas the rate of change of $\sigma_a{}^{*2}$ with respect to the model error variance is a function of $\sigma_q{}^2$.

We next illustrate the relationship between model error variance and the analysis error variance. We firstly set $\beta^e = 1$, $\sigma_b{}^2 = 0.5$, and $\sigma_{ob}{}^2 = 0.5$, which leads to $k = 0.5$. Figure 7.3 shows the analysis error variance both for when the model error is not accounted for and when the model error is accounted for with the combined error statistics. Specifically, Figure 7.3 shows the relationship between these analysis error variances and the size of the model error variance. Figure 7.3 clearly shows that when the model error variance $\sigma_q{}^2$ increases the analysis error variance $\sigma_a{}^2$ increases at a constant rate and is unbounded. Whereas the analysis error variance increases $\sigma_a{}^{*2}$ also increases as $\sigma_q{}^2$ increases, but at a slower rate and is bounded above by the background error variance $\sigma_b{}^2$. This property is shown clearly in Figure 7.3. For all sizes of model error present, the analysis $x^a{}_0{}^*$ (7.33) has improved accuracy statistics when compared with the background x^b. However, whether the analysis $x^a{}_0$ (7.32) is an improvement from the background x^b is dependent on the size of the model error. With significantly large model error, specifically $\sigma_q{}^2 > (\beta^{e2} \sigma_b{}^2 + \sigma_{ob}{}^2)$, the analysis $x^a{}_0$ (7.32) is of less accuracy than the background x^b. This results from this experiment will be known as the 'control run' for further experiments in this section.
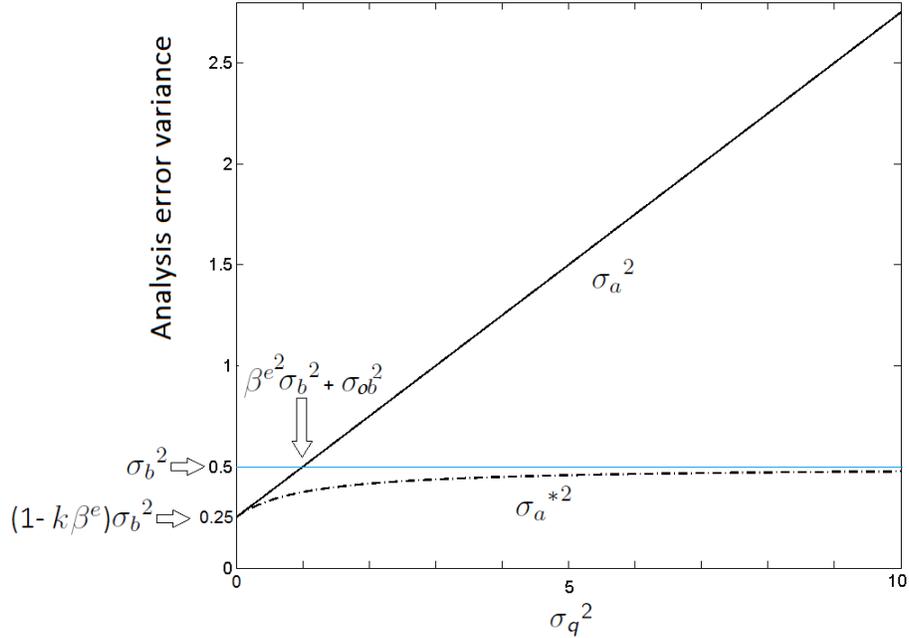
Figure 7.3: Analysis error variance $\sigma_a{}^2$ (black solid line) and $\sigma_a{}^{*2}$ (black dotted line) for the case where the erroneous model is $x_1 = \beta^e x_0$ with erroneous scalar constant $\beta^e = 1$. Both the background error variance $\sigma_b{}^2 = 0.5$ (blue line) and observation error variance $\sigma_{ob}{}^2 = 0.5$, which leads to $k = 0.5$.

We know it is not just the size of the model error that the analysis accuracy depends on, but also the ratio between the accuracy of the background and observations. We show the results from the control run in Figure 7.4 (black lines), with $0 \leq \sigma_q{}^2 \leq 1$, where $r = 1$. We then increase the ratio so that $r = 2$, by increasing the background error variance from the control so that $\sigma_b{}^2 = 1$, with the results shown in blue in Figure 7.4. We then decrease the ratio so that $r = \frac{1}{2}$, by increasing the observation error variance from the control so that $\sigma_{ob}{}^2 = 1$, with the results shown in red in Figure 7.4. It is clearly shown in Figure 7.4 that the larger r is the greater the difference is between the analysis error variances $\sigma_a{}^2$ and $\sigma_a{}^{*2}$.

What should also be observed is that when no model error is present, the same accuracy is obtained for both the 'blue' and 'red' conditions. However, when model

error is present, the analysis accuracy is more sensitive to an increase to the background error than an increase to the observation error. The rate of change of the analysis error variance with respect to an increase in model error variance is larger when $r = 2$ (blue lines) than when $r = \frac{1}{2}$ (red lines) as shown in Figure 7.4. Let us relate these results to the strong constraint 4DVar cost function (2.19). When there is an increase to the background error variance, more weight is given to the $\mathcal{J}_{ob}$ term comparing the observation and model evolved state in observation space, this term contains the erroneous model. However when there is an increase to the observation error variance, less weight is given to the $\mathcal{J}_{ob}$ term comparing the observation and model evolved state in observation space, which contains the erroneous model. Therefore, an increase to the background error variance has a more significant detrimental effect on the analysis accuracy, than an increase to the observation error variance (both when the model error is accounted for and when the model error is not).

We highlight that when random error is present in a model and not accounted for in the 4DVar process, that if the model error is sufficiently large, it is possible that a more accurate analysis can be obtained even with less accurate observations. When model error is present and not accounted for in 4DVar, the observation error covariance matrix $\hat{\mathbf{R}}$ in the $\mathcal{J}_{ob}$ term of the cost function accounts only for observation error in the comparison of the model evolved state and observations. Therefore, if the observations are very accurate, the variances in $\hat{\mathbf{R}}$ are specified to be very small and therefore the estimation of the initial conditions is adjusted so that the model evolved initial state is in the close vicinity of the observations. However, the fact the model evolving the initial conditions is erroneous is not accounted for, means the analysis may be far from the true initial conditions. We demonstrate this property, with the erroneous scalar model, in Figure 7.4 where the black solid line ($\sigma_{ob}{}^2 = 0.5$) crosses both the red dotted and red solid lines ($\sigma_{ob}{}^2 = 1$) at $\sigma_q{}^2 = 0.5$ and $\sigma_q{}^2 = 0.6$ respectively. When the black solid line (not accounting for model error) crosses the

respective red lines, the analysis error variance is higher even though the observations are more accurate.
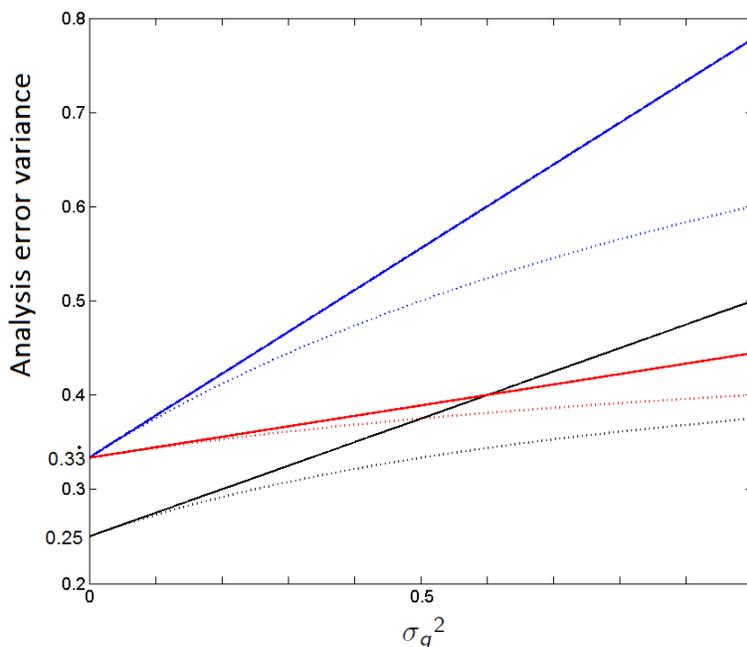
Figure 7.4: Analysis error variance $\sigma_a{}^2$ (solid lines) and $\sigma_a{}^{*2}$ (dotted lines) for the case where the erroneous model is $x_1 = \beta^e x_0$ with erroneous scalar constant $\beta^e = 1$. Black lines: $\sigma_b{}^2 = \sigma_{ob}{}^2 = 0.5$, $r = 1$. Red lines: $\sigma_b{}^2 = 0.5$, $\sigma_{ob}{}^2 = 1$, $r = \frac{1}{2}$. Blue lines: $\sigma_b{}^2 = 1$, $\sigma_{ob}{}^2 = 0.5$, $r = 2$.

Let us refer back to the equation (7.39) where we stated that the difference between the analysis error variances $\sigma_a{}^2 - \sigma_a{}^{*2}$ is also dependent on the value of the erroneous model constant β^e. We firstly investigate the effect of increasing and decreasing this model constant β^e has on the analysis accuracy of $x^a{}_0$ (7.32) and $x^a{}_0{}^*$ (7.33). We show the analysis accuracy results from the control run (black lines) with $\beta^e = 1$ in Figure 7.5, with $0 \le \sigma_q{}^2 \le 1$. Equations (7.34) and (7.35) have been used to plot these lines. These can be compared with the corresponding results when increasing the erroneous model constant to $\beta^e = 2$ (red lines) and decreasing the model constant to $\beta^e = \frac{1}{2}$ (blue lines). Again when model error is present, the accuracy of $x^a{}_0{}^*$ (7.33) is improved when compared to the accuracy of $x^a{}_0$ (7.32). However, the analysis error

variance differs depending on the size of the model constant β^e. The larger the model constant β^e, the smaller the analysis error variance.
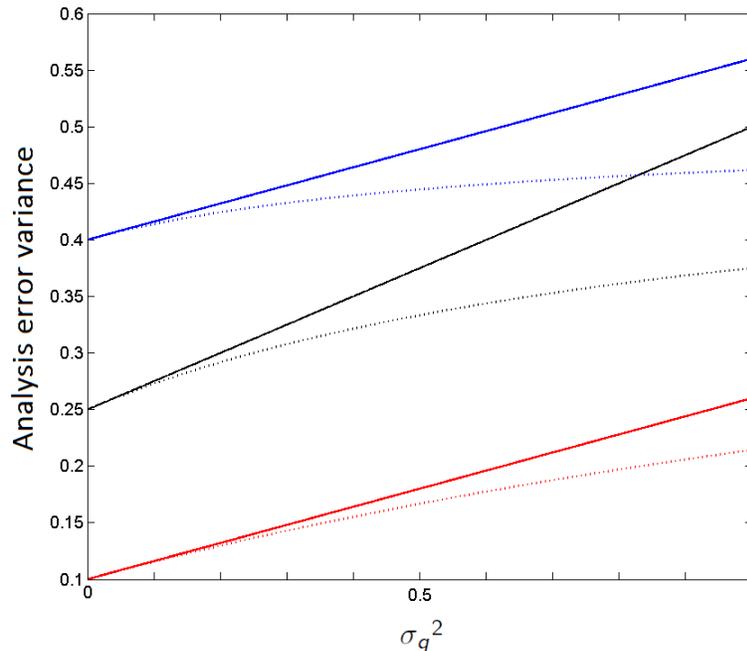
Figure 7.5: Analysis error variance $\sigma_a{}^2$ (solid lines) and $\sigma_a{}^{*2}$ (dotted lines) for the case where the erroneous model is $x_1 = \beta^e x_0$. Both the background error variance $\sigma_b{}^2 = 0.5$ and observation error variance $\sigma_{ob}{}^2 = 0.5$. Black lines: $\beta^e = 1$. Red lines: $\beta^e = 2$. Blue lines: $\beta^e = \frac{1}{2}$.

We are also interested in the effect the specification of the model constant has on the difference between the variances $\sigma_a{}^2 - \sigma_a{}^{*2}$. We plot difference $\sigma_a{}^2 - \sigma_a{}^{*2}$ term (7.36) with respect to changing values of the model constant β^e, shown in Figure 7.6, where the model error variance is fixed at $\sigma_q{}^2 = 0.1$ and the background error and observation error variances are fixed at $\sigma_b{}^2 = \sigma_{ob}{}^2 = 0.5$. We are aware that models often contain fixed constants, but wish to evaluate the effect the size of the model constant has on the analysis accuracy with this set up. The maximum difference in the analysis error variances $\sigma_a{}^2 - \sigma_a{}^{*2}$, shown in Figure 7.6, occurs when $\beta^e = 0.73$ (2.s.f). We expect the difference between $\sigma_a{}^2 - \sigma_a{}^{*2}$ to decrease as the model constant

β^e increases, as can be seen in Figure 7.6 when $\beta^e > 0.73$ (2.s.f). This is because as the model constant increases, the ratio of the model error with the model constant decreases. So the effect the model error has on the evolution of the model variable becomes less significant.

Both $\sigma_a{}^2$ and $\sigma_a{}^{*2}$ decrease as the model constant β^e increases. This is due to the experimental set up we are using. Specifically as β^e gets larger and larger, then the initial condition will become smaller and smaller as the model error variance is held constant. In practice we would expect the model error variance to increase as the erroneous model error constant increases. However, what is interesting is the increase in the difference $\sigma_a{}^2 - \sigma_a{}^{*2}$ from $\beta^e = 0$ until the maximum at $\beta^e = 0.73$ (2.s.f). When the size of the erroneous model constant β^e is insignificant in comparison with the size of the model error, little improvement can be made to the background model state, when accounting for model error, as little information is gained from the comparison of the model state and observations. This means the analysis error variance $\sigma_a{}^{*2}$ will be of a similar size to $\sigma_b{}^2$. When the size of the erroneous model constant β^e is insignificant in comparison with the size of the model error, obviously when not accounting for model error very little improvement can be made to the background (if any), so the analysis will have a error variance $\sigma_a{}^2$ of a similar size to $\sigma_b{}^2$ as well. The specification of the model constant β^e which creates the maximum difference $\sigma_a{}^2 - \sigma_a{}^{*2}$ (7.36) will increase as the background becomes more accurate and either or both the observations and model both become less accurate. In general model coefficients are of course pre-specified in atmosphere and ocean operational models, but is interesting to note that with different model coeffients, there will different improvements to the analysis accuracy when replacing $\hat{\mathbf{R}}$ with $\hat{\mathbf{R}}^*$ in the strong constraint 4DVar cost function.
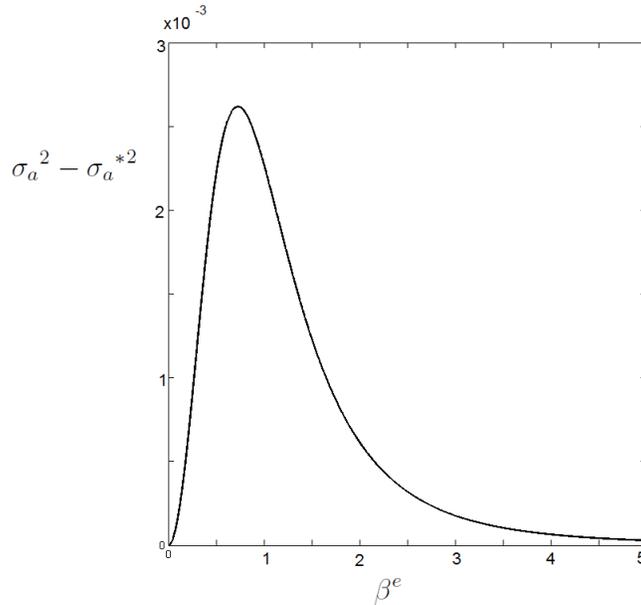
Figure 7.6: Difference in the analysis error variances $\sigma_a{}^2 - \sigma_a{}^{*2}$ for varying values of the model constant β^e. The model error variance is fixed at $\sigma_q{}^2 = 0.1$ and both $\sigma_b{}^2 = \sigma_{ob}{}^2 = 0.5$.

We restate our hypothesis that the analysis $x^a{}_0{}^*$ (7.27) is statistically more accurate than the analysis $x^a{}_0$ (7.23), by using $\hat{\mathbf{R}}^*$ as opposed to $\hat{\mathbf{R}}$ in the strong constraint 4DVar cost function. We next demonstrate the application of our developed method, to account for random error in a model and hence improve the analysis accuracy, firstly with use of the linear advection equation and secondly with an idealized coupled atmosphere-ocean model.

7.4   Numerical experiments: Linear advection equation

We now demonstrate that the replacement of the observation error covariance matrix $\hat{\mathbf{R}}$, with combined error statistics in $\hat{\mathbf{R}}^*$, leads to an analysis of greater statistical accuracy when random error is present in the linear advection equation. We use the

161

numerical set up as described previously in Section 7.2.2. We take our erroneous linear model to be the time-stepping solution of the linear advection equation over the spatial domain $x \in [0, 10)$, as described in Section 5.1.1, with the spatial step $\Delta x = 0.1$ and time-step $\Delta t = 0.1$. The true model state at each time t_i, differs from the erroneous model state by the vector of random error $\boldsymbol{\eta}_i$, as defined by equation (7.1). We use an assimilation window length of eight time-steps. We assume we have direct observations of all spatial points in the vector $\mathbf{y}_i$, with $\mathbf{H} = \mathbf{I}$, every two time-steps (i=2, 4, 6, 8).

We take the true initial conditions, as described in Section 5.1.1, and run the erroneous model over 8 time-steps by adding random error $\boldsymbol{\eta}_i$ at each time-step produced as noise using the variances in $\mathbf{Q}_i = 0.01\mathbf{I}$ (spatially uncorrelated). We repeat this process to obtain a sample of 100 model trajectories, shown in Figure 7.7 at the final time t_8. The linear advection equation run with no random error present is also shown in Figure 7.7 (black dotted line) at time t_8. The general behaviour of the passive tracer is maintained. However, this level of random error in the model leads to significant variations in the shape of the passive tracer.
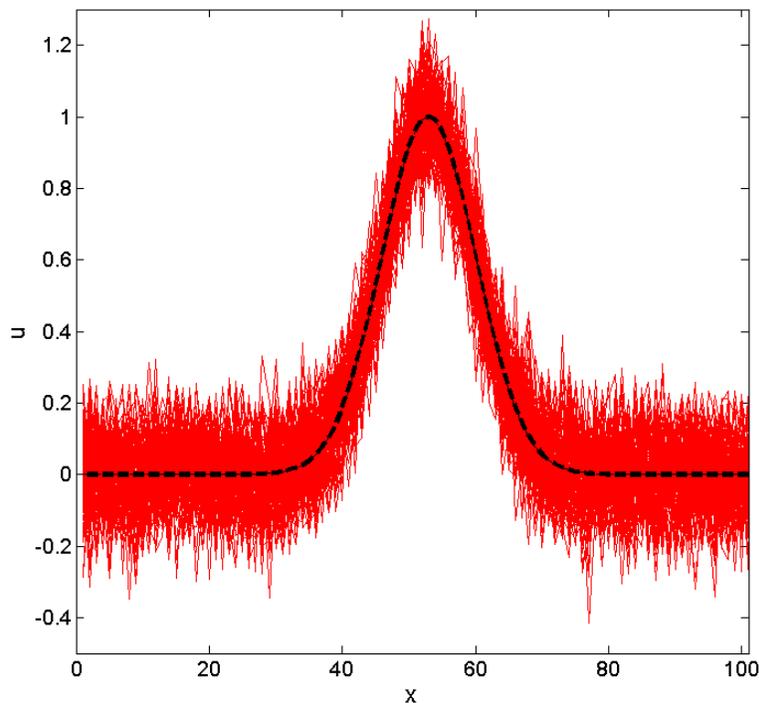
Figure 7.7: The linear advection equation run from from t_0 to t_8 with a vector of random error $\boldsymbol{\eta}_i$ ($\mathbf{Q}_i = 0.01$) added at each time-step (red lines). Results shown at time t_8 for a sample of 100 independent model runs. The linear advection equation run from from t_0 to t_8 with no random error present (black dotted line).

We now define two data assimilation methods that will be performed and compared.

- Method 1: the evaluation of the analysis $\mathbf{x}_0^a$ using the 4DVar cost function (2.19) with no changes,

- Method 2: the evaluation of the analysis $\mathbf{x}_0^{a*}$ by replacing $\hat{\mathbf{R}}$ in the 4DVar cost function (2.19) with the combined error covariance matrix $\hat{\mathbf{R}}^*$ (7.11).

The elements of $\hat{\mathbf{R}}^*$ are calculated using equation (7.11). The RMSE (Root Mean Square Error) for a sample of analysis outputs from Method 1 and Method 2 are

163

Conditions	Covariance matrix	Variance
A	$\mathbf{Q}_i$	0.01
A	$\mathbf{B}$	0.04
A	$\mathbf{R}_i$	0.04
B	$\mathbf{Q}_i$	0.01
B	$\mathbf{B}$	0.04
B	$\mathbf{R}_i$	0.0016
C	$\mathbf{Q}_i$	0.04
C	$\mathbf{B}$	0.04
C	$\mathbf{R}_i$	0.04

Table 7.2: List of data assimilation conditions for error covariance matrices used in strong constraint 4DVar with the linear advection equation with random error. Both $\mathbf{Q}_i$ and $\mathbf{R}_i$ are diagonal matrices, whereas $\mathbf{B}$ is constructed using the SOAR function with correlation length scale $\mathcal{L} = 0.4$.

calculated and compared. We conduct the following steps 100 times to produce a sample size of 100 analysis values $\mathbf{x}^a{}_0$ and 100 analysis values $\mathbf{x}^a{}_0{}^*$ for Method 1 and Method 2 respectively:

1. Produce a background vector $\mathbf{x}^b$ by adding noise, consistent with the statistics prescribed in $\mathbf{B}$, to the true initial state $\mathbf{x}^t{}_0$.

2. Produce a random model error vector $\boldsymbol{\eta}_i$ for each time $t_1, ..., t_8$ using the statistics specified in the model error covariance matrix $\mathbf{Q}_i$ (i=1,...,8).

3. Produce a vector of observations $\mathbf{y}_i$ at each observation time t_i, (i =2, 4, 6, 8) by evaluating the true model state vector $\mathbf{x}^t{}_i$ with equation (7.1) and adding noise in proportion to the error statistics specified in $\mathbf{R}_i$ (i=2,4,6,8).

4. Compute both the analysis $\mathbf{x}^a_0$ and the analysis $\mathbf{x}^{a*}_0$ using 4DVar as described by Method 1 and Method 2 respectively.

The resulting analysis RMSEs, using conditions A as specified in Table 7.2, are shown in the left two bars in Figure 7.8. When accounting for the model error, in Method 2

with $\hat{\mathbf{R}}^*$, a more accurate analysis is obtained. The calculated combined error covariance matrix $\hat{\mathbf{R}}^*$ at each observation time is diagonal with the increasing variances. In this case $\sigma_{ob}^{*2} = 0.06$ at t_2, $\sigma_{ob}^{*2} = 0.08$ at t_4, $\sigma_{ob}^{*2} = 0.10$ at t_6 and $\sigma_{ob}^{*2} = 0.12$ at t_8. This represents the increase in uncertainty of the model trajectory throughout the assimilation window, due to the errors present in the model. The data assimilation scheme therefore puts less weight on the comparison of the observations with the model evolved state as time increases and hence allows a more accurate estimation of the initial conditions. We illustrate the analysis outputs (Method 1 and Method 2) from one data assimilation cycle (from the sample) in Figure 7.9.
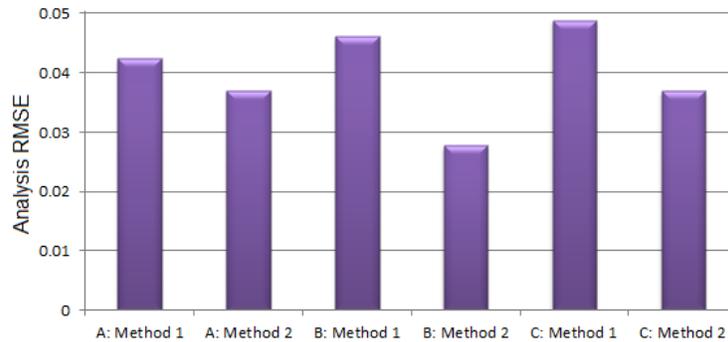
Figure 7.8: Results comparing the analysis RMSE (at time t_0 only) from a sample of 100 data assimilation runs, for both Method 1 (not accounting for the model error) and Method 2 (accounting for model error with the combined error covariance matrix $\hat{\mathbf{R}}^*$ (7.11)).
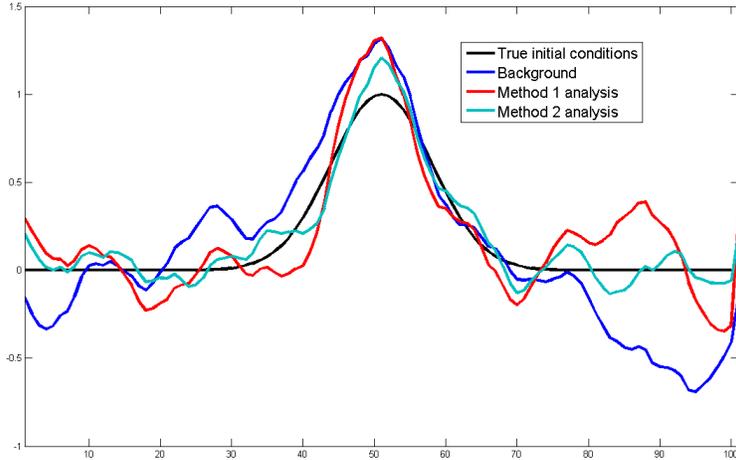
Figure 7.9: Results comparing the analysis from both Method 1 (not accounting for the model error) and Method 2 (accounting for the model error with the combined error covariance matrix $\hat{\mathbf{R}}^*$ (7.11)).

Work in Section 7.3 showed that, for a scalar model, there is a more significant increase in analysis accuracy (of x_0^{a*} compared with x_0^a) when the observations increase in accuracy (in comparison with the background accuracy) and when the size of the model error increases. These conclusions are not limited to models of a scalar nature and we demonstrate that these properties also hold when using the linear advection equation with random error in 4DVar. Firstly, we reduce the standard deviations of the observation errors by a factor of five so that $\mathbf{R}_i = 0.0016\mathbf{I}$ ($i =$2, 4, 6, 8), as shown by conditions B in Table 7.2. The corresponding combined error covariance matrix at each observation time is diagonal with the variances in the combined error matrix increasing with time. In this case $\sigma_{ob}^{*2} = 0.0216$ at t_2, $\sigma_{ob}^{*2} = 0.0416$ at t_4, $\sigma_{ob}^{*2} = 0.0616$ at t_6 and $\sigma_{ob}^{*2} = 0.0816$ at t_8. We ran a sample of 100 4DVar experiments, as described above, using both Method 1 and Method 2. The resulting analysis RMSE for Method 1 and analysis RMSE for Method 2 are shown in the central two bars in Figure 7.8. These results demonstrate that an increase in observation accuracy leads to a more significant increase in analysis accuracy, when accounting for the model

error as opposed to not. It is perhaps interesting to remark that when the model error is not accounted for (Method 1), this increase in observation accuracy causes a degradation in the analysis $\mathbf{x}^a{}_0$, as can be seen when comparing Method 1 results from experiment A with Method 1 results from experiment B. This is because, with a decrease in observation variance, the trajectory of the erroneous model is confined to be closer to the observations, even at the end of the assimilation window. When there is significant error present in the model, the erroneous model trajectory should not be in the close vicinity of the observations towards the end of the assimilation window and if it is this can cause degradation to the estimation of the initial conditions. However, when accounting for the model error in the analysis $\mathbf{x}^a{}_0{}^*$ in Method 2, an improvement is made to the analysis when the observations are of increased accuracy. This is because the model error is accounted for appropriately in the combined error matrix, meaning that the erroneous model trajectory is allowed to depart further from the observations in order to best estimate the analysis.

Next we increase the standard deviation of the model error variance by a factor of two so that $\mathbf{Q}_i = 0.04\mathbf{I}$ $(i = 1, ..., 8)$, as stated by conditions C in Table 7.2. The corresponding combined error covariance matrix at each observation time is diagonal with the variances in the combined error matrix increasing over time. In this case $\sigma_{ob}{}^{*2} = 0.12$ at t_2, $\sigma_{ob}{}^{*2} = 0.20$ at t_4, $\sigma_{ob}{}^{*2} = 0.28$ at t_6 and $\sigma_{ob}{}^{*2} = 0.36$ at t_8. The resulting analysis RMSEs are shown in Figure 7.8 (right two bars), where the analysis $\mathbf{x}^a{}_0{}^*$ from Method 2 is again more accurate than when not accounting for the model error in the Method 1 analysis $\mathbf{x}^a{}_0$. When not accounting for this significant increase in model error, there is a significant decrease in analysis accuracy (Method 1). However, when accounting for this significant increase in model error, using the combined error covariance matrix, this has minimal effect on the accuracy of the analysis $\mathbf{x}^a{}_0{}^*$, as can be seen by comparing the analysis RMSE from experiment A with the analysis RMSE from experiment C. We have demonstrated that the work in Section 7.3 for a scalar

167

model; showing the significant increase in analysis accuracy when the observations increase in accuracy (in comparison with the background accuracy) and when the size of the model error increases, also holds when using a non scalar erroneous linear model.

The application of our developed method, replacing $\hat{\mathbf{R}}$ with $\hat{\mathbf{R}}^*$ (7.11), has been demonstrated to be successful in obtaining an analysis of greater accuracy, when the model used in 4DVar is an erroneous linear model. We next extend our investigation by applying the developed method with an erroneous model of a nonlinear nature.

7.5 Numerical experiments: Idealized coupled atmosphere-ocean model

We now demonstrate that use of combined model error and observation error statistics can account for random error in a model of a nonlinear nature and hence improve the analysis accuracy. We use the discretized solution to the idealized coupled atmosphere-ocean model, as described in Section 5.2, as our nonlinear erroneous model of the form,

$$\mathbf{x}_i = \mathcal{M}^e_{\{i-1\}\to i}(\mathbf{x}_{i-1}) \qquad i = 1, 2, .., 50 \tag{7.42}$$

where $\mathbf{x}_i = (x_i\ y_i\ z_i\ w_i\ v_i)^T$ is the model state vector consisting of both the atmosphere state variables x_i, y_i, z_i and ocean variables w_i, v_i at time t_i. The true model $\mathcal{M}_{\{i-1\}\to i}$ is as follows,

$$\begin{aligned}
\mathbf{x}^t{}_i &= \mathcal{M}_{\{i-1\}\to i}(\mathbf{x}^t{}_{i-1}) & i = 1, 2, ..., 50 \\
&= \mathcal{M}^e{}_{\{i-1\}\to i}(\mathbf{x}^t{}_{i-1}) + \boldsymbol{\eta}_i & (7.43)
\end{aligned}$$

where the vector of model error $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$.

For experiments in this section we use an assimilation window length of 50 time-steps, where the model time-step length is $\Delta t = 0.01$. We define the true initial conditions; $x^t{}_0 = -3.4866$, $y^t{}_0 = -5.7699$, $z^t{}_0 = 18.341$, $w^t{}_0 = -10.7175$ and $v^t{}_0 = -7.1902$, which are on the coupled model attractor. We initially specify the model error covariance matrix $\mathbf{Q}_i$ at each time t_i ($i = 1, 2, ..., 50$) to be diagonal, with variances along the diagonal set to 0.02, 0.02, 0.2, 0.01, 0.01. With this level of model error variance, the general behaviour of the coupled model is maintained, but there are significant variations in the model trajectories of both the atmospheric and oceanic model state variables throughout the time window. This property can be seen in Figure 7.10, where 100 possible 'true' model trajectories (red) have been evaluated using equation (7.43) run from the same initial conditions. Each vector of model error $\boldsymbol{\eta}_i$ has been independently produced as noise using the statistics specified in $\mathbf{Q}_i$. The model trajectories of the coupled model $\mathcal{M}^e{}_{\{i-1\}\to i}$ with no model error vectors added (7.42) are also shown in Figure 7.10 (black lines).
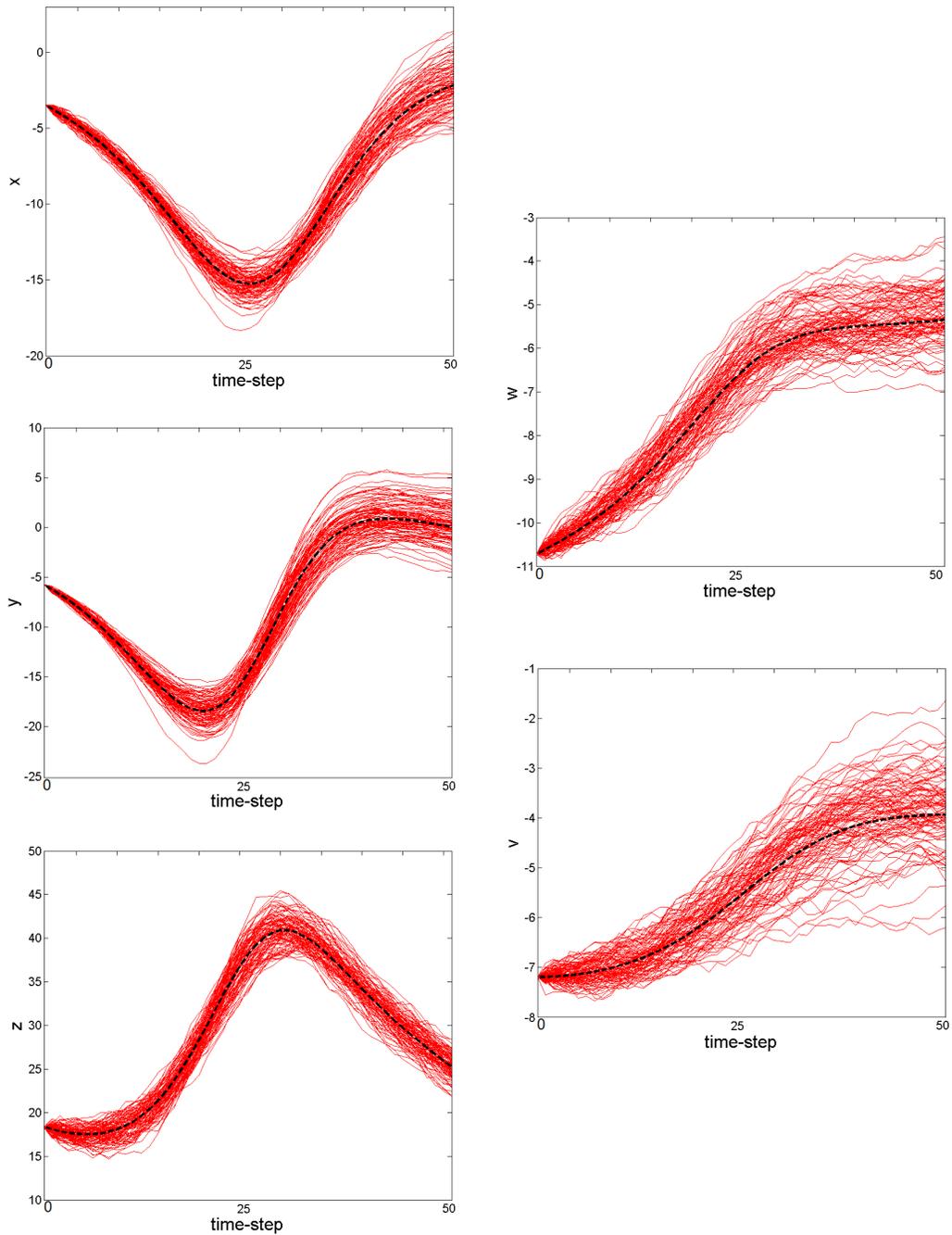
Figure 7.10: Idealized coupled model (7.43) run 100 times, from the same initial conditions, with random error present at each time-step (red lines). The model error covariance matrix $\mathbf{Q}_i$ at each time t_i $(i = 1, 2, ..., 50)$ is diagonal, with variances along the diagonal set to 0.02, 0.02, 0.2, 0.01, 0.01. Idealized coupled model (7.42) run from the same initial conditions, where no vectors of model error have been added (black dotted lines).

The forms of the erroneous model (7.42) and true model (7.43) are the same as those used in the derivation of the combined model error and observation error covariance matrix $\hat{\mathbf{R}}^*$ in Section 7.1, with the exception that here nonlinear model equations are used, as opposed to linear model matrices. One of the key objectives in this section is to show that the theory developed with linear models, earlier in the chapter, is also successfully applicable using models of a nonlinear nature.

We initially specify a diagonal background error covariance matrix $\mathbf{B}$ with standard deviations approximately 10% of the true initial conditions (specifically variances 0.1, 0.3, 3.4, 1.1 and 0.52 along the diagonal). We have direct observations of each model state variable present every 10 time-steps with observation operator $\mathbf{H}_i = \mathbf{I}$. The observation error covariance matrix $\mathbf{R}_i$ is diagonal at each observation time t_i in the assimilation window, initially set with standard deviations approximately 2% of the maximum absolute value of each respective variable. Specifically $\mathbf{R}_i$ contains the values 0.09, 0.09, 0.81, 0.04, 0.04 along the diagonal. Later in this section we investigate the sensitivity of the results to the specification of $\mathbf{B}$, $\mathbf{R}_i$ and $\mathbf{Q}_i$.

We have discussed the capabilities of NWP centres, such as ECMWF, for estimating the diagonal entries of the background model error covariance matrix evolved using the model matrix and subsequently mapped to observation space in Section 3.2 using the randomisation method. If our developed method to compute $\widetilde{\mathbf{R}}^*$ (7.22) was to be implemented operationally using similar randomisation techniques, then only the diagonal elements would be specified in the combined error covariance matrix. Therefore, we wish to show in our numerical experiments, with the idealized erroneous model (7.42), that even when only the diagonal elements of the combined error covariance matrix are calculated and used within the data assimilation process, improvements to the analysis accuracy can be obtained. Specifically, this ignores the presence of time, spatial and multivariate cross correlations in both the observation error and model error.

7.5.1 Combined model error and observation error statistics

The 4DVar cost function (2.9) is formulated upon the basis that there is no error present in the model. Our experiments involve using the erroneous model (7.42) in this 4DVar cost function (2.9). We define three methods in Table 7.3, that vary depending on the error covariance matrix used in the $\mathcal{J}_{ob}$ term of the 4DVar cost function (2.9). These are next summarised as follows.

- Method 1: aims to minimise the 4DVar cost function (2.9) with no changes and therefore does not account for model error.

- Method 2: replaces $\hat{\mathbf{R}}$ in the 4DVar cost function (2.9) with the diagonal entries of the combined error covariance matrix $\hat{\mathbf{R}}^*$ (7.11).

- Method 3: replaces $\hat{\mathbf{R}}$ in the 4DVar cost function (2.9) with the diagonal entries of the estimated matrix $\widetilde{\mathbf{R}}^*$ (7.22) from sample innovation data.

Figure 7.11 compares the diagonal entries (variances) of the three matrices $\hat{\mathbf{R}}$, $\hat{\mathbf{R}}^*$ and $\widetilde{\mathbf{R}}^*$ at the observation times t_i ($i = 10, 20, 30, 40, 50$). The observation error covariance matrix $\mathbf{R}_i$ is static throughout the window, whereas the variances in the combined error covariance matrix $\hat{\mathbf{R}}^*$ computed directly with the formulae (7.11) in Method 2 are significantly larger as time increases, accounting for the uncertainty in the model state trajectories caused by the errors in the model. Note that the variances estimated in the combined model error and observation error covariance matrix $\widetilde{\mathbf{R}}^*$ with Method 3 do not always increase with time, as can be seen for the atmospheric variables x, y and z in Figure 7.11. These variances have been estimated using the nonlinear innovations in the diagnostic present in equation (7.22). Whether $\hat{\mathbf{R}}^*$ (Method 2) or $\widetilde{\mathbf{R}}^*$ (Method 3) contain the most appropriate combined model error and observation error statistics to use in the 4DVar cost function is an area of future

172

work. One hypothesis that should be tested is that the the estimate $\widetilde{\mathbf{R}}^*$ uses the innovation diagnostics and therefore better recognises the spread of the model error at each specific observation time. However, the estimate $\widetilde{\mathbf{R}}^*_{(i,i)}$ uses the linerization of the nonlinear model around the background model state trajectory (as opposed to the true model trajectory) and this could have a detrimental effect on the estimation of the combined statistics. Another key point to note is that the variances evaluated in $\mathbf{R}^*_{(i,i)}$ and estimated in $\widetilde{\mathbf{R}}^*_{(i,i)}$ are much closer in value for the ocean variables, than the atmospheric variables. The reason for this is that the theory for the derivation of $\mathbf{R}^*_{(i,i)}$ (7.11) and estimation of $\widetilde{\mathbf{R}}^*_{(i,i)}$ (7.22) is based on the use of linear models and the ocean variables w and v are of a less chaotic nature than the atmospheric variables x, y and z, that are of a highly nonlinear chaotic nature. In the next section we perform strong constraint 4DVar not accounting for the error in the model (Method 1) and compare the analysis outputs with strong constraint 4DVar performed when accounting for error in the model (Method 2 and Method 3).
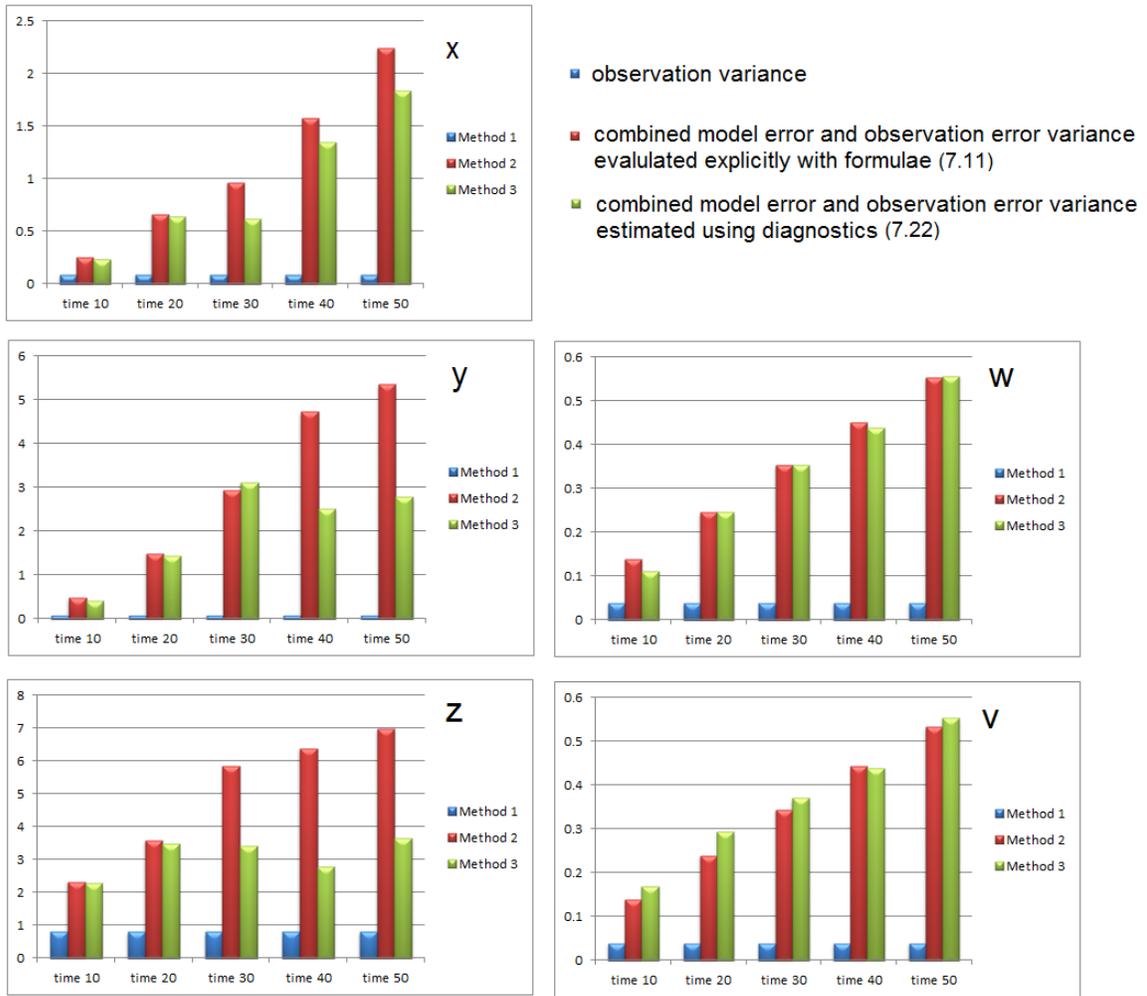
Figure 7.11: Diagonal elements of $\hat{\mathbf{R}}^*$ as described in Method 2 (red) and $\widetilde{\mathbf{R}}^*$ as described in Method 3 (green), compared with the observation error covariance matrix $\mathbf{R}_i$ as in Method 1 (blue), for each of the variables in the idealized coupled nonlinear model, at all observation time-steps t_{10}, t_{20}, t_{30}, t_{40}, t_{50}. Method 1, Method 2 and Method 3 are described in Table 7.3.

Method	Error covariance matrix in $\mathcal{J}_{ob}$	Details
Method 1	$\hat{\mathbf{R}}$	Use of unamended observation error covariance matrices $\mathbf{R}_i$ in the 4DVar cost function (2.9).
Method 2	$\hat{\mathbf{R}}^*$	Replacement of $\mathbf{R}_i$ with diagonal elements of $\hat{\mathbf{R}}^*_i$ in the 4DVar cost function (2.9). The elements of $\hat{\mathbf{R}}^*_i$ are calculated directly with equation (7.11), which requires linearization of the nonlinear model (7.43) around the true model state trajectory.
Method 3	$\widetilde{\mathbf{R}}^*$	Replacement of $\mathbf{R}_i$ with diagonal entries of $\widetilde{\mathbf{R}}^*_i$ in the 4DVar cost function (2.9). The diagonal entries of $\widetilde{\mathbf{R}}^*_i$ are estimated with equation (7.22), which requires linearization of the nonlinear model (7.42), this is conducted around the background model state trajectory. The estimation of the entries in $\widetilde{\mathbf{R}}^*_i$ are evaluated using a sample size of 1000 innovations at each observation time t_i (i=10,20,30,40,50).

Table 7.3: Three strong constraint 4DVar methods used with the erroneous idealized coupled atmosphere-ocean model.

7.5.2 Data assimilation results

We perform data assimilation with Method 1, Method 2 and Method 3 (see Table 7.3) and compare the analysis and subsequent analysis trajectories by computing the RMSE (Root Mean Square Error) from a sample of analysis outputs. We conduct the following steps 100 times to produce a sample size of 100 analysis values for each of Method 1, Method 2 and Method 3 respectively:

1. Produce a background vector $\mathbf{x}^b$ by adding noise, consistent with the statistics prescribed in $\mathbf{B}$, to the true initial state $\mathbf{x}^t_0$.

2. Produce a random model error vector $\boldsymbol{\eta}_i$ for each time $t_1, ..., t_{50}$ using the statistics specified in the model error covariance matrix $\mathbf{Q}_i$ (i=1,...,50).

3. Produce a vector of observations $\mathbf{y}_i$ at each observation time $t_{10}, t_{20}, t_{30}, t_{40}$ and t_{50} by evaluating the true model state vector $\mathbf{x}^t_i$ with equation (7.43) and adding noise in proportion to the error statistics specified in $\mathbf{R}_i$ (i=10,20,30,40,50).

4. Compute the three analysis outputs as described by Method 1, Method 2 and Method 3 in Table 7.3 respectively.

The results show a significant reduction in analysis error for all model state variables when combined model error and observation error statistics are used in the strong constraint 4DVar cost function, compared with observation error statistics only, as shown in Figure 7.12 (top). When compared to the observation error variances (used in Method 1), both the error variances in $\hat{\mathbf{R}}^*$ (Method 2) and the estimated error variances in $\widetilde{\mathbf{R}}^*$ (Method 3) give less weight to the comparison of the model evolved state with observations, due to the uncertainty in the model trajectory. The reduction in analysis RMSE, when accounting for the model error, is largest for the atmospheric variables y and z. We hypothesise that this is down to two factors: the first of which is ratio between the background error and observation error variances and the second of which is the size of the model error. We will investigate this further in the upcoming experiments. Accounting for errors in the model with the combined error statistics reduces the number of minimisation iterations required to reach the tolerance level. This is demonstrated in this case with the number of minimisation iterations reducing from 12 for Method 1, to 10 for both Method 2 and Method 3 respectively. This is because the increase in variances in the 4DVar cost function, leads to the erroneous model trajectory no longer being so tightly constrained to the observations and therefore a solution is more easily found.

An increase in analysis accuracy does not necessarily lead to an increase in forecast accuracy, as demonstrated in Figure 7.12 (bottom) which shows the RMSE of the analysis trajectories throughout the assimilation window. A decrease in forecast RMSE is only obtained for atmospheric variable z. When we include the model error statistics in the combined error covariance matrix, we are increasing the values of the variances from the observation error variances. Therefore, we are letting the

analysis model trajectory depart further from the observations and this can enable the analysis trajectory to differ further from the true state trajectory.
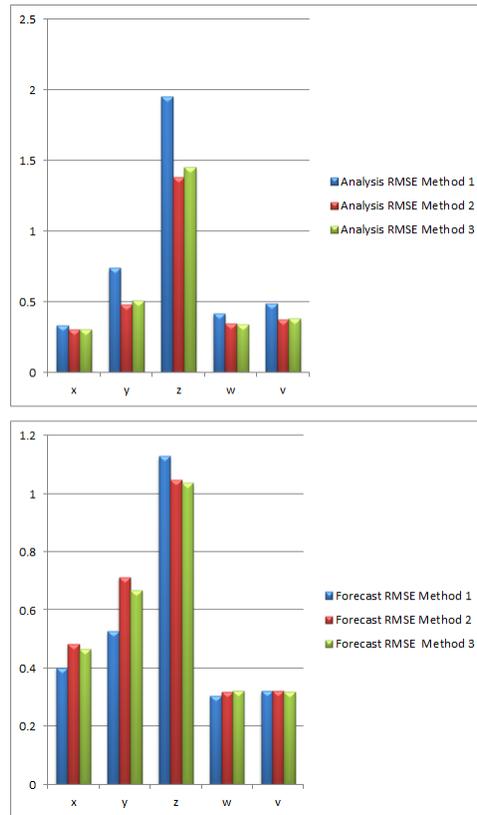
Figure 7.12: Analysis RMSE and the subsequent RMSE of the analysis trajectories over the assimilation window. Results are from applying Method 1, Method 2 and Method 3 (as described in Table 7.3) over a sample of 100 data assimilation runs.

Work with an erroneous scalar model in Section 7.3 showed that there was the most improvement to the analysis accuracy, when replacing observation error statistics in the cost function with combined error statistics, in the presence of: large model error variance, and a large background error variance in comparison with the observation error variance. Next we investigate whether these characteristic features hold with use of the erroneous nonlinear model (7.42) in strong constraint 4DVar. Work from

Conditions	Covariance matrix	Variance for x	Variance for y	Variance for z	Variance for w	Variance for v
A	$\mathbf{Q}_i$	0.02	0.02	0.2	0.01	0.01
A	$\mathbf{B}$	0.1	0.3	3.4	1.1	0.52
A	$\mathbf{R}_i$	0.09	0.09	0.81	0.04	0.04
B	$\mathbf{Q}_i$	0.005	0.005	0.05	0.0025	0.0025
C	$\mathbf{Q}_i$	0.08	0.08	0.8	0.04	0.04
D	$\mathbf{Q}_i$	0.5	0.5	5	0.25	0.25
E	$\mathbf{B}$	0.004	0.012	0.136	0.044	0.0208
E	$\mathbf{R}_i$	2.25	2.25	20.25	1	1

Table 7.4: List of error covariance matrices used in strong constraint 4DVar with the erroneous idealized coupled atmosphere-ocean model. Condition A is as described at the start of Section 7.5. When error covariance matrices are not explicitly stated for Conditions in this table, they are as defined by Conditions A.

now on compares the resulting analysis accuracy when firstly using Method 1 (with $\mathbf{R}_i$ in the 4DVar cost function) with that from Method 3 (replacement of $\mathbf{R}_i$ with diagonal entries of $\widetilde{\mathbf{R}}^*$). This is because we make the assumption that we do not know the model error covariance matrices explicitly, which is realistic of operational models. We define conditions A to be the error covariance matrices as described at the start of this section and restate these in Table 7.4.

We investigate the effect of increasing and decreasing the size of the model error on the analysis accuracy. We previously defined $\mathbf{Q}_i$ at each time t_i using conditions A in Table 7.4. We reduce the standard deviations of the model error for each of the model state variables by a factor of two and label these variances B in Table 7.4. Whereas in condition C in Table 7.4 we increase the model error standard deviations by a factor of two. Conditions D in Table 7.4 are the initial model error standard deviations increased by a factor of five. For each of the conditions B, C and D we recalculate the diagonal entries in $\widetilde{\mathbf{R}}_i^*$ at observation times ($i = 10, 20, 30, 40, 50$) with a sample size of 1000 innovation vectors and subsequently conduct 100 strong constraint 4DVar runs. Each individual data assimilation run uses independent model error vectors, vectors of observations and background vectors, and minimises the 4DVar cost function using

both Method 1 and Method 3. This produces 100 analysis values for both Method 1 and Method 3. From these samples of analysis outputs both the analysis RMSE and forecast RMSE (in the assimilation window) are evaluated for both Method 1 and Method 3 with results shown in Figure 7.13. For all specifications of the model error variance covariance $\mathbf{Q}_i$, we achieve a lower analysis RMSE accounting for the model error (Method 3), as opposed to when not (Method 1). Comparing the levels of model error within the data assimilation process, where B is lowest and is the D highest, we demonstrate that the larger the model error, the more significant the increase in accuracy is when accounting for the model error (Method 3), as opposed to when not (Method 1). The number of iterations that the minimisation algorithm performs is lower when accounting for model error, as opposed to not accounting for model error, as can be seen in Figure 7.13. We also note, similarly to the previous experiment, an increase in analysis accuracy does not necessarily lead to an increase in forecast accuracy. In the results shown in Figure 7.13 there is a degradation to the forecast RMSE (calculated throughout the assimilation window), when accounting for model error, for all variables excluding z (conditions B and C). With conditions D, there was a degraded forecast for all model state variables when accounting for model error.

The last of the experiments was conducted with very large model error variances (conditions D in Table 7.4). The decrease in analysis RMSE, when using the estimated combined error variances in Method 3, is very significant, see Figure 7.13. We re-emphasise that this remarkable improvement to the analysis accuracy, accounting for the model error has been made with the diagonal entries of the estimated matrix $\widetilde{\mathbf{R}}^*$ (7.22) replacing the observation error variances, was without the requirement of explicitly specifying the model error statistics in $\mathbf{Q}_i$ (although in our 'toy' experiments we sample the model error vectors from $\mathbf{Q}_i$). We also have disregarded any cross covariance information in $\widetilde{\mathbf{R}}^*$ as we are only using the diagonal entries in these experiments. The impact of including this information in $\widetilde{\mathbf{R}}^*$ could produce even

179

further improvement to the analysis accuracy and is an area of further work.
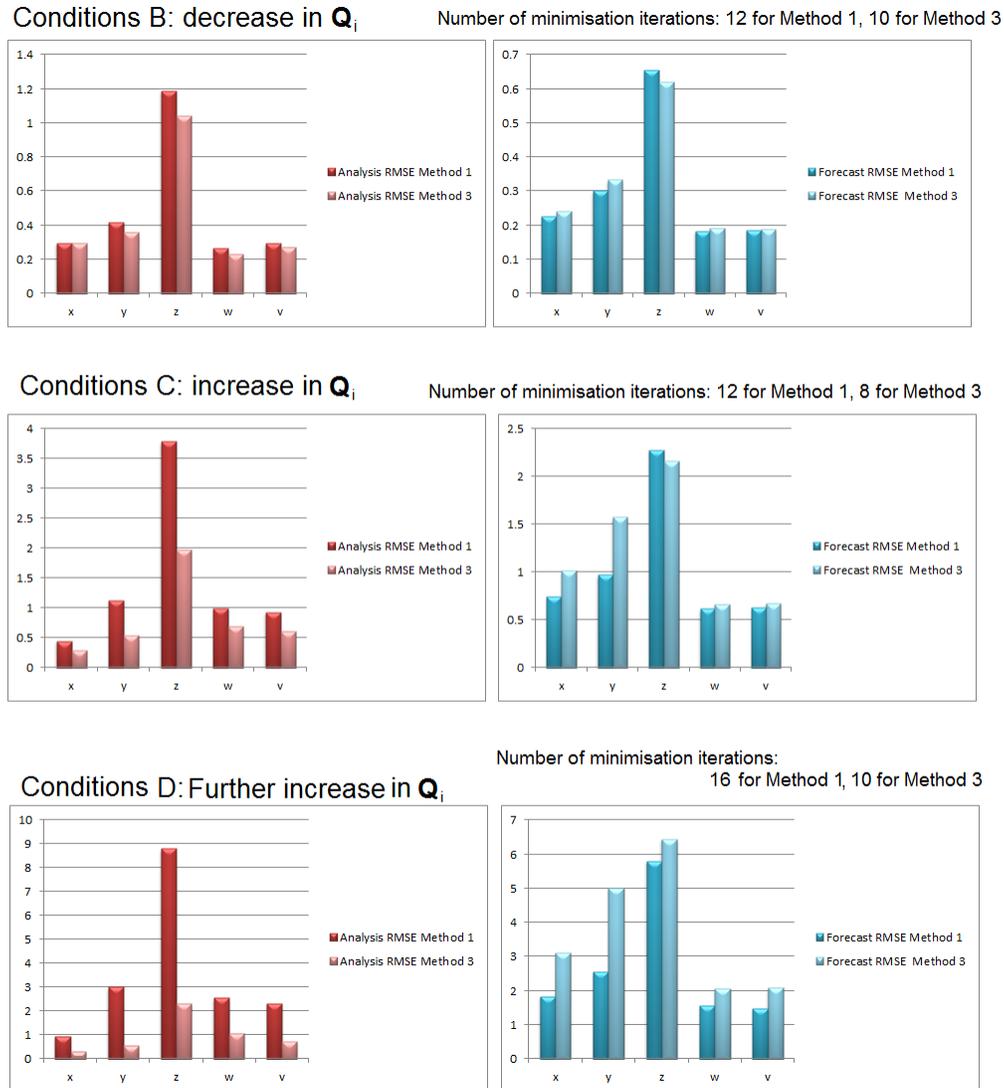
Figure 7.13: Analysis RMSE and the subsequent RMSE of the analysis trajectories over the assimilation window (Method 1 and Method 3) over a sample of 100 data assimilation runs. The data assimilation conditions for B, C and D are defined in Table 7.4.

The ratio between the accuracy of the background error and observation error is also important when investigating the difference between the accuracy of the analysis obtained when accounting for model error, as opposed to not. We demonstrate this

property by both increasing the accuracy of the background model state (by reducing the standard deviations in $\mathbf{B}$ by a factor of five) and decreasing the accuracy of the observations (by increasing the standard deviations in $\mathbf{R}_i$ by a factor of five), see conditions E in Table 7.4. The observations are now far less accurate than the background. We set $\mathbf{Q}_i$ back to conditions A as stated in Table 7.4. We recalculate the diagonal entries of $\widetilde{\mathbf{R}}_i^*$, with the error covariance matrices as specified in conditions E using a sample size of 1000. We subsequently conduct 100 strong constraint 4DVar runs for both Method 1 and Method 3. Each individual data assimilation run uses independent model error vectors, vectors of observations and background vectors. This produces 100 analysis values for both Method 1 and Method 3. Although accounting for the error in the model produces an analysis of higher statistical accuracy for all model state variables, the significance of this increase is reduced when the ratio $r = \frac{\sigma_b{}^2}{\sigma_{ob}{}^2}$ is significantly reduced for each of the model state variables, as shown in Figure 7.14. For case A, the ratio $r = \frac{\sigma_b{}^2}{\sigma_{ob}{}^2}$ for each of the model state variables x, y, z, w and v is 1.11, 3.33, 4.20, 27.5 and 13 respectively. Whereas for case E the ratios r are far smaller at 0.002, 0.005, 0.007, 0.044 and 0.0208 for each of the respective model state variables x, y, z, w and v.

For a general 4DVar problem using a model with random error present at each time-step, the larger the ratio $r = \frac{\sigma_b{}^2}{\sigma_{ob}{}^2}$ for each of the model state variables, the more weight is given to the comparison of the observations and erroneous model evolved state. This leads to a more significant difference in analysis accuracy, when accounting for model error with the combined error statistics, as opposed to not.

181
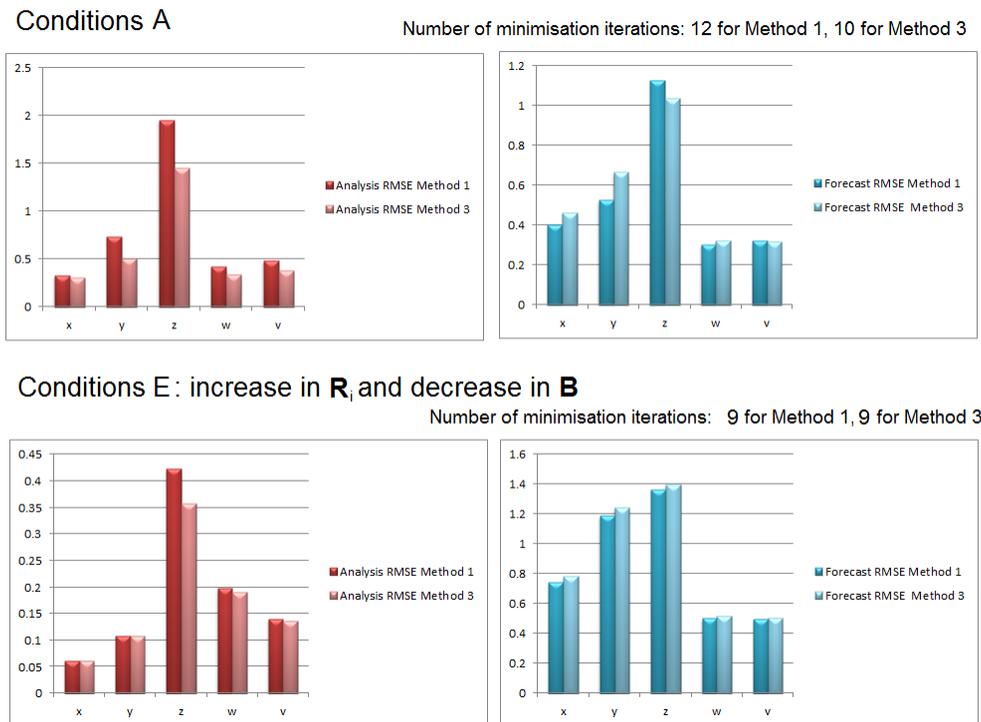
Figure 7.14: Analysis RMSE and the subsequent RMSE of the analysis trajectories over the assimilation window (Method 1 and Method 3) over a sample of 100 data assimilation runs. The data assimilation conditions for A and E are defined in Table 7.4.

7.6 Summary

In this chapter we have developed a combined model error and observation error covariance matrix. When a model with random error at each time-step is used within strong constraint 4DVar, replacing the observation error covariance matrix with this combined error covariance matrix, produces a statistically better estimate of the initial state. We have noted that this does not necessarily lead to improvements in the forecast accuracy. We have shown that when this replacement is conducted the strong constraint 4DVar diagnostics are upheld. We have developed a method to estimate this combined model error and observation error matrix without the need for the ex-

plicit specification of the model error statistics. The theory for both the development and estimation of the combined model error and observation error covariance matrix was derived with use of erroneous linear model matrices. However, we have demonstrated the successful application of our developed methods with use of an erroneous nonlinear model. In the next chapter we change our focus from improving analysis accuracy, to improving coupled model forecasts.

Chapter 8

Improving an erroneous coupled model forecast

In Chapter 3 we discussed the use of coupled atmosphere-ocean models at operational NWP centres to produce seasonal to interannual forecasts. These coupled models do not perfectly describe the true dynamics of the coupled system as they contain model errors. Bulk formulae used to represent fluxes at the atmosphere-ocean interface require the specification of coupling parameters. The specification of these coupling parameter values are often tuned as they do not have a physical meaning [55]. It has previously been shown that the effect model bias has on a simple coupled forecast may be reduced through estimation of both atmospheric and oceanic parameter values along with the coupled model state using an ensemble Kalman Filter method [130] [131], even if the source of the model bias does not originate from the parameters. In Section 8.1 we introduce general nonlinear perfect coupled model equations and subsequently general nonlinear erroneous coupled model equations. In Section 8.2 we demonstrate the effect model bias can have on a coupled model forecast with use of a

biased idealized coupled atmosphere-ocean model. In Section 8.3, we extend the idea
of coupled model state estimation, to simultaneously estimating coupling parameters
along with the coupled model state in a 4DVar context, with the objective that
the coupling parameter estimation can mitigate the effect of the model errors on
the forecast. In Section 8.4 we demonstrate how coupling parameter estimation can
compensate for model bias present in an idealized coupled atmosphere-ocean model
and hence improve the accuracy of the coupled forecast. In Section 8.5 we demonstrate
that the developed method is not successfully applicable to compensate for error of
a random nature at each time-step. However, in Section 8.6 we demonstrate how
coupling parameter estimation can compensate for another type of systematic error
in the coupled model, namely static parameter errors present in an idealized coupled
atmosphere-ocean model and hence improve the accuracy of the coupled forecast.

8.1 Erroneous coupled atmosphere-ocean model

In Section 3.1.2 we discussed the presence of errors in operational NWP coupled
atmosphere-ocean models and the difficulties in eliminating these errors from the
model equations. Therefore, we seek to develop data assimilation methods to mitigate
the effect these errors have on coupled model forecasts. We consider a perfect coupled
model of the form,

$$\mathbf{x}^t_i = \mathcal{M}_{\{i-1\}\rightarrow i}(\mathbf{x}^t_{i-1}, \boldsymbol{\alpha}^t) \qquad i = 1, 2, ... \tag{8.1}$$

where the model operator $\mathcal{M}_{\{i-1\}\rightarrow i} : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ describes perfectly the evolution of
the state vector from time t_{i-1} to time t_i. The true values of atmosphere and ocean
variables, at time t_i, are represented in the true coupled model state vector $\mathbf{x}^t_i \in \mathbb{R}^m$.

The model's optimal scalar coupling parameters are represented in the vector $\boldsymbol{\alpha}^t \in \mathbb{R}^l$ and are assumed to be time invariant. We next consider the situation where the best known description of the true coupled dynamics is erroneous and of the form,

$$\mathbf{x}_i = \mathcal{M}^e_{\{i-1\}\rightarrow i}(\mathbf{x}_{i-1}, \boldsymbol{\alpha}^t) \tag{8.2}$$

$$= \mathcal{M}_{\{i-1\}\rightarrow i}(\mathbf{x}_{i-1}, \boldsymbol{\alpha}^t) + \boldsymbol{\eta}_i \qquad i = 1, 2, ... \tag{8.3}$$

where the model operator $\mathcal{M}^e_{\{i-1\}\rightarrow i} : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ erroneously describes the coupled model dynamics. The erroneous model consists of the summation of the true model dynamics from time t_{i-1} to time t_i and a vector of model error $\boldsymbol{\eta}_i \in \mathbb{R}^m$ at time t_i. We note that here (8.2) we assume the 'optimal' values of the coupling parameters $\boldsymbol{\alpha}^t$ are known. In numerical experiments later in this chapter we remove this assumption. Previously, we have added a vector of model error $\boldsymbol{\eta}_i$ to the erroneous model equations, see (2.24) and (7.1), however here we add the vector of model error to the true model equations (8.3). The inconsistencies in this notation is purely to present workings in each respective chapter in the most clear manor. Note that above where the additive model error is present in the equation for the true model, this can be rearranged so that it is present in the erroneous model with the only difference being that the model error will be subtracted.

For numerical experiments in this chapter we use the idealized coupled atmosphere-ocean model [87], discretized in time with time-step $\Delta t = 0.01$, as described in Section 5.2. The model state vector $\mathbf{x}_i$ consists of the atmospheric state variables x, y and z and oceanic state variables w and v at time t_i. There is one scalar coupling parameter α present in this idealized coupled model, which can be interpreted as describing the heat flux at the atmosphere-ocean interface [42]. In Section 5.2.3 we demonstrated that the specification of the coupling parameter α can have a significant impact on

186

the behaviour of both the atmospheric and oceanic model state trajectories over time.

8.2 Biased coupled atmosphere-ocean model forecast

We let the discretized idealized coupled atmosphere-ocean model, as described in Section 5.2, with $\alpha^t = 1$ be our 'true' reference model (8.1). We consider the situation where the model error vector in equation (8.3) is constant model bias at each time t_i and hence simplify the notation $\boldsymbol{\eta}_i = \boldsymbol{\eta}$. The 'true' model is spun up for 10^4 time steps starting from $(x, y, z, w, v)^T = (1, 1, 1, 1, 1)^T$ to give true initial conditions $(x^t{}_0, y^t{}_0, z^t{}_0, w^t{}_0, v^t{}_0)^T = (-3.4866, -5.7699, 18.3410, -10.7175, -7.1902)^T$ on the model attractor. We introduce bias into the model equations (8.3) by denoting the vector of constant bias $\boldsymbol{\eta} = (5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2})^T$. This high level of bias in the model creates model state trajectories that follow the true general behaviour at the start of the window (approximately up to 50 time-steps), however this level of bias leads to significant differences between the 'true' and 'biased' model trajectories over time. This behaviour is shown in Figure 8.1, where the 'true' model state trajectories (black) and the 'biased' model state trajectories (pink) are shown for each of the atmosphere and ocean variables over 750 time-steps starting from the same true initial conditions. This demonstrates the point that even if the true model state initial conditions are known and used to initialise a coupled model, bias present in a coupled model can over time lead to significant errors in a coupled forecast. The objective of work in this chapter is to mitigate the detrimental effect that model error can have on the accuracy of a coupled model forecast.

Let us reference back to Figure 5.7 presented in Section 5.2.3 where the 'true' reference model (8.1) was run multiple times from the same initial conditions but with

varying specifications of the coupling parameter α. With just small perturbations to the coupling parameter (specifically 3% in Figure 5.7) the model state trajectories of all the atmospheric and oceanic variables deviated from the 'true' run over time. With this in mind for a generic coupled system, we aim to mitigate the effect of model bias in a coupled forecast, by allowing the model's coupling parameters to deviate from their original 'optimal' values. To do so we will formulate a 4DVar method to simultaneously estimate the coupled atmosphere and ocean model state initial conditions and coupling parameters.
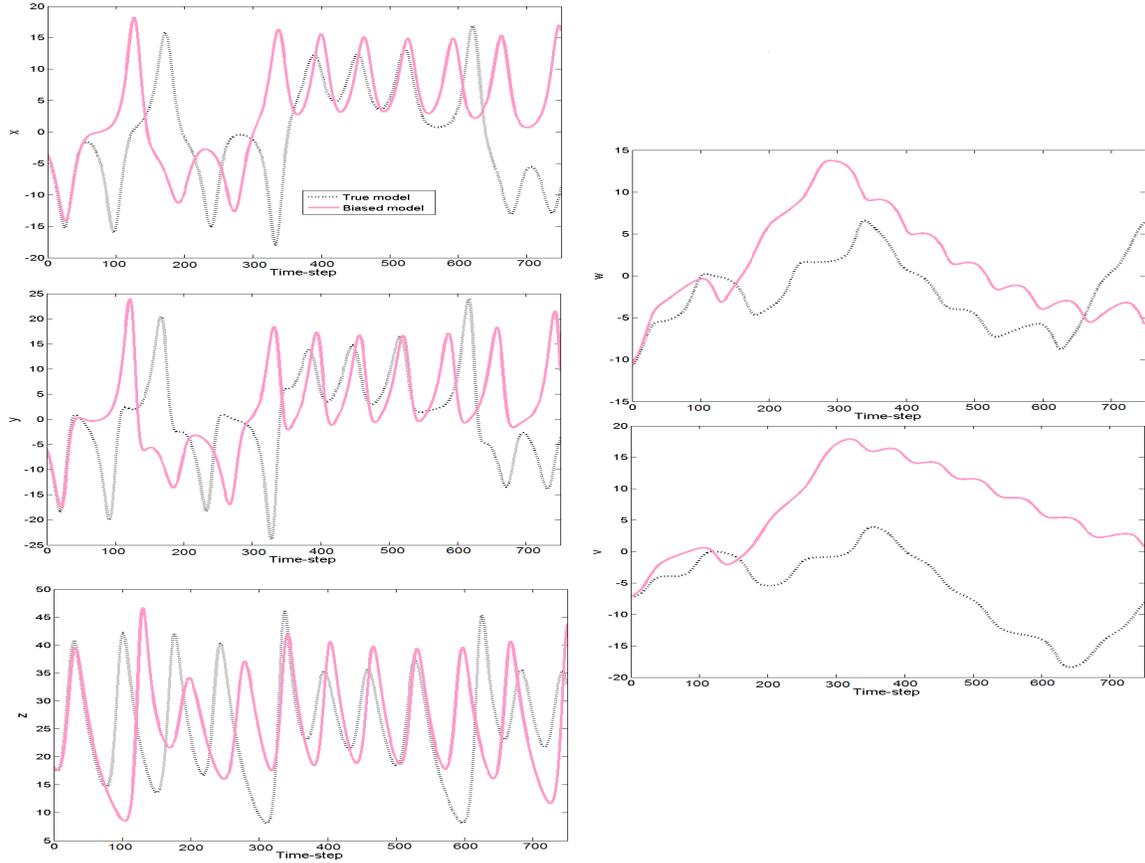
Figure 8.1: Model state trajectories produced using the 'true' idealized coupled model (black dotted lines), model state trajectories produced using the biased idealized coupled model (pink lines), shown for atmospheric variables x, y and z and ocean variables w and v. Vector of constant bias $\boldsymbol{\eta} = (5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2})^T$ present at each time-step t_i in the biased coupled model run. Both models initialised from the same true initial conditions $(x^t{}_0, y^t{}_0, z^t{}_0, w^t{}_0, v^t{}_0)^T = (-3.4866, -5.7699, 18.3410, -10.7175, -7.1902)^T$.

8.3 Joint coupled state-coupling parameter estimation

In operational coupled models the 'optimal' values of coupling parameters are often unknown, with model experts currently specifying best guesses in bulk formula for

189

fluxes at the atmosphere-ocean interface. As coupling parameters help describe fluxes at the atmosphere-ocean interface, the behaviour of the coupled system, particularly near the interface, can be sensitive to the specification of the coupling parameters, as was demonstrated with the idealized coupled model Section 5.2.3.

In Section 4.3, we outlined work previously conducted on both adjusting fluxes at the atmosphere-ocean interface and estimating model parameters, with the aim to compensate for model bias and hence improve coupled model forecasts. Operational centres are moving towards coupled data assimilation, where the initial conditions of both the atmospheric and oceanic variables are simultaneously estimated. To date, no prior work has involved using strong constraint 4DVar to simultaneously estimate coupling parameters along with a coupled model state with the aim for the coupling parameter estimation to compensate for bias present in a coupled model and hence improve the coupled forecast. We next develop a strong constraint 4DVar scheme for this purpose.

8.3.1   Coupled 4DVar with coupling parameter estimation

The vector $\boldsymbol{\alpha}^t \in \mathbb{R}^l$ present in the 'true' coupled model (8.1) contains the 'optimal' values of the scalar coupling parameters, by which we mean the values of the coupling parameters which best represent the processes that they portray. We aim to allow the values of the coupling parameters to deviate from these 'optimal' values in order to compensate for the errors in the model and hence push the model forecast trajectories closer to the 'true' model state trajectories. To do so we add a term onto the strong constraint 4DVar cost function (2.9) and we augment the 4DVar control vector to

include the model's coupling parameters along with the coupled model state as follows,

$$\mathcal{J}(\mathbf{x}_0, \boldsymbol{\alpha}) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i))$$
$$+ \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^t)^T \mathbf{F}^{-1}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^t), \tag{8.4}$$

where the difference between the coupling parameters that lead to forecast error minimisation in the assimilation window and the original 'optimal' values $\boldsymbol{\alpha}^t$ is considered to be normally distributed around a zero mean with the covariance matrix $\mathbf{F} \in \mathbb{R}^{l \times l}$. This difference term is assumed to be uncorrelated with the both the background errors and observation errors. The cost function (8.4) is to be minimised with respect to both the initial coupled state vector $\mathbf{x}_0$ and the vector of coupling parameters $\boldsymbol{\alpha}$, subject to satisfying the coupled model,

$$\mathbf{x}_i = \mathcal{M}^e{}_{\{i-1\} \to i}(\mathbf{x}_{i-1}, \boldsymbol{\alpha}). \tag{8.5}$$

Note that this erroneous model (8.5) is the same as the erroneous model specified in equation (8.2) except for the replacement of $\boldsymbol{\alpha}^t$ with $\boldsymbol{\alpha}$. It is this replacement which is vital for the performance of this scheme, as we aim for the analysis of the coupling parameters to compensate for the bias present in the model. This method requires the specification of $\mathbf{F}$, which for operational models would rely on advice from a model expert. Currently statistical information for certain atmospheric parameters is provided by model experts in order to account for the uncertainty in the assigned model parameter values from which ensemble forecasts are produced at operational centres, such as the Met Office [19].

The resulting analysis of the model state $\mathbf{x}_0^a$ from the minimisation of the cost function (8.4) is dependent on the analysis of the coupling parameters $\boldsymbol{\alpha}^a$ as they occur within the model equations. Likewise, the analysis of the coupling parameters are dependent on the analysis of the model state, as the model trajectories are dependent on the specification of the initial model state. We next investigate the compositional structure of the analysis of the coupled model state and analysis of the coupling parameters for a simple erroneous model.

8.3.2 Coupled state-coupling parameter analysis

We wish to take a simplistic situation where the erroneous nonlinear model (8.2) contains just one scalar coupling parameter α^t. We do so in order to derive an explicit analysis which we can study. We assume we have a vector of observations $\mathbf{y}$ at a single time t_1. The true coupled model dynamics from time t_0 to time t_1 are described by a model of the form (8.1) where the optimal scalar coupling parameter α^t is present in one matrix element. We minimise the amended strong constraint 4DVar cost function (8.4) with respect to both the initial coupled state vector $\mathbf{x}_0$ and the scalar coupling parameter α, subject to the coupled model (8.5). As we have just one scalar parameter α^t present in the model, the covariance matrix $\mathbf{F}$ in the cost function (8.4) is simply replaced with the variance $\sigma_\alpha{}^2$. The resulting analysis of the initial coupled model state is as follows,

$$\mathbf{x}_0^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathcal{H}\left(\mathcal{M}^\alpha(\mathbf{x}^b)\right)), \tag{8.6}$$

where the gain matrix $\mathbf{K} = \mathbf{BM}^{\alpha T}\mathbf{H}^T \left(\mathbf{R} + \mathbf{HM}^\alpha\mathbf{BM}^{\alpha T}\mathbf{H}^T\right)^{-1}$ and the corresponding analysis of the scalar coupling parameter is as follows,

$$\alpha^a \;\; = \;\; \alpha^t + \mathbf{k}^\alpha \left(\mathbf{y} - \mathcal{H}(\mathcal{M}^\alpha(\mathbf{x}^a{}_0)) \right), \tag{8.7}$$

where the gain vector $\mathbf{k}^\alpha = \sigma_\alpha{}^2 \left(\frac{\partial \mathcal{M}^e}{\partial \alpha}(\mathbf{x}^a{}_0) \right)^T \mathbf{H}^T \mathbf{R}^{-1}$. The model $\mathcal{M}^\alpha = \mathcal{M}^e{}_{0\to1}(\mathbf{x}_0, \boldsymbol{\alpha}^a)$, therefore is the erroneous model (8.2) but with the true value of the coupling parameter α^t replaced with the analysis of the coupling parameter α^a. We have assumed the tangent linear hypothesis [17] holds, for both the nonlinear observation operator $\mathcal{H}$ and nonlinear system equations $\mathcal{M}^\alpha$, as described previously in Section 2.3.1. In summary, the linear observation operator $\mathbf{H}$ is the first order term in the expansion of the Taylor series of $\mathcal{H}(\mathbf{x} + \delta\mathbf{x})$ and the tangent linear model $\mathbf{M}^\alpha$ is the first order term in the expansion of the Taylor series of $\mathcal{M}^\alpha(\mathbf{x} + \delta\mathbf{x})$. Note that in the evaluation of (8.7) we have assumed that the coupling parameter is not present in the observation operator.

We should note that the departure of the coupling parameter analysis (8.7) from the original 'optimal' value α^t is dependent on the specification of the variance $\sigma_\alpha{}^2$. We also highlight this departure is also dependent on the comparison of the observation vector with the model state analysis evolved using the biased model (and then mapped to observation space). We hypothesise that the deviation of the analysis of coupling parameters from their prior 'optimal' values can compensate for the model bias and pull the model trajectories closer to the observations and hence improve the coupled model state forecast. We next investigate the potential improvements to coupled forecasts this scheme can produce, by implementing this scheme in numerical experiments.

8.4 Numerical experiments compensating for model bias

In this section we apply the coupling parameter estimation scheme to the case when the model used in 4DVar has error of the form (8.3). In this section, the error present in the model (8.3) is that of bias. The 'true' and biased idealized coupled models are as described in Section 8.2 with the prior 'optimal' coupling parameter $\alpha^t = 1$ and the model bias vector at each time-step set to $\boldsymbol{\eta} = (5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2}, 5 \times 10^{-2})^T$. This high level of model bias creates biased model state trajectories that deviate significantly from the 'true' model state trajectories, as can be seen Figure 8.1. We consider direct observations, with linear observation operator $\mathbf{H}_i = \mathbf{I}$ at all observation times t_i, of all atmosphere and ocean variables to be available every 10 time-steps over a total assimilation window length of 250 time-steps. Later in this section we will vary the assimilation window length and investigate the impact this has on the results. The observation error covariance matrix $\mathbf{R}_i$ is diagonal at each observation time t_i in the assimilation window, initially set with variances $1.6 \times 10^{-5}, 1.6 \times 10^{-5}, 6.4 \times 10^{-5}, 10^{-6}, 10^{-6}$ along the diagonal. The standard deviations of the observation errors are relative to the amplitude of each of the model state variables. We assume we have a background model state vector $\mathbf{x}^b$ of the form (2.5) where initially we specify a diagonal background error covariance matrix $\mathbf{B}$ with the variances $4 \times 10^{-3}, 4 \times 10^{-3}, 4 \times 10^{-3}, 10^{-3}, 10^{-3}$ along the diagonal. The reasoning for the specification of the background variance for z not being larger than those for x and y (as was for the observation variances) is that the atmospheric equation for z (5.4) is the only equation that does not contain the coupling parameter. Therefore, initial experiments will investigate whether the bias in the forecast of z can be compensated implicitly through the estimation of α, where we assume a

reasonably accurate background model state is available. Later, we will investigate the effect of decreasing the accuracy of the background on our results. As there is only one coupling parameter α present in the idealized coupled model (5.4), the covariance matrix $\mathbf{F}$ in the cost function is now replaced by the error variance $\sigma_\alpha{}^2$. We set the variance of the coupling parameter to $\sigma_\alpha{}^2 = 10^{-4}$ in the cost function (8.4), to allow the coupling parameter to deviate from the true value in order to try to mitigate the effect of the model bias. Experiments in this section will produce values for the background model state and observations by taking the true model state values and adding random error consistent with the variances specified in $\mathbf{B}$ and $\mathbf{R}_i$ respectively. We will compare the analysis and subsequent forecast accuracy from two strong constraint 4DVar methods.

- Method 1: Use of the biased model (8.2) in the minimisation of the strong constraint 4DVar cost function (2.9) to best estimate the initial coupled model state only. This analysis of the coupled model state will subsequently be used to produce a coupled forecast with the biased coupled model.

- Method 2: Minimisation of the amended strong constraint 4DVar cost function (8.4) to simultaneously estimate the coupling parameter α along with the initial coupled model state. The analysis of the coupled model state and analysis of the coupling parameter will subsequently be used to initialize a coupled forecast.

We will repeat the performance of both Method 1 and Method 2 to obtain a sample of 100 data assimilation analysis outputs for both methods. On each cycle an independent background vector and independent vectors of observations will be used. Both data assimilation schemes will be performed with the same set of background and observational data on each cycle. The accuracy of the coupled forecasts from both methods will be compared through both the assimilation window and over a further forecast window length of 500 time-steps. Later in this section we will investigate the

effect the length of the forecast window has on the results. We compute the accuracy of the forecasts by calculating the RMSE (Root Mean Square Error) between the analysis model trajectories and the true model trajectories over the sample.

8.4.1 Improved forecast accuracy

With the experimental set up as just described, the resulting RMSE's in the analysis, assimilation window and forecast window are shown in plots (a), (b) and (c) respectively in Figure 8.2. Comparing the results from Method 1 and Method 2 in plots (b) and (c), estimating the coupling parameter along with the coupled model state has reduced the RMSE for each model state variable in the assimilation window and hence improved the coupled model forecast in the subsequent forecast window. The coupling parameter estimation has partially compensated for the model bias and thus improved the model forecast. In doing so the analysis of the coupling parameter averaged at 0.2486 below the prior 'optimal' value of $\alpha^t = 1$, with all coupling parameter values from the sample of experiments $\alpha^a < 1$. We remark that when simultaneously estimating the coupling parameter along with the model state (Method 2), this does not necessarily provide a more accurate model state analysis, than when not estimating the coupling parameter (Method 1). This result is demonstrated in plot (a) in Figure 8.2. Let us restate that the objective of the coupling parameter estimation is to mitigate the effect of the model bias over the assimilation window and hence provide an improved forecast.

The most significant improvement the coupling parameter estimation has made to the forecast accuracy is that to the ocean variable v , as shown in Figure 8.2. This is because the 'true' trajectory and 'biased' trajectory for v have the most significant deviations, even when the initial conditions are specified to be the same, as can be seen in Figure 8.1. Another important outcome to note is that the z equation (5.4) does

not explicitly contain the coupling parameter, but interaction with the atmospheric variables x and y which do contain the coupling parameter still enables the bias in z to be partially compensated for.

To illustrate how the developed scheme works we present the results in the assimilation window from one sample run of both Method 1 and Method 2 in Figure 8.3. In this case the analysis of the coupling parameter $\alpha^a = 0.7481$. The analysis of the coupling parameter has partially compensated for the bias in the model enabling the true model trajectories to be better represented within the assimilation window, particularly in this case for the atmospheric variables x, y and z. We present Figure 8.3 to emphasise that the data assimilation methods rely on observations over the assimilation window to represent the true behaviour of the model state variables. It is therefore the observations that are allowing the coupling parameter estimation to compensate for the bias in the model. In Section 8.4.2 we investigate the significance the success of this scheme has on accurate observations.

The forecast duration for which coupled 4DVar with coupling parameter estimation can compensate for the bias in the model is the next area of our investigation. The coupling parameter is not 'correcting' the model bias, in fact an additional error to the prior 'optimal' coupling parameter value is made in order to counteract the bias in the model. Therefore the model with the analysis of the coupling parameters present is not expected to be an exact representation of the true dynamics, but is expected to improve on the forecast subsequent to the assimilation window for a certain period of time. We now investigate the effect of extending the forecast period in our numerical experiment from 500 time-steps to 1000 time-steps on our results. The forecast RMSE's for this longer period are shown in plot (d) in Figure 8.2. The comparison we are interested in here is between the plot (c) where the forecast is of length 500 time-steps with plot (d) where the forecast is of length 1000 time-steps. With this forecast window extension, a more accurate forecast is obtained with cou-

197

pling parameter estimation (Method 2 results compared with Method 1 results) for atmospheric variable z and both ocean variables w and v, however there is a slight degradation to the atmospheric forecasts of x and y. The length of time that coupling parameter estimation can compensate for the bias and improve a coupled forecast is model dependent. We hypothesise that the less chaotic the nature of the model variables, the longer the coupled 4DVar with coupling parameter estimation scheme can improve the forecast accuracy, as opposed to coupled model state estimation only. This is an area of further work, as discussed in Section 9.2 of this thesis.
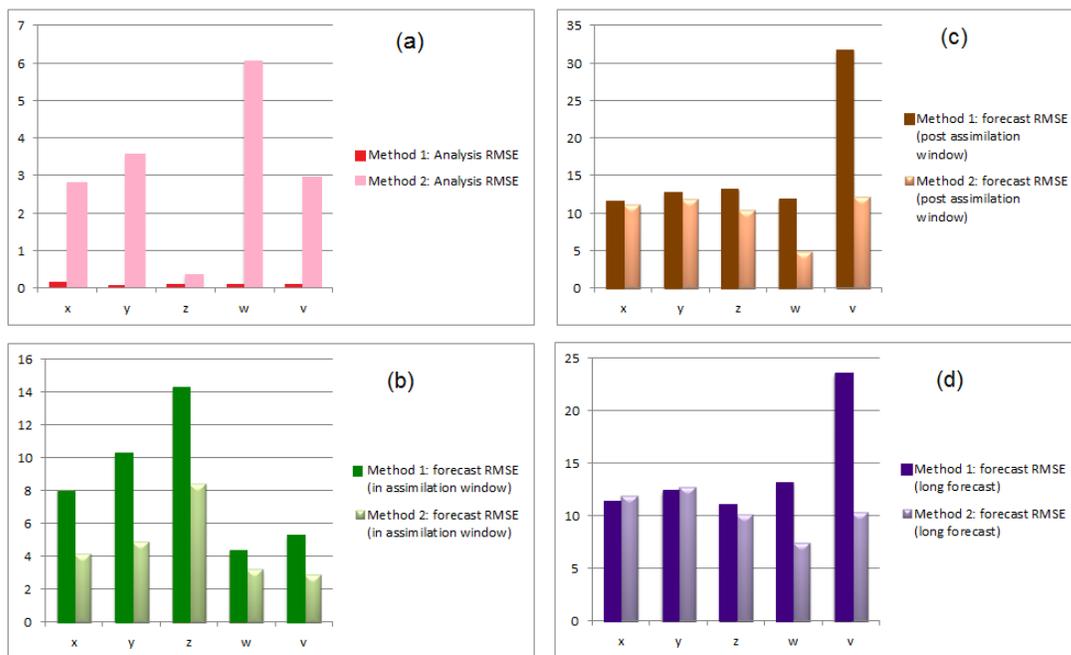
Figure 8.2:   Comparison of Method 1 and Method 2 from a sample of 100 data assimilation runs for both methods with the biased idealized coupled model. The figure shows the resulting RMSE in the: (a) analysis (b) assimilation window (250 time-steps) (c) forecast window (500 time-steps) (d) forecast window (1000 time-steps). The analysis of the coupling parameter α^a resulted in an average of 0.2486 below $\alpha^t = 1$.
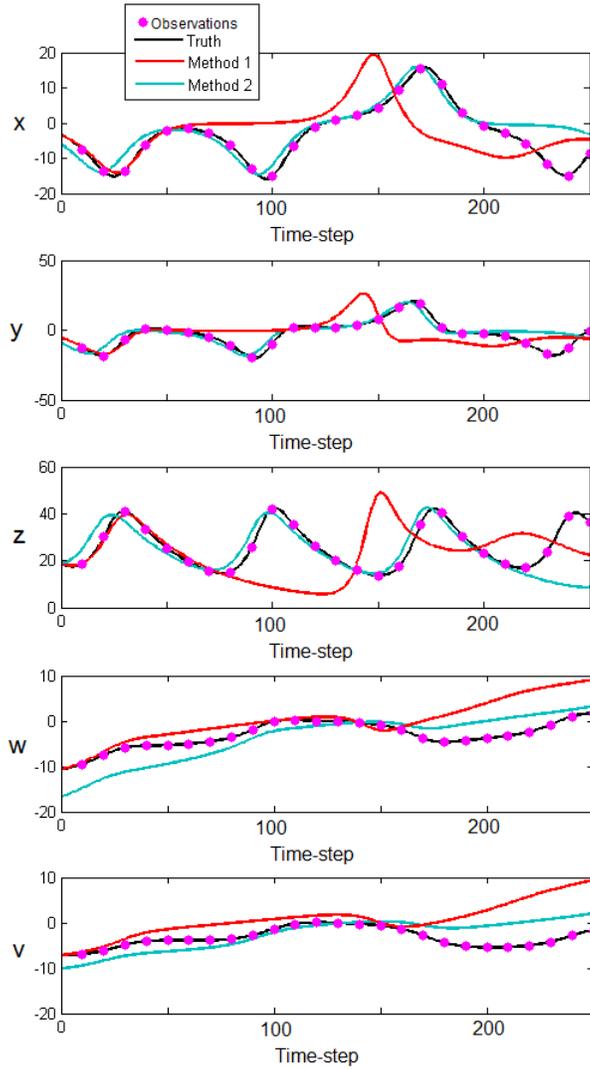
Figure 8.3: Atmosphere and ocean variables in the assimilation window from one sample run of both Method 1: Strong constraint 4DVar used to estimate the coupled model state initial conditions only (red), Method 2: simultaneous estimation of the coupled model state and coupling parameter α (turquoise), both schemes using the biased idealized coupled model. Method 2 analysis of coupling parameter $\alpha^a = 0.7481$.

8.4.2 Reliance on observations

The observations throughout the assimilation window in the previous experiment were of high accuracy and hence represented the general behaviour of the true model trajectories, as can be seen in Figure 8.3. We demonstrate that when there is a significant degradation in observation accuracy, that coupling parameter estimation can no longer successfully compensate for bias present in a coupled model to produce an improved coupled forecast. We illustrate this point by using the same experimental conditions as specified in Section 8.4, except we significantly increase the observation error standard deviations by a factor of 10, so the observations are no longer as tightly constrained to the true model state trajectories. We conduct strong constraint 4DVar to best estimate the initial coupled model state (Method 1) and also conduct strong constraint 4DVar with simultaneous estimation of the coupling parameter along with the initial coupled model state (Method 2). The resulting RMSE's (from a sample size of 100) throughout the assimilation and forecast windows are shown in Table 8.1 and Table 8.2 respectively for both data assimilation schemes. Comparing the results in Table 8.1 and Table 8.2, it is only the atmospheric variables x and y that have a slight reduction in RMSE throughout the assimilation window when simultaneously estimating the coupling parameter. It is only the oceanic variable w that has a reduction in forecast RMSE when simultaneously estimating the coupling parameter. However, what should be noted is that this reduction in observation accuracy has put more weight on the background term in the cost function and the estimation of the coupling parameter has lead to an increase in the analysis accuracy for each of the model state variables.

When we compare the analysis RMSE when using coupled 4DVar with coupling parameter estimation (Method 2) in Table 8.2 with those in Figure 8.2 (pink bars), we observe that a significant degradation in observational accuracy has significantly

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	0.5871	5.6339	11.2629
y	0.7848	7.1913	12.3034
z	0.6354	12.2803	10.7122
w	2.8783	3.6373	11.6071
v	1.5192	4.3968	29.4670

Table 8.1: Reduced observational accuracy. Strong constraint 4DVar used to estimate the coupled model state initial conditions only (Method 1), using the biased idealized coupled model. Average number of minimisation iterations: 53.

increased the statistical accuracy of the analysis. This is because the results shown in Figure 8.2 have originated from the combination of a very high observational accuracy with a high level of model bias. The combination of the initial model state analysis and coupling parameter analysis that optimally minimises the cost function (8.4) is that of an inaccurate initial model state. However, when the observational accuracy is significantly reduced, this allows the analysis trajectories to deviate further from the 'true' trajectories. In this case, less weight is put on the observational term in the cost function and more weight is given to the background model state and background coupling parameter terms. This leads to an improvement in statistical accuracy of the analysis of the initial model state, as shown in Table 8.2.

In general for the estimation of coupling parameters to successfully compensate for the bias in a model, the scheme needs to be provided with observations that describe the behaviour of the true model trajectories to a reasonable degree of accuracy. The inability of the coupling parameter estimation to compensate for the bias in the assimilation window, leads to an inability of the accuracy of the coupled forecast to be significantly improved in the forecast window.

Operationally, observations of the ocean are available less frequently in time than observations of the atmosphere. The ocean is a slow moving system, so as long

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	0.0989	5.5055	11.6348
y	0.0622	6.9893	12.9469
z	0.2044	12.3463	11.5105
w	0.7552	3.6637	11.3132
v	0.0464	4.5058	31.5242

Table 8.2: Reduced observational accuracy. Strong constraint 4DVar with coupling parameter estimation (Method 2) using the biased idealized coupled model. Average number of minimisation iterations: 42. The analysis of the coupling parameter α^a resulted in an average of 0.2148 below $\alpha^t = 1$.

as there are observations at a sufficient number of points in time throughout an assimilation window, the general behaviour of the true ocean model state variables should still be reasonably represented. This is related to the idea of 'observability', as discussed by Navon [90]. With less frequent ocean observations (than observations of the atmosphere) the coupled 4DVAR with coupling parameter estimation scheme should still be able to compensate for bias in a coupled model. We demonstrate this with a numerical experiment where we use the experimental set up as described at the start of the Section 8.4 except for a reduction in the ocean observation frequency (to every 20 time-steps as opposed to 10 time-steps). The resulting analysis and forecast RMSE's from a sample of runs with both Method 1 and Method 2 are shown in Table 8.3 and Table 8.4 for Method 1 and Method 2 respectively. Estimation of the coupling parameter (Method 2) partially compensates for the bias in the model and leads to a reduction in the RMSE's for all of the model state variables in the assimilation window, when compared to coupled model state estimation only (Method 1). This leads to an increase in accuracy in the subsequent forecast for all model state variables. However, what should be noted is that the reduction in the frequency of ocean observations has reduced the extent to which the coupling parameter can compensate for the model bias. This is because the true ocean variables are now

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	0.0906	6.7084	11.4883
y	0.0905	8.7133	12.5332
z	0.2970	13.1567	11.8780
w	1.0573	4.0636	11.8999
v	0.5378	5.0298	30.8169

Table 8.3: Reduction in ocean observations (from every 10 time-steps to every 20 time-steps). Strong constraint 4DVar used to estimate the coupled model state initial conditions only (Method 1), using the biased idealized coupled model. Average number of minimisation iterations: 12.

represented at fewer points in time by the observations and in our experimental set up we have a high level of model bias in each of the model equations, as described in Section 8.2. In this experiment the analysis of the coupling parameter α^a resulted in an average of 0.1271 below $\alpha^t = 1$, however with more frequent ocean observations a more accurate forecast was obtained with the analysis of the coupling parameter α^a being an average of 0.2486 below $\alpha^t = 1$, see Figure 8.2. To summarise for the scheme to work most optimally, it requires the general behaviour of the true model dynamics to be well represented over time to enable the coupling parameter estimation to push the variables trajectories away from the biased trajectories and towards the truth. Therefore, the less frequently the observations occur, the less the coupling parameter estimation scheme can compensate for the bias and hence improve the coupled forecast skill.

Both in this experiment and the previous experiments in this section, the bias in the atmospheric variable z was compensated for implicitly through the estimation of the coupling parameter α in the model equations for the atmospheric variables x and y that are present in the model equation for z, see (5.4). We next investigate how reliant the success of our developed method is on the specification of an accurate background model state for variables that do not explicitly include the coupling parameter.

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	0.0620	6.4960	11.1985
y	0.0651	8.4198	12.2770
z	0.0934	12.9402	11.4234
w	0.0946	3.9482	10.9192
v	0.0488	4.8082	30.0434

Table 8.4: Reduction in ocean observations (from every 10 time-steps to every 20 time-steps). Coupled 4DVar with coupling parameter estimation (Method 2) using the biased idealized coupled model. Average number of minimisation iterations: 12. The analysis of the coupling parameter α^a resulted in an average of 0.1271 below $\alpha^t = 1$.

8.4.3 Background model state accuracy

Next we look at changes to the background $\mathbf{x}^b$. Reduction in background model state accuracy leads to more weight being given to the observation term in the cost function. When estimating the coupling parameter simultaneously along with the coupled model state, this leads to more flexibility in the estimation of the coupled model state analysis which can lead to an analysis of the coupled state and coupling parameter that gives trajectories that fit closer to the vicinity of the observations and hence a more accurate forecast. However, we need to be cautious, as a decrease in background model state accuracy does not always lead to a statistically more accurate coupled forecast when using coupled 4DVar with coupling parameter estimation, as opposed to coupled model state estimation only. We demonstrate an exception to this property now with the experimental set up as described at the start of Section 8.4, but increase the background error standard deviation for the atmospheric variable z by a factor of two. Specifically we specify a diagonal background error covariance matrix $\mathbf{B}$ with the variances $4 \times 10^{-3}, 4 \times 10^{-3}, 1.6 \times 10^{-2}, 10^{-3}, 10^{-3}$ along the diagonal. The resulting analysis RMSE's and forecast RMSE's (over a forecast window length of 500 time-steps) from a sample of data assimilation runs are shown in Table 8.5

204

and Table 8.6 for Method 1 and Method 2 respectively. The coupled 4DVar with coupling parameter scheme (Method 2) still produces an overall statistically more accurate forecast than just estimating the coupled model state (Method 1), however the increase is less significant and there is in fact a slight degradation in forecast accuracy for the atmospheric variable x and little increase for the atmospheric variable y. The analysis RMSE increased significantly for Method 1 and the atmospheric variable z in Method 2 when the background error variance for z was increased, due to the coupled nature of the model and the large model bias present. Also the increase in background error variance led to the analysis of the coupling parameter α^a being further from the 'optimal' value, specifically 0.3731 away from $\alpha^t = 1$. This changed the behaviour of the coupled model trajectories to an extent that the forecast of the model state variable x was degraded.

In general, when the level of bias in a model is high, the coupled 4DVar with coupling parameter estimation scheme requires both an accurate background and accurate observations in order to represent the behaviour of the true state and enable the analysis of the coupling parameters to compensate for the bias in the model. This is even more important for model variables in which coupling parameters are not present, as in these cases the coupling parameters are implicitly compensating for the bias in the model. We next investigate whether the coupled 4DVar with the coupling parameter scheme can compensate for model bias of even greater amplitude.

8.4.4 Amplitude of model bias

We have demonstrated in Section 8.4.1 that even with a high level of model bias, the coupled 4DVar with coupling parameter estimation scheme can improve the coupled forecast accuracy, as opposed to coupled model state estimation only. However, we demonstrate there are limitations on the size of the model bias that this scheme can

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	6.5404	8.2967	11.2699
y	3.4261	10.1293	12.6484
z	7.0395	13.6261	12.6444
w	2.6612	4.0681	10.0414
v	1.6834	4.7229	26.4128

Table 8.5: Reduction in background accuracy for atmospheric variable z. Strong constraint 4DVar used to estimate the coupled model state initial conditions only (Method 1), using the biased idealized coupled model. Average number of minimisation iterations: 55.

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	4.8447	6.9316	11.5792
y	3.0861	8.1755	12.6444
z	4.1530	11.2501	11.4129
w	4.5622	3.6189	7.8518
v	2.2604	3.7921	21.1961

Table 8.6: Reduction in background accuracy for atmospheric variable z. Strong constraint 4DVar with coupling parameter estimation (Method 2) using the biased idealized coupled model. Average number of minimisation iterations: 73. The analysis of the coupling parameter α^a resulted in an average of 0.3731 below $\alpha^t = 1$.

successfully compensate for. We now quadruple the model bias amplitude so that $\boldsymbol{\eta} = (0.2, 0.2, 0.2, 0.2, 0.2)^T$ at each model time-step. With the numerical set up as described at the start of Section 8.4, with the exception of the increased model bias, the resulting RMSE's from a sample of 100 data assimilation runs for both Method 1 and Method 2 are shown in Figure 8.4. Comparing the results in Figure 8.4 plot (b) for Method 1 and Method 2, the coupling parameter estimation is shown only to be successful in reducing the RMSE error throughout the assimilation window for the atmospheric variable y and ocean variable v. The model bias is of such a significantly large amplitude that the coupling parameter estimation has not succeeded in compensating for the model bias and improving the forecast accuracy for all the model state variables, as can be seen in Figure 8.4 plot (c). There is only a significant reduction in the forecast RMSE for the ocean variable v when estimating the coupling parameter along with the coupled model sate. In order to obtain this reduction the analysis of the coupling parameter is on average 1.1257 above the original 'optimal' value of $\alpha^t = 1$. With such a large change in the specification of the coupling parameter, the general behaviour of the coupled model is not maintained and therefore a degraded forecast is obtained for all variables except x and v. We note that there is a substantial reduction in analysis RMSE when simultaneously estimating the coupling parameter along with the coupled model state, as can be seen in plot (a) in Figure 8.4. The increase in model bias has reduced the ability of the coupling parameter to compensate for the bias across the forecast window, but the coupling parameter estimation is able to produce an analysis of increased accuracy.
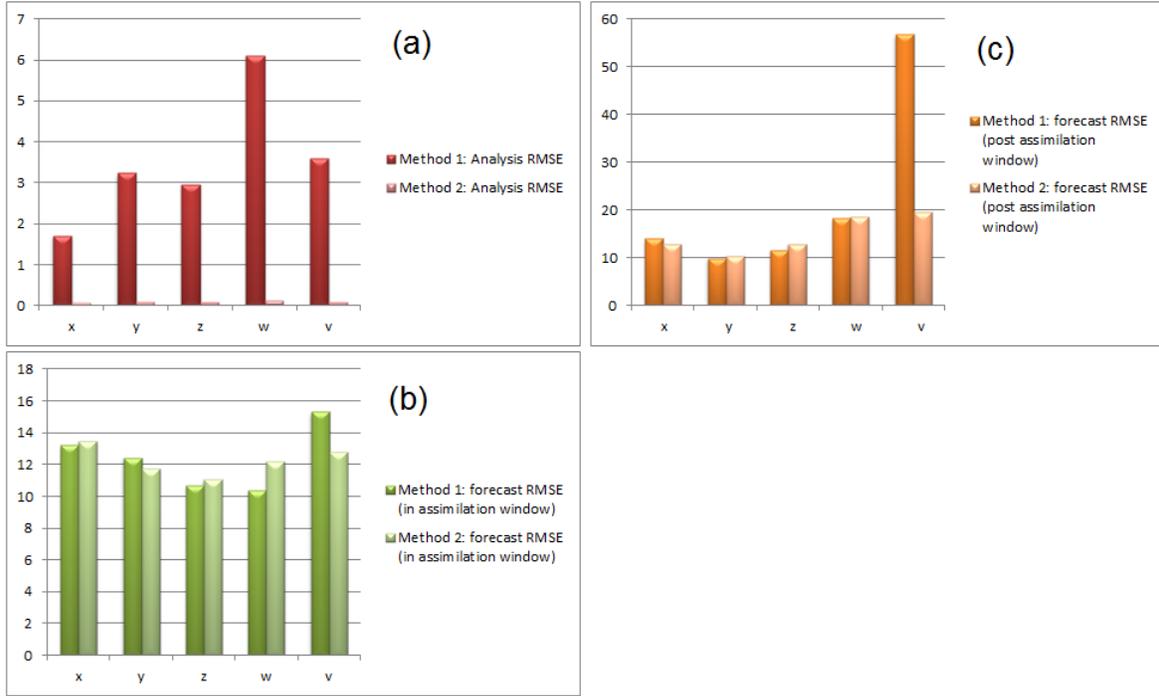
Figure 8.4: Increase in model bias to $\boldsymbol{\eta} = (0.2, 0.2, 0.2, 0.2, 0.2)^T$ at each time-step. Comparison of Method 1 and Method 2 from a sample of 100 data assimilation runs for both methods with the biased idealized coupled model. The Figure shows RMSE errors in the: (a) analysis (b) assimilation window (250 time-steps) and (c) forecast window (500 time-steps). The analysis of the coupling parameter α^a resulted in an average of 1.1257 above $\alpha^t = 1$.

In general, the coupled 4DVar with coupling parameter estimation scheme should be used with caution if the model bias is at such a level that the analysis of the coupling parameters, to compensate for such a high level of bias, are values that are out of a realistic range (that for operational coupled systems would need to be specified by a model expert). This brings us to the next topic of discussion. The cost function (8.4) requires the specification of the 'optimal' coupling parameters $\boldsymbol{\alpha}^t$ and the covariance matrix $\mathbf{F}$. For a generic case, it is obvious that if the variances specified in $\mathbf{F}$ are significantly small, the analysis of coupling parameters are limited to small regions around $\boldsymbol{\alpha}^t$. This can prevent the estimation of coupling parameters

from successfully compensating for bias present in a coupled model. However, caution also needs to be taken not to specify the variances in $\mathbf{F}$ so large that they enable the analysis of the coupling parameters to be so far from the original 'optimal' values $\boldsymbol{\alpha}^t$ to the extent that they change the general behaviour of the coupled system.

8.4.5 Background coupling parameter

In operational coupled atmosphere-ocean models, coupling parameters are best guesses and therefore may not be equal to the 'optimal' values $\boldsymbol{\alpha}^t$. We will denote these coupling parameter best guesses as background coupling parameters $\boldsymbol{\alpha}^b$. Therefore, if this scheme were to be operationally implemented, it may be the case that $\boldsymbol{\alpha}^t$ would be replaced by $\boldsymbol{\alpha}^b$ in the 4DVar cost function (8.4).

We demonstrate using the biased idealized coupled atmosphere-ocean model that when a background coupling parameter α^b is used, as opposed to the 'optimal' value α^t, that the coupled 4DVar with coupling parameter scheme can still successfully compensate for bias in a model and improve a coupled model forecast. With the numerical set up as described at the start of Section 8.4, with the exception that we now use the background coupling parameter $\alpha^b = 1.2$ (20% error) in the biased idealized coupled model and cost function (8.4), we perform a sample of Method 1 and Method 2 minimisations. The resulting RMSEs in the analysis and subsequent model trajectories through both the assimilation and forecast windows are shown in Table 8.7 and Table 8.8 for Method 1 and Method 2 respectively. The coupled 4DVar with coupling parameter scheme is still successful in compensating for the bias in the model and hence improving the coupled forecast (Method 2), in comparison with only estimating the model state (Method 1). The analysis of the coupling parameter α^a resulted in an average of 0.2109 below $\alpha^t = 1$.

To summarise work in Section 8.4, we have demonstrated, with the use of a bi-

209

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	0.5952	8.3758	9.5465
y	0.6222	10.0940	11.2541
z	0.4603	13.5288	13.1486
w	3.6353	4.8505	7.2853
v	2.1834	5.7438	19.6055

Table 8.7: Background coupling parameter in model $\alpha^b = 1.2$. Strong constraint 4DVar used to estimate the coupled model state initial conditions only (Method 1), using the biased idealized coupled model.

Model state variable	Analysis RMSE	RMSE assimilation window	RMSE forecast window
x	3.0074	2.6209	8.1993
y	3.5099	4.2142	9.6016
z	0.6620	5.2283	9.4486
w	6.7301	3.5181	6.0533
v	3.2969	3.0371	16.0080

Table 8.8: Background coupling parameter in model $\alpha^b = 1.2$. Strong constraint 4DVar with coupling parameter estimation (Method 2) using the biased idealized coupled model. The analysis of the coupling parameter α^a resulted in an average of 0.2109 below $\alpha^t = 1$.

ased idealized coupled model, estimation of a coupling parameter along with the initial coupled model state can compensate for bias present in the coupled model and subsequently improve the coupled model forecast, as opposed to coupled model state estimation only. We have shown the reliance the success of the scheme has on how accurately the background and observations represent the behaviour of the true system. We have demonstrated the results are also dependent on the size of the bias present in the coupled model. We next investigate whether this scheme can compensate for error of a random nature in a coupled atmosphere-ocean model.

8.5 Numerical Experiments compensating for random error

We now take the situation where we have a vector of random model error $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$ at each time-step in our coupled model (8.3). We demonstrate that the coupled 4DVar with coupling parameter estimation scheme is not suited to compensating for error of this nature and explain why. We take the experimental set as described in Section 8.4, with the exception that the idealized coupled atmosphere-ocean model is no longer biased and instead has random error at each time-step. We specify the model error covariance matrix $\mathbf{Q}_i$ to be diagonal, specifically with the variances 5×10^{-4} along the diagonal (so that the standard deviation is less than half of the model bias we set in the previous section). We run a experiment with 100 sample runs of Method 1 and Method 2. Both methods use the same set of background vectors, model error vectors and vectors of observations in each data assimilation run. The resulting RMSE's in the analysis, assimilation window and forecast window are shown in plots (a), (b) and (c) in Figure 8.5. The estimation of the coupling parameter (in Method 2) reduces the RMSE in both the analysis and the analysis trajectories throughout the assimilation

window (than that of Method 1). The analysis of the coupling parameter α^a resulted in an average absolute error of 0.0106 from $\alpha^t = 1$, with approximately half the analysis outputs from the sample above $\alpha^t = 1$ and the other half below $\alpha^t = 1$. However, this does not lead to an overall improvement in forecast accuracy subsequent to the end of the assimilation window, as can be seen in Figure 8.5 plot (c).
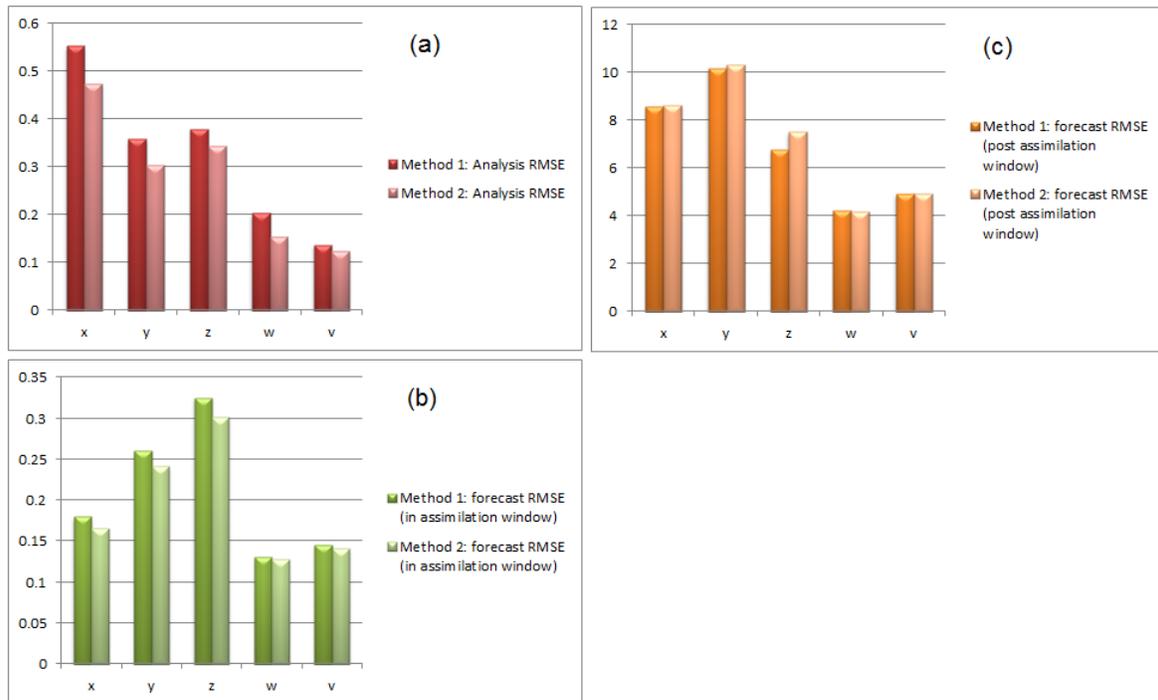
Figure 8.5:   Random error $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_i)$ present in the idealized coupled model at each time-step. Comparison of Method 1 and Method 2 from a sample of 100 data assimilation runs for both methods. The Figure shows RMSE errors in the: (a) analysis (b) assimilation window (250 time-steps) and (c) forecast window (500 time-steps). Average number of minimisation iterations: 13 (Method 1), 53 (Method 2). The analysis of the coupling parameter α^a resulted in an average of 0.0106 above $\alpha^t = 1$.

When the coupling parameter estimation was shown to be successful in compensating for the model bias and subsequently improving the coupled model forecast in Section 8.4, it did so by using the background and observations throughout the assimilation window to describe the behaviour of the true coupled system and estimating

an analysis of the coupling parameter to best fit the coupled model trajectory to this information. The model error in this case was model bias, therefore of a systematic nature, which then continued to be of a systematic nature into the forecast window. This systematic nature of the model error allowed the coupling parameter estimation to amend the model in such a way to reduce the impact of the model bias on the forecast. However, in this section we are dealing with model error of a random nature at each time-step, which is unpredictable. In general, when error of a random nature is present in a coupled model, coupling parameter estimation may indeed lead to a reduction in error within the assimilation window, but the coupling parameter analysis is then not expected to lead to an improvement in forecast accuracy.

We next investigate whether the coupled 4DVar with coupling parameter scheme can improve a coupled model forecast where the coupled model has a different form of systematic error present, as opposed to model bias.

8.6 Numerical Experiments compensating for static parameter error

Parameterization schemes are used to represent processes which are not modelled explicitly. As previously discussed in Section 3.1.2, the specification of atmospheric and oceanic parameter values is often a source of model error in coupled atmospheric-oceanic models. We define the 'optimal' atmospheric and oceanic parameter values to be the specifications that best represent the processes. When both the 'optimal' and erroneous parameter values are time invariant, this leads to error in a coupled model of a systematic nature. For numerical experiments presented in this section we use the idealized coupled atmosphere-ocean model [87], discretized in time as described in Section 5.2, as our erroneous model equations (8.2). We denote the erroneous atmospheric model parameter $\sigma_m = 10$ and oceanic model parameter $k_m = 0.1$. In

this section we no longer consider model bias or error of a random nature to be present in the coupled model. We create our reference 'true' model by perturbing the atmospheric parameter σ_m and the oceanic parameter k_m using a method known as stochastic representation of model uncertainties [23]. This parameter perturbation method has previously been used for ensemble forecasting, with the perturbations taken from a uniform distribution. We use this method to create our true model parameters σ_t and k_t, but instead of using a uniform distribution, we use the following Gaussian distributions,

$$
\begin{aligned}
\sigma_t &= \gamma_\sigma \sigma_m, & \gamma_\sigma &\sim \mathcal{N}(\mathbf{1}, \frac{1}{144}), \\
k_t &= \gamma_k k_m, & \gamma_k &\sim \mathcal{N}(\mathbf{1}, \frac{1}{36}).
\end{aligned}
\tag{8.8}
$$

Once evaluated σ_t and k_t are held constant in our reference 'true' coupled model (8.1). The atmospheric parameter σ_t is varied at a 'low' level with standard deviation $\frac{1}{12}$ and the oceanic parameter k_t is varied at a 'medium' level with standard deviation $\frac{1}{6}$. We specify the true initial conditions $x^t_0 = -3.4866$, $y^t_0 = -5.7699$, $z^t_0 = 18.3410$, $w^t_0 = -10.7175$, $v^t_0 = -7.1902$, as defined at the start of Section 8.2. Figure 8.6 shows an example of the idealized coupled model run for 750 time-steps, firstly with the true model parameters $\sigma_t = 8.1176$ and $k_t = 0.1144$ (calculated as described above) and secondly with the erroneous model parameters $\sigma_m = 10$ and $k_m = 0.1$, where both models have been initialised with the same true initial conditions. There are significant levels of error in the model state trajectories for all of the variables, with the atmospheric variables on the wrong side of the model attractor for the majority of the second half of the time window and the ocean variables drifting from the true values.

214
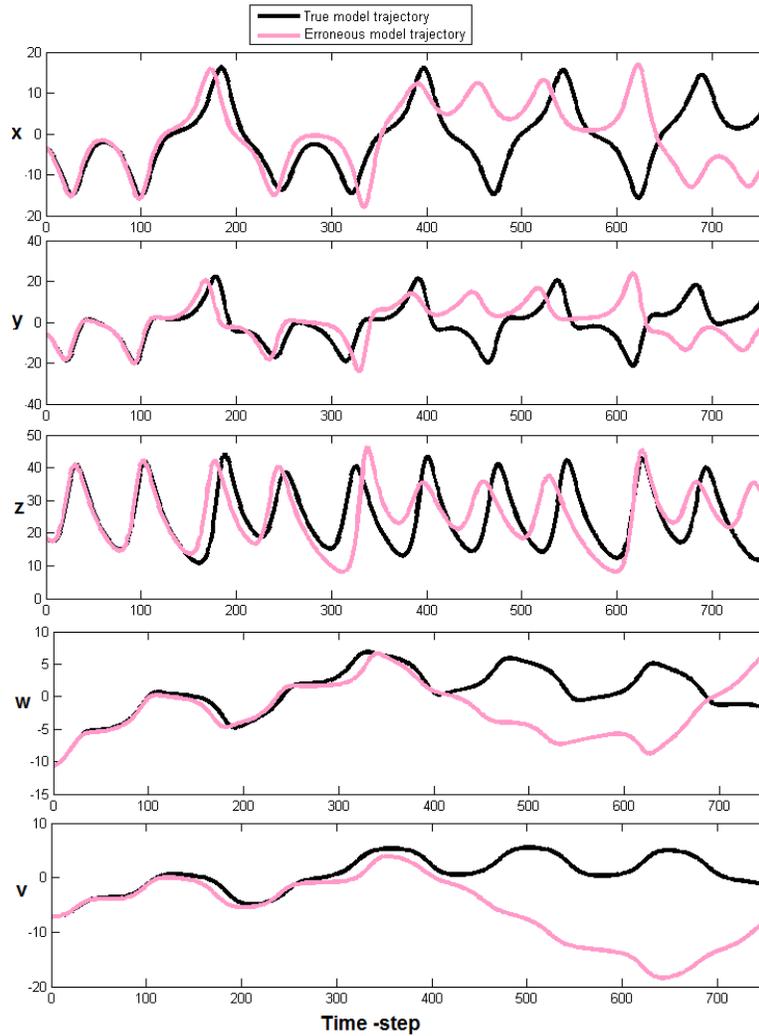
Figure 8.6: True idealized coupled model trajectories with $\sigma_t = 8.1176$ and $k_t = 0.1144$ (black), erroneous idealized coupled model trajectories with $\sigma_m = 10$ and $k_m = 0.1$ (pink). Both models initialised with the same initial conditions $(x^t{}_0, y^t{}_0, z^t{}_0, w^t{}_0, v^t{}_0)^T = (-3.4866, -5.7699, 18.3410, -10.7175, -7.1902)^T$.

We will next investigate the ability of the coupled 4DVar with coupling parameter estimation scheme to compensate for the erroneous atmospheric and oceanic param-eter values in the idealized coupled model. We emphasise that is not the atmospheric and oceanic parameter values we will be estimating, it is the coupled model state and

the coupling parameter. We will use the same data assimilation set up as described at the start of Section 8.4, with the exception that the source of the error in the coupled model is now parameter errors as described above, as opposed to model bias. We will also extend the assimilation window length to 500 time-steps. Looking at Figure 8.6, we can see this time length of 500 time-steps allows the parameter errors to create a significant difference in the model state trajectories, particularly towards the end of the window. We also increase the variance of the coupling parameter to $\sigma_\alpha{}^2 = 10^{-2}$ in the cost function (8.4), to allow the coupling parameter to deviate from the true value in order to try to mitigate the effect of the parameter errors across this longer window. To recap, the error covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$ are set as defined as conditions A in Table 8.9.

We perform data assimilation with Method 1 (coupled 4DVar) and Method 2 (coupled 4DVar with coupling parameter estimation) and compare the analysis and subsequent analysis trajectories by computing the RMSE (Root Mean Square Error) from a sample of analysis outputs. We conduct the following steps 100 times.

1. Produce a background vector $\mathbf{x}^b$ by adding noise, consistent with the statistics prescribed in $\mathbf{B}$, to the true initial state $\mathbf{x}^t{}_0$.

2. Compute the true model parameters σ_t and k_t as described in (8.8).

3. Produce a vector of observations $\mathbf{y}_i$ at each observation time t_i by evaluating the true model state vector $\mathbf{x}^t{}_i$ with equation (8.1) with the true coupling parameters σ_t and k_t present and subsequently adding noise in proportion to the error statistics specified in $\mathbf{R}_i$.

4. Use the coupled model (8.2) with erroneous parameter values $\sigma_m = 10$ and $k_m = 0.1$ in the data assimilation methods (Method 1 and Method 2) to compute the analysis outputs.

216

5. Initialize the coupled model with the respective analysis outputs to produce a coupled model forecast (over 500 time-steps) for both Method 1 and Method 2.

8.6.1 Results

With the experimental set up as just described, the results comparing the performance of Method 1 (estimating the coupled model state only) and Method 2 (simultaneously estimating the coupled model state and the coupling parameter) are shown in Figure 8.7. We observe that the coupling parameter estimation has partially compensated for the atmosphere and ocean parameter errors in the assimilation window, with a reduction in RMSE as shown in plot (b) in Figure 8.7. This has lead to improvements in the subsequent forecast accuracy of each of the model state variables, as can be seen in plot (c) in Figure 8.7. This illustrates the ability of the coupling parameter estimation to partially compensate for the static atmospheric and oceanic parameter errors and produce analysis trajectories that are a closer representation of the true behaviour of the model state variables. The analysis of the coupling parameter α^a resulted in an average absolute error of 0.0526, with the values of α^a equally distributed above and below the true value α^t. Again, we note that an increase in forecast accuracy does not necessarily lead to an increase in analysis accuracy, which can be seen when comparing the analysis RMSE's in plot (a) between Method 1 and Method 2 for the ocean variable v. The least significant increase in forecast accuracy, comparing the forecast RMSE from Method 1 and Method 2, is for the atmospheric variable z. This is because the model equation for this variable (5.4) does not contain either of the erroneous parameter values σ_m and k_m and also does not contain the coupling parameter α.

We next demonstrate that the increase in forecast accuracy, from coupling parameter estimation, does not remain for an indefinite time. We have extended the forecast

window by 500 time-steps to a total length of $1,000$ time-steps and calculated the corresponding forecast RMSE's which are shown in plot (d) in Figure 8.7. With this increase in the forecast window length, the forecast accuracy is no longer improved with coupling parameter estimation for the atmospheric variable z.
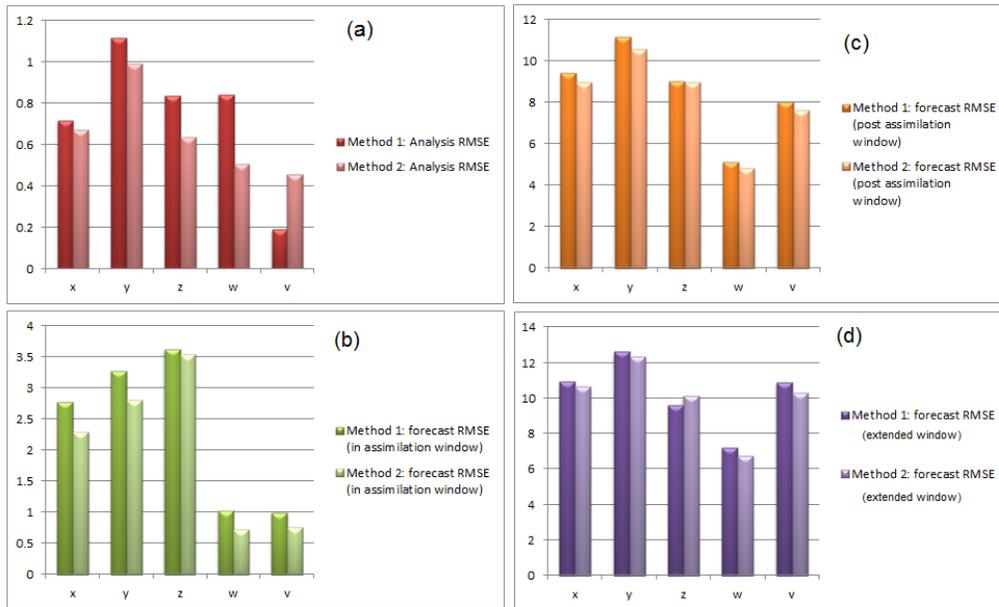
Figure 8.7: Parameter errors in σ_m and k_m in idealized coupled model. Comparison of Method 1 and Method 2 from a sample of 100 data assimilation runs for both methods. The Figure shows RMSE's in the: (a) analysis (b) assimilation window (500 time-steps) (c) forecast window (500 time-steps) and the (d) extended forecast window (1000 time-steps). Average number of minimisation iterations: 53 (Method 1), 80 (Method 2). The analysis of the coupling parameter α^a resulted in an average absolute error of 0.0526.

We next conduct numerical experiments to investigate how the following changes in the data assimilation set-up conditions affect the results.

- Conditions A: No changes (see Table 8.9).

- Conditions B: Reduction in the accuracy of both the atmosphere observations and ocean observations (see Table 8.9).

- Conditions C: Reduction in the frequency of the ocean observations,

  - from every 10 time-steps to every 20 time-steps.

- Conditions D: Reduction in the accuracy of the background model state (see Table 8.9).

- Conditions E: Increase in the amplitude of the atmosphere and ocean parameter errors,

  - σ_m varied at a 'medium' level with standard deviation $\frac{1}{6}$ and k_m varied at a 'high' level with standard deviation $\frac{1}{3}$ (8.8).

- Conditions F: The true coupling parameter is unknown to the data assimilation process,

  - instead we set the background coupling parameter $\alpha^b = 1.2$.

We present results from a sample size of 100 data assimilation runs using both Method 1 (estimation of the coupled model state only) and Method 2 (coupled 4DVar with coupling parameter estimation) for each set of conditions (A-F). We summarise the results in Figure 8.8, where the overall combined coupled model state % increase in RMSE is presented, from using Method 2 as opposed to Method 1. The RMSE has been computed with all variables for Method 1 and Method 2 respectively. Subsequently the % increase from using Method 2 as opposed to Method 1 has been computed. The presence of negative values indicate that the coupling parameter estimation has compensated for the model error and hence increased the forecast accuracy. The results shown in Figure 8.8 for conditions A use the same data as presented in Figure 8.7, which we use as our control conditions and results. The results in plot (c) Figure 8.8 show that use of all the conditions (A-F) result in an increase in forecast accuracy, however the extent of which varies depending on the specification

of the conditions. It should be noted again that the increase in forecast accuracy does not necessarily originate from an increase in analysis accuracy, as can be seen in plot (a) in Figure 8.8. The reason for this is that this method aims to use coupling parameter estimation to compensate for the bias in the model. At the initial time there is no effect from the bias in the model as the model has not yet been evolved.

For the coupling parameter to compensate for the atmosphere and ocean parameter errors, accurate observations are required in order to represent the true coupled dynamics. We next demonstrate that a decrease in observational accuracy has a detrimental effect on the success of the developed scheme. Using conditions B as stated in Table 8.9, where the observation error standard deviations have been increased by a factor of 10, the accuracy of both the atmosphere and ocean observations have been degraded leading to the coupled 4DVar with coupling parameter scheme (Method 2) producing analysis trajectories of less accuracy throughout the assimilation window, than when not estimating the coupling parameter (Method 1), as shown in plot (b) in Figure 8.8. There is still however a small increase in forecast accuracy post the assimilation window, shown in Figure 8.8, as the coupled model behaviour towards the end of the assimilation window is better captured with the analysis of the coupling parameter (than with the true value) when there is model error present, which subsequently leads to forecast trajectories that begin closer to the true conditions. The general result is that the more accurate the observations are, the more successfully the coupling parameter can compensate for the atmosphere and ocean parameter errors.

Conditions	Covariance matrix	Variance for x	Variance for y	Variance for z	Variance for w	Variance for v
A, B, C, E, F	**B**	4×10^{-3}	4×10^{-3}	4×10^{-3}	10^{-3}	10^{-3}
A, C, D, E, F	$\mathbf{R}_i$	1.6×10^{-5}	1.6×10^{-5}	6.4×10^{-5}	10^{-6}	10^{-6}
B	$\mathbf{R}_i$	1.6×10^{-3}	1.6×10^{-3}	6.4×10^{-3}	10^{-4}	10^{-4}
D	**B**	4×10^{-1}	4×10^{-1}	4×10^{-1}	10^{-1}	10^{-1}

Table 8.9: Error covariance matrices used in Method 1 and Method 2 with the idealized coupled atmosphere-ocean model with parameter errors. Condition A is as used in the initial experiment at the start of Section 8.6.1 .
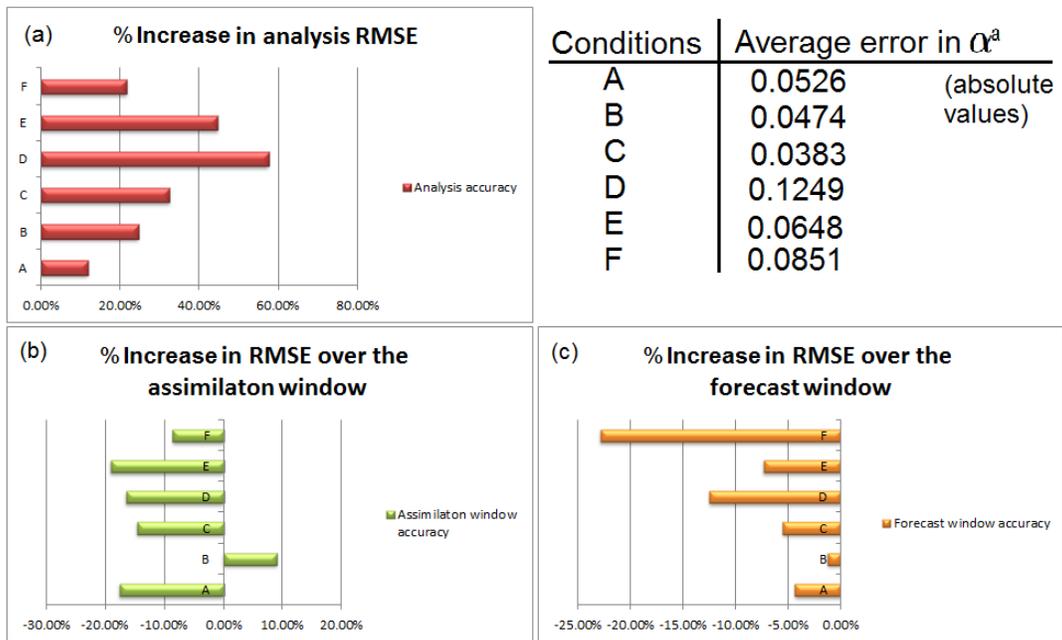
Figure 8.8:   Parameter errors in σ_m and k_m in the idealized coupled model. Comparison of Method 1 and Method 2 from a sample of 100 data assimilation runs. The Figure shows overall % increase in RMSE in the: (a) analysis (b) assimilation window (500 time-steps) and (c) forecast window (500 time-steps) from using Method 2 as opposed to Method 1.

Operationally, observations of the ocean are available less frequently in time than observations of the atmosphere. The ocean is a slow moving system, so as long as there are observations at a sufficient number of points in time throughout an assimilation window, the general behaviour of the true ocean model state variables should still

221

be reasonably represented. With less frequent ocean observations (than observations of the atmosphere) the coupled 4DVAR with coupling parameter estimation scheme should still be able to compensate for static atmosphere and ocean parameter errors in a coupled model. We demonstrate this with a reduction in the ocean observation frequency (to every 20 time-steps as opposed to 10 time-steps) as in 'Conditions C'. There is a significant % reduction RMSE in the resulting RMSE's when estimating the coupling parameter along with the model state (Method 2), as opposed to estimating the model state only (Method 1), as shown in Figure 8.8. When comparing the results in Figure 8.8 when using conditions C as opposed to conditions A, i.e. a reduction in ocean observation frequency, we observe that the reduction in % RMSE is less in the assimilation window, however it is slightly larger in the forecast window.

We next demonstrate that increasing the error variances of the background coupled model state can lead to a more significant improvement in the forecast accuracy, when using Method 2 as opposed to Method 1. We significantly increase the background error standard deviations by a factor of 10 for all atmosphere and ocean model state variables, as stated in conditions D in Table 8.9. The resulting RMSE % reduction is shown in Figure 8.8. Comparing the results from this experiment using conditions D with those from the control experiment using conditions A, we observe a much larger % reduction RMSE in the coupled forecast when estimating the coupling parameter along with the coupled model state using conditions D, see plot (c). Increasing the background variance, puts less weight on the background term in the 4DVar cost function. When estimating the coupling parameter along with the coupled model state in order to compensate for model error, this can lead to a significant decrease in analysis accuracy, as can be seen in the plot (a) in Figure 8.8. However, the decrease in weight of the background term in the cost function can allow the coupled model state trajectories to better fit the vicinity of the observations towards the end of the assimilation window and with coupling parameter estimation capture the

general behaviour of the coupled model with more accuracy in the latter part of the assimilation window which can lead to a subsequent improvement in the forecast accuracy. Let us remark that the analysis of the coupling parameter α^a resulted in an average absolute error of 0.1249 from the true value α^t, which is a significantly larger error than the control run, in order to fit the coupled model state analysis trajectories in close vicinity to the observations in the latter part of the assimilation window.

The increase in forecast accuracy simultaneously estimating the coupling parameter along with the model state, as opposed to coupled model state estimation only, increases with an increase in amplitude of the parameter errors. We demonstrate this result by increasing the standard deviations of the atmospheric and oceanic parameter errors by two, as stated in conditions E, where σ_m is now varied at a 'medium' level with standard deviation $\frac{1}{6}$ and k_m is now varied at a 'high' level with standard deviation $\frac{1}{3}$ (8.8). The % reduction in RMSE results are shown in Figure 8.8 for these conditions, where in comparison to the control run (conditions A), there are increased % reductions in the RMSE's in both the assimilation and forecast window. This is because when not accounting for model error (Method 1), the data assimilation scheme is unable to closely fit the erroneous model state trajectories to the observations. Whereas, the coupling parameter estimation enables the model state trajectories to better fit to the close vicinity of the observations.

So far work in this section has assumed that the 'optimal' value of the coupling parameter in the idealized coupled atmosphere-ocean model (5.4) is known, specifically $\alpha^t = 1$. However, in operational coupled atmosphere-ocean models coupling parameters are best guesses and therefore may not be equal to the 'optimal' values. We demonstrate that even when the 'optimal' coupling parameter α^t in the idealized coupled model and hence in the cost function (8.4) is replaced with a 'background' coupling parameter α^b, that Method 2 is still successful in compensating for the errors in the model and hence improving the coupled forecast. We replace $\alpha^t = 1$ with

223

$\alpha^b = 1.2$ in the erroneous idealized coupled model and denote these conditions F. The results in Figure 8.8 show that the coupled 4DVar with coupling parameter scheme (Method 2) is still successful in compensating for the parameter errors in the model and hence improving the coupled forecast, in comparison to when estimating the coupled model state only (Method 1). The analysis of the coupling parameter α^a resulted in an average of 0.0851 away from $\alpha^t = 1$, which is a significant increase from the 'control' run using conditions A. The % reduction in forecast RMSE has significantly increased with these conditions F in comparison to the control conditions A. This is because there is now error in the coupling parameter α^b as well as the atmosphere and ocean model parameters σ_m and k_m and when not accounting for the parameter errors (Method 1), this causes a significant degradation in forecast accuracy.

Coupled 4DVar with coupling parameter estimation can improve a coupled model forecast, when compared with the forecast initialised with the estimation of the initial coupled model state only. We have shown, with use of an idealized coupled atmosphere-ocean model, that coupling parameter estimation can compensate for both model bias and time invariant atmosphere and ocean parameter errors, which can then subsequently improve a coupled forecast. We have shown that our developed scheme is not able to mitigate the effect random errors present in a coupled model have on the coupled forecast.

We predict that coupled 4DVar with a coupling parameter estimation scheme can compensate for other types of systematic error, as well as model bias and static parameter errors that we have focused on, and this is discussed in Section 9.2, further work, of this thesis. Further work also includes how the covariance matrix $\mathbf{F}$ in the 4DVar cost function (8.4) should be specified for operational coupled models. Other proposed further work is to consider the estimation of time varying coupling parameters in order to better compensate for the error within the assimilation window. We define time varying coupling parameters as coupling parameters that are not

224

static over the full assimilation window, but instead are static for shorter periods within the window. Numerical experiments conducted in this section have used the idealized coupled atmosphere-ocean model [87]. Further work involves extending this investigation to use more complex coupled atmosphere-ocean models which are of a more realistic nature.

8.7   Summary

Coupled atmosphere-ocean models at operational NWP centres are used to produce seasonal to interannual forecasts. Bulk formulae represent fluxes at the atmosphere-ocean interface in the coupled model equations, where these formulae require the specification of coupling parameters. Operational centres are moving towards coupled data assimilation, where the initial conditions of both the atmospheric and oceanic variables are simultaneously estimated. We have extended the idea of coupled model state estimation to also simultaneously estimating coupling parameters in the strong constraint 4DVar framework. We have demonstrated, with use of an idealized coupled atmosphere-ocean model, how estimation of a coupling parameter can compensate for both model bias and static atmosphere and ocean parameter errors present in a coupled model and hence improve the accuracy of the coupled forecast.

Chapter 9

Conclusions

Variational data assimilation methods aim to best estimate the state of a system, with practical applications including both atmospheric and oceanic forecasting. We acknowledge in this thesis that models are representations of true dynamical systems and therefore consist of incorrect equations. The inconsistencies these incorrect equations have with the truth can be of a random or systematic nature. In this thesis we define these inconsistencies as 'model error'. In particular we focus on the fact that coupled atmosphere-ocean models contain model error [104] [54] [34]. In Chapter 3 we outlined current variational data assimilation methods used at operational NWP centres for both forecasting and reanalysis. A current objective of operational weather centres is how to better account for model error in the data assimilation process.

In chapter 3 we described the use of consistency diagnostics, that assume a correct model, as quality checks for the specification of both background error and observation error covariance matrices. In Chapter 4 we explained the current difficulties in specifying model error covariance matrices for operational models at NWP centres and highlighted the negative effects the use of weak constraint 4DVar with inaccurately specified model error covariance matrices can have on forecasts. In Chapter

4, we also examined data assimilation methods that have been formulated to compensate for bias present in coupled atmosphere-ocean models, with the objective to improve coupled model forecasts. Compiling this background research provided us with the motivation for the work in this thesis. The key objectives of this thesis were to develop methods in two key areas; firstly, to verify and refine an estimated model error covariance matrix and secondly, to account for model error in the data assimilation process without the need for the explicit specification of the model error statistics. The latter of these tasks is split into two sub areas; accounting for model error to improve the estimation of the initial conditions and accounting for model error to improve the accuracy of a coupled atmosphere-ocean model forecast. We next summarise the work carried out in this thesis and the subsequent conclusions drawn and then suggest ideas for further work.

9.1 Conclusions

A set of four diagnostics tools were developed by Desroziers et al. [38] as consistency checks for the specification of both background error and observation error covariance matrices. These diagnostic tools were not formulated to include model evolution and hence when used with model evolution, do not account for any errors present in the model. Practical use of these diagnostic tools at operational weather centres has been conducted with model evolution included [2]. We know that operational NWP models contain error and work in Chapter 6 involved investigating the effect random model error has on the diagnostic tools. One strong constraint 4DVar diagnostic equation has been derived previously which accounts for random error present in a model [33] [2]. In Chapter 6 we developed three further strong constraint 4DVar diagnostic equations that account for random error present in a model. We conducted comparisons of the diagnostic tools that account for random model error with the set

of diagnostic tools that assume a perfect model and observed an additional term involving the model error covariance matrix present in each respective diagnostic equation. We emphasized that caution should be taken, if random model error is or could be present, when using any of the four diagnostic equations that assume a perfect model.

In Chapter 6, we next demonstrated that the set of four diagnostic equations, that account for the presence of random error in a model, can be used as quality checks to verify the consistency of an estimated model error covariance matrix, in observation space, with both background error and observation error covariance matrices. We subsequently described how the strong constraint 4DVar diagnostic tools have the potential, under certain conditions, to be used to refine an estimated model error covariance matrix.

Weak constraint 4DVar relies on the accurate specification of a model error covariance matrix. In Chapter 6 we developed four diagnostic equations specifically for the weak constraint formulation of 4DVar. We subsequently demonstrated the use of these weak constraint 4DVar diagnostic equations to verify the consistency of an estimated model error covariance matrix, in observation space, with both background error and observation error covariance matrices. The theory developed in this chapter is not specific to erroneous models of atmospheric and oceanic dynamics and can be used with any erroneous model where the error is of a random nature at each time-step.

In Chapter 4 we described data assimilation techniques developed to acknowledge that the comparison between observations and the model state, in the presence of representativity error, is not optimal without amending the cost function [56]. We then discussed the relationship between representativity error and model error and subsequently identified the fact that methods to deal with representativity error had the potential to be further developed to account for model error. In Chapter 7

we developed a combined model error and observation error covariance matrix that accounts for the errors in the comparison of observations with a model evolved initial state (mapped to observation space), where the model is of an erroneous nature with random error at each time-step. When a model with random error at each time-step is used within strong constraint 4DVar, replacing the observation error covariance matrix with this combined error covariance matrix, produces a statistically better estimate of the initial state. This replacement ensures the 4DVar problem is formulated to be mathematically correct, in order to best estimate the initial state with use of an erroneous model of this nature.

The equation we derived for the combined model error and observation error covariance matrix (7.9) includes the specification of model error covariance matrices. As previously discussed, these are often unknown and therefore we developed a method, with use of diagnostic tools, to estimate the combined model error and observation error covariance matrix without the need for explicit specification of the model error statistics. We subsequently demonstrated the successful application of our developed 4DVar method with the estimated combined model error and observation error covariance matrix. These numerical experiments involved use of an erroneous idealised nonlinear coupled model, where the coupled model analysis was of higher accuracy when using the estimated combined error variances in the cost function, as opposed to the observation error variances alone. Similarly to Chapter 6, the theory developed in Chapter 7 is not specific to erroneous models of atmospheric and oceanic dynamics and can be used with any erroneous model where the error is of a random nature at each time-step.

In Chapter 8 we focused our attention on coupled atmosphere-ocean models, which at operational NWP centres are used to produce seasonal to interannual forecasts. Bulk formulae represent fluxes at the atmosphere-ocean interface in the coupled model equations, where these formulae require the specification of coupling parameters. Op-

erational centres are moving towards coupled data assimilation, where the initial conditions of both the atmospheric and oceanic variables are simultaneously estimated. We extended the idea of coupled model state estimation to also simultaneously estimating coupling parameters by augmenting the strong constraint 4DVar control vector to include the model's coupling parameters along with the model state initial conditions. We demonstrated, with use of an idealized coupled atmosphere-ocean model, how estimation of a coupling parameter can compensate for both model bias and static atmosphere and ocean parameter errors present in a coupled model and hence improve the accuracy of the coupled forecast. We investigated and presented the conditions under which coupling parameter estimation can best compensate for these systematic errors in the model. These conditions include accurate observations of the model state variables to describe the general behaviour of the true model trajectories throughout the assimilation window.

9.2   Further work

Methods developed in this thesis have been formulated with the future objective to be of practical use at operational NWP centres. Prior to possible implementation, we suggest the following ideas for further work.

Work in Chapter 6 involved the derivation of diagnostic equations that accounted for model error of a random nature. We developed and demonstrated use of both strong constraint and weak constraint 4DVar consistency diagnostic equations (6.8), (6.10), (6.12), (6.13) and (6.29), (6.33), (6.35), (6.36) respectively. It was assumed that samples of innovation vectors that include samples of model error were available for use in the diagnostic equations. In 'toy' models we can create the model error vectors required as we have the model error covariance matrix from which we can sample vectors of random error. However, this is certainly not the case for operational

models of the atmosphere and ocean. Therefore, how to obtain the required samples of innovation vectors operationally is an area of further work.

Of course there is no 'true' forecast model of the atmosphere and ocean. Operational weather centres have forecast models that best describe the atmosphere and ocean as well as they can, but contain errors. What operational centres do also have are ensemble prediction systems that aim to represent random error in a model forecast using stochastic physics [23]. Therefore, we suggest use of these ensemble prediction systems to obtain the innovation data. We have two suggested options to do so: Firstly, obtaining a sample of background vectors at a unique time and using an ensemble prediction system to evolve the sample of background vectors to the subsequent time. A sample of observation vectors are required at this subsequent time. Secondly, again using the ensemble prediction system to evolve a sample of background vectors, but using innovations across multiple times and therefore allowing use of observational data from different times and considering a time-averaged model error covariance matrix. The first of these methods would be advantageous as this would allow for the model error covariance matrix to evolve with each time-step. However, this is dependent on the availability of the sample observation data at a unique time. The next step is to work with weather centres to discuss the availability of innovation data.

Further work also involves the possible simplification and the possible use of the weak constraint 4DVar consistency diagnostics (6.29), (6.33), (6.35) and (6.36) we derived in Chapter 6 to enable an estimated model error covariance matrix to be refined.

Work in Chapter 7 involved the derivation of a combined model error and observation error covariance matrix. Both an explicit calculation and an estimation (using diagnostic tools) of the combined model error and observation error covariance matrix were conducted in Chapter 7 with a nonlinear erroneous idealized coupled

atmosphere-ocean model. Further work should be conducted to investigate which of these combined covariance matrices, explicit calculation or estimation with diagnostic tools, is the most successful combined model error and observation error statistics to use in the 4DVar cost function with a nonlinear erroneous model to improve the analysis accuracy the most. Although, it should be kept in mind that if this were to be conducted operationally the method used would be estimation (using the diagnostic tools) as the model error covariance matrix is unknown.

Work in Chapter 7 assumed that the true background error statistics were known in the estimation of the combined model error and observation error covariance matrix. Operationally background error statistics are estimated and therefore are likely to contain inaccuracies. Further work involves the investigation of the possible detrimental implications an inaccurately specified background error covariance matrix could have on the estimation of the combined error statistics. Further to this, investigations should be conducted into the implications an inaccurately specified background error covariance matrix has on the analysis accuracy when using estimated combined model error and observation error covariance statistics in 4DVar.

Demonstrations we conducted in Chapter 7 to show the success of the developed method, to account for model error and hence improve the analysis, involved use of an erroneous idealised nonlinear coupled model, where only the diagonal entries of the combined model error and observation error matrix were evaluated and used in the process. An investigation into the significance of including cross covariance information within the combined error matrix should be conducted. We hypothesise that the inclusion of this cross covariance information will further improve the analysis accuracy. Practical limitations on specifying very large full error covariance matrices should be kept in mind and therefore investigation into which cross covariance information is most important to include and why should be conducted. Further work should also involve implementation of the developed method, to both estimate the

combined statistics and then subsequently use these in the strong constraint 4DVar cost function, with an erroneous coupled atmosphere-ocean model of a more complex nature.

Although the objectives of strong constraint 4DVar and weak constraint 4DVar are different, an interesting comparison would be the analysis accuracy and the subsequent forecast accuracy from strong constraint 4DVar, strong constraint 4VDar with combined model error and observation error statistics and weak constraint 4DVar. This would allow a summary of which method could be most appropriate to use depending on the objective of the assimilation. We hypothesise that strong constraint 4DVar with combined error statistics would be most appropriate when the the objective is to best estimate the initial model state, however to improve the accuracy of the analysis trajectory across the assimilation window weak constraint 4DVar should be used.

Work in Chapter 8 of this thesis involved coupling parameter estimation, in the strong constraint 4DVar framework, to compensate for both model bias and static atmosphere and ocean parameter errors present in a coupled model and hence improve the accuracy of the coupled forecast. The investigation into how best to specify the variances for the coupling parameters for operational coupled models should be conducted. Further work should involve contact with operational weather centres to assess which coupling parameters are most suitable for inclusion in the scheme. We suggest that future work in this area should also include estimation of time varying coupling parameters with the aim to better compensate for the error in the model over time. Specifically, whether splitting an assimilation window into sub intervals, where coupling parameters were estimated and stayed constant only in each subwindow, could further improve the accuracy of the analysis trajectory. It may be possible characteristic features could be identified, for example certain values of coupling parameters compensating for errors during the day and other values of coupling

parameters better compensating for the errors over night. This information could then be applied to improve coupled forecasts further and possibly for a longer length of time.

We hypothesise that the less chaotic the nature of a model, the longer the coupled 4DVar with coupling parameter estimation scheme can improve the forecast accuracy. Testing this statement further is an area of suggested work. Demonstrations in Chapter 8 used an idealized coupled atmosphere-ocean model, firstly that was biased and secondly, that contained static atmosphere and ocean parameter errors. Future work involves extending this investigation to compensate for other types of systematic error present in a coupled model and the combination of multiple forms of systematic errors present in a coupled model. We suggest that future work should also involve implementation of the coupled 4DVar with coupling parameter estimation scheme with an erroneous coupled atmosphere-ocean model of a more complex nature. The complex coupled model under consideration could contain multiple coupling parameters to be estimated along with the coupled model state.

Bibliography

[1] C.D. Ahrens. *Meteorology today: An introduction to Weather, Climate, and the Environment.* Cengage Learning, 2012.

[2] E. Andersson. Modelling the temporal evolution of innovation statistics. In *Recent developments in data assimilation for atmosphere and ocean, ECMWF,* pages 153–164, 2003.

[3] E. Andersson, M. Fisher, R. Munro, and A. McNally. Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation scheme, and the explanation of a case of poor convergence. *Quarterly Journal of the Royal Meteorological Society,* 126(565):1455–1472, 2000.

[4] M.A. Balmaseda and D. Anderson. Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters,* 36(1), 2009.

[5] M.A. Balmaseda, K. Mogensen, and A.T. Weaver. Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society,* 139(674):1132–1161, 2013.

[6] M.A. Balmaseda, A. Vidard, and D.L.T. Anderson. The ECMWF Ocean Analysis System: ORA-S3. *Monthly Weather Review,* 136(8):3018–3034, 2008.

[7] R.N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970, 2008.

[8] R.N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error covariance statistics. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1971–1996, 2008.

[9] S. Bélair, J. Mailhot, C. Girard, and C.P. Vaillancourt. Boundary layer and shallow cumulus clouds in a medium-range forecast of a large-scale weather system. *Monthly weather review*, 133(7):1938–1960, 2005.

[10] M.J. Bell, R.M. Forbes, and A. Hines. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *Journal of Marine Systems*, 25(1):1–22, 2000.

[11] M.J. Bell, M.J. Martin, and N.K. Nichols. Assimilation of data into an ocean model with systematic errors near the equator. *Quarterly Journal of the Royal Meteorological Society*, 130(598):873–893, 2004.

[12] C.M. Biouele. *Earth's Atmosphere Dynamic Balance Meteorology*. Scientific Research Publishing, Inc. USA, 2015.

[13] J. Bjerknes. A possible response of the atmospheric Hadley circulation to equatorial anomalies of ocean temperature. *Tellus*, 18(4):820–829, 1966.

[14] E. Blayo, M. Bocquet, E. Cosme, and L.F. Cugliandolo. Advanced data assimilation for geosciences. In *International Summer School-Advanced Data Assimilation for Geosciences*, page 608. Oxford University Press, 2014.

[15] E.W. Blockley, M.J. Martin, A.J. McLaren, A.G. Ryan, J. Waters, D.J. Lea, I. Mirouze, K.A. Peterson, A. Sellar, and D. Storkey. Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new global FOAM forecasts. *Geoscientific Model Development*, 7(6):2613–2638, 2014.

[16] Space Studies Board. *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*. National Academies Press, 2007.

[17] F. Bouttier and P. Courtier. Data assimilation concepts and methods March 1999. *Meteorological training course lecture series, ECMWF*, 2002.

[18] P.A. Bouttier, E. Blayo, J.M. Brankart, P. Brasseur, E. Cosme, J. Verron, and A. Vidard. Toward a data assimilation system for NEMO. *Mercator Ocean Quarterly Newsletter*, (46):24–30, 2012.

[19] N.E. Bowler, A. Arribas, K.R. Mylne, K.B. Robertson, and S.E. Beare. The MOGREPS short-range ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 134(632):703–722, 2008.

[20] J.M. Brankart, G. Candille, F. Garnier, C. Calone, A. Melet, P.A. Bouttier, P. Brasseur, and J. Vernon. A generic approach to explicit simulation of uncertainty in the NEMO model. *Geoscientific Model Development*, 8(5):1285–1297, 2015.

[21] G.B. Brassington, M.J. Martin, H.L. Tolman, S. Akella, M. Balmeseda, C.R.S. Chambers, E. Chassignet, J.A. Cummings, Y. Drillet, P.A.E.M Jansen, et al. Progress and challenges in short-to medium-range coupled prediction. *Journal of Operational Oceanography*, 8(2):239–258, 2015.

[22] M. Buehner, P. Gauthier, and Z. Liu. Evaluation of new estimates of background-and observation-error covariances for variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3373–3383, 2005.

[23] R. Buizza, M. Milleer, and T.N. Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, 125(560):2887–2908, 1999.

[24] A. Carrassi and S. Vannitsem. Accounting for model error in variational data assimilation: A deterministic formulation. *Monthly Weather Review*, 138(9), 2010.

[25] F. Castelli, D. Entekhabi, and E. Caporali. Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance. *Water Resources Research*, 35(10):3115–3125, 1999.

[26] D.M. Causon and C.G. Mingham. *Introductory Finite Difference Methods for PDEs*. Bookboon, 2010.

[27] S.E. Cohn, N.S. Sivakumaran, and R. Todling. A fixed-lag kalman smoother for retrospective data assimilation. *Monthly Weather Review*, 122(12):2838–2867, 1994.

[28] P. Courtier, J.N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.

[29] J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. In *Math-*

ematical Proceedings of the Cambridge Philosophical Society, volume 43, pages 50–67. Cambridge Univ Press, 1947.

[30] R. Daley. Estimating model-error covariances for application to atmospheric data assimilation. *Monthly weather review*, 120(8):1735–1746, 1992.

[31] R. Daley. *Atmospheric Data Analysis*. Number 2. Cambridge university press, 1993.

[32] R. Daley. Estimating observation error statistics for atmospheric data assimilation. In *Annales geophysicae*, volume 11, pages 634–647. Copernicus, 1993.

[33] D.P. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly weather review*, 123(4):1128–1145, 1995.

[34] D.P. Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343, 2005.

[35] D.P. Dee, S.M. Uppala, A.J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M.A. Balmaseda, G. Balsamo, P. Bauer, et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011.

[36] J.C. Derber. A variational continuous assimilation technique. *Monthly weather review*, 117(11):2437–2446, 1989.

[37] G. Desroziers, L. Berre, and B. Chapnik. Objective validation of data assimilation systems: diagnosing sub-optimality. In *Proceedings of ECMWF Workshop on diagnostics of data assimilation system performance*, pages 15–25, 2009.

[38] 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(613):3385–3396, 2005.

[39] G. Desroziers and S. Ivanov. Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 127(574):1433–1452, 2001.

[40] F.X. Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38(2):97–110, 1986.

[41] C.J. Donlon, M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer. The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, 116:140–158, 2012.

[42] M.A. Dubois and P. Yiou. Testing asynchronous coupling on simple "ocean-atmosphere" dynamic systems. *Climate dynamics*, 15(1):1–7, 1999.

[43] ECMWF. Operational implementation part II: Data assimilation, IFS documentation. Technical report, 2013.

[44] A. El-Said. *Conditioning of the Weak-Constraint Variational Data Assimilation Problem for Numerical Weather Prediction*. PhD thesis, University of Reading, 2015.

[45] G. Evensen. *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media, 2009.

[46] M. Fisher. Minimization algorithms for variational data assimilation. *Recent Developments in Numerical Methods for Atmospheric Modelling*, pages 364–385, 1998.

[47] M. Fisher. Background error covariance modelling. In *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, pages 45–63, 2003.

[48] M. Fisher and P. Courtier. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF*, 1995.

[49] W. Fu, J. She, and M. Dobrynin. A 20-year reanalysis experiment in the Baltic Sea using three-dimensional variational (3DVAR) method. *Ocean Science*, 8(5):827–844, 2012.

[50] L.S. Gandin and R. Hardin. Objective analysis of meteorological fields. *Jerusalem: Israel program for scientific translations*, 242, 1965.

[51] P. Gauthier, M. Tanguay, S. Laroche, S. Pellerin, and J. Morneau. Extension of 3DVAR to 4DVAR: Implementation of 4DVAR at the Meteorological Service of Canada. *Monthly weather review*, 135(6):2339–2354, 2007.

[52] A.E. Gill. *Atmosphere-ocean dynamics*, volume 30. Academic press, 1982.

[53] R.J. Graham, M.Gordon, P.J. McLean, S. Ineson, M.R. Huddleston, M.K. Davey, A. Brookshaw, and R.T.H. Barnes. A performance comparison of coupled and uncoupled versions of the Met Office seasonal prediction general circulation model. *Tellus A*, 57(3):320–339, 2005.

[54] A.K. Griffith and N.K. Nichols. Adjoint methods in data assimilation for estimating model error. *Flow, turbulence and combustion*, 65(3-4):469–488, 2000.

[55] K. Haines. Coupled atmosphere-ocean data assimilation. In *Seminar on Data assimilation for atmosphere and ocean*, pages 249–264. ECMWF, 2012.

[56] D. Hodyss and N. Nichols. The error of representation: basic understanding. *Tellus A*, 67, 2015.

[57] A. Hollingsworth and P. Lönnberg. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A*, 38(2):111–136, 1986.

[58] M.R. Huddleston, M.J. Bell, M.J. Martin, and N.K. Nichols. Assessment of wind-stress errors using bias corrected ocean data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 130(598):853–871, 2004.

[59] N.B. Ingleby. The statistical structure of forecast errors and its representation in the Met. Office global 3-D variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 127(571):209–231, 2001.

[60] M. Janisková, J.N. Thépaut, and J.F. Geleyn. Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Monthly weather review*, 127(1):26–45, 1999.

[61] T. Kadowaki. A 4-dimensional variational assimilation system for the JMA Global Spectrum Model. *CAS/JAC WGNE Research Activities in Atmospheric and Oceanic Modelling*, 34:1–17, 2005.

[62] R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[63] E. Kalnay. *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, 2003.

[64] D. Kondrashov, C. Sun, and M. Ghil. Data assimilation for a coupled ocean-atmosphere model. Part II: Parameter estimation. *Monthly Weather Review*, 136(12):5062–5076, 2008.

[65] W. Lahoz, B. Khattatov, and R. Menard. *Data assimilation: making sense of observations.* Springer Science & Business Media, 2010.

[66] P. Laloyaux, M. Balmaseda, D. Dee, K. Mogensen, and P. Janssen. A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 2015.

[67] V.E. Larson, D.P. Schanen, M. Wang, M. Ovchinnikov, and S. Ghan. PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km. *Monthly Weather Review*, 140(1):285–306, 2012.

[68] M. Latif, T. Stockdale, J. Wolff, G. Burgers, E. Maier-Reimer, M.M. Junge, K. Arpe, and L. Bengtsson. Climatology and variability in the ECHO coupled GCM. *Tellus A*, 46(4):351–366, 1994.

[69] A.S. Lawless. A note on the analysis error associated with 3D-FGAT. *Quarterly Journal of the Royal Meteorological Society*, 136(649):1094–1098, 2010.

[70] A.S. Lawless. Variational data assimilation for very large environmental problems. *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences*, 13:55, 2013.

[71] A.S. Lawless, S. Gratton, and N.K. Nichols. An investigation of incremental 4D-Var using non-tangent linear models. *Quarterly Journal of the Royal Meteorological Society*, 131(606):459–476, 2005.

[72] A.S. Lawless and N.K. Nichols. Inner-loop stopping criteria for incremental four-dimensional variational data assimilation. *Monthly Weather Review*, 134(11):3425–3435, 2006.

[73] D.J. Lea, J.P. Drecourt, K. Haines, and M.J. Martin. Ocean altimeter assimilation with observational and model bias correction. *Quarterly Journal of the Royal Meteorological Society*, 134(636):1761–1774, 2008.

[74] D.J. Lea, I. Mirouze, M.J. Martin, R.R. King, A. Hines, D. Walters, and M. Thurlow. Assessing a new coupled data assimilation system based on the Met Office coupled atmosphere–land–ocean–sea ice model. *Monthly Weather Review*, 143(11):4678–4694, 2015.

[75] D.J. Lea, I. Mirouze, M.J. Martin, A. Shelly, A. Hines, and P. Sykes. Assessment of a weakly coupled atmosphere-land-ocean-ice data assimilation system. Technical report, Met Office, 2013.

[76] J.M. Lewis and J.C. Derber. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus A*, 37(4):309–322, 1985.

[77] J.M. Lewis, S. Lakshmivarahan, and S. Dhall. Dynamic data assimilation: a least squares approach. *Cambridge University Press*, 13, 2006.

[78] Z. Li, Y. Chao, J.C. McWilliams, and K. Ide. A three-dimensional variational data assimilation scheme for the regional ocean modeling system. *Journal of Atmospheric and Oceanic Technology*, 25(11):2074–2090, 2008.

[79] Z. Li, S. Feng, Y. Liu, W. Lin, M. Zhang, T. Toto, A.M. Vogelmann, and S. Endo. Development of fine-resolution analyses and expanded large-scale forcing properties: 1. Methodology and evaluation. *Journal of Geophysical Research: Atmospheres*, 120(2):654–666, 2015.

[80] A.C. Lorenc. Analysis methods for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 112(474):1177–1194, 1986.

[81] E.N. Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.

[82] J. Lu and W.W. Hsieh. On determining initial conditions and parameters in a simple coupled atmosphere-ocean model by adjoint data assimilationl. *Tellus A*, 50(4):534–544, 1998.

[83] L. Magnusson, M.A. Balmaseda, S. Corti, F. Molteni, and T. Stockdale. Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, 41(9-10):2393–2409, 2013.

[84] M.J. Martin, M.J. Bell, and N.K. Nichols. *Treatment of systematic errors in sequential data assimilation*. Ocean Applications, Meteorological Office, 1999.

[85] M.J. Martin, A. Hines, and M.J. Bell. Data assimilation in the FOAM operational short-range ocean forecasting system: A description of the scheme and its impact. *Quarterly Journal of the Royal Meteorological Society*, 133(625):981–995, 2007.

[86] T. Mochizuki, H. Igarashi, N. Sugiura, S. Masuda, N. Ishida, and T. Awaji. Improved coupled GCM climatologies for summer monsoon onset studies over Southeast Asia. *Geophysical research letters*, 34(1), 2007.

[87] F. Molteni, L. Ferranti, T.N. Palmer, and P. Viterbo. A dynamical interpretation of the global response to equatorial Pacific SST anomalies. *Journal of climate*, 6(5):777–795, 1993.

[88] T. Müller and H. Müller. *Modelling in natural sciences: design, validation and case studies*. Springer Science & Business Media, 2013.

[89] J. Murphy, V. Kattsov, N. Keenlyside, M. Kimoto, G. Meehl, V. Mehta, H. Pohlmann, A. Scaife, and D. Smith. Towards prediction of decadal climate variability and change. *Procedia Environmental Sciences*, 1:287–304, 2010.

[90] I.M. Navon. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of atmospheres and oceans*, 27(1):55–79, 1998.

[91] N.K. Nichols. Data assimilation: Aims and basic concepts. In *Data Assimilation for the Earth System*, pages 9–20. Springer, 2003.

[92] N.K. Nichols. Treating model error in 3-D and 4-D data assimilation. In *Data Assimilation for the Earth System*, pages 127–135. Springer, 2003.

[93] P.R. Oke and P. Sakov. Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 25(6):1004–1017, 2008.

[94] P. Okely, O. Alves, D. Hudson, and Y. Yin. Towards coupled data assimilation: Coupled covariance structures. *CAWCR Research Letters*, (7):4–9, 2011.

[95] S.K. Park and L. Xu. *Data assimilation for atmospheric, oceanic and hydrologic applications*, volume 2. Springer Science & Business Media, 2013.

[96] A. Persson. User guide to ECMWF forecast products. 2001.

[97] E. Polak and G. Ribiere. Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle, série rouge*, 3(1):35–43, 1969.

[98] F. Rabier, H. Järvinen, E. Klinker, J.F. Mahfouf, and A. Simmons. The ECMWF operational implementation of four-dimensional variational assimi-

lation. I: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126(564):1143–1170, 2000.

[99] 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(623):347–362, 2007.

[100] S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, 2015.

[101] L. Rezende. *Chronology of science*. Infobase Publishing, 2006.

[102] S. Saha, S. Moorthi, H.L Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, et al. The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, 91(8):1015–1057, 2010.

[103] Y. Sasaki. Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, 98(12):875–883, 1970.

[104] R. Sausen, K. Barthel, and K. Hasselmann. Coupled ocean-atmosphere models with flux correction. *Climate Dynamics*, 2(3):145–163, 1988.

[105] G. Shutts. A stochastic convective backscatter scheme for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 141(692):2602–2616, 2015.

[106] D.M. Smith, A.A. Scaife, G.J. Boer, M. Caian, F.J. Doblas-Reyes, V. Guemas, E. Hawkins, W. Hazeleger, L. Hermanson, C.K. Ho, et al. Real-time multi-model decadal climate predictions. *Climate dynamics*, 41(11-12):2875–2888, 2013.

[107] A. Soloviev and R. Lukas. *The near-surface layer of the ocean: structure, dynamics and applications*, volume 48. Springer Science & Business Media, 2013.

[108] J.H. Steele, S.A. Thorpe, and K.K. Turekian. *Marine Ecological Processes: A derivative of the Encyclopedia of Ocean Sciences*. Academic Press, 2010.

[109] D.J. Stensrud. *Parameterization schemes: keys to understanding numerical weather prediction models*. Cambridge University Press, 2009.

[110] 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. *Quarterly Journal of the Royal Meteorological Society*, 140(681):1236–1244, 2014.

[111] R.H. Stewart. Introduction to physical oceanography. *Department of Oceanography, Texas A & M University*, 2005.

[112] D. Storkey, E. W. Blockley, R. Furner, C. Guiavarch, D. Lea, M.J. Martin, R. M. Barciela, A. Hines, P. Hyder, and J.R. Siddorn. Forecasting the ocean state using NEMO: The new FOAM system. *Journal of operational oceanography*, 3(1):3–15, 2010.

[113] A. Stuart and A.R. Humphries. *Dynamical systems and numerical analysis*, volume 2. Cambridge University Press, 1998.

[114] N. Sugiura, T. Awaji, S. Masuda, T. Mochizuki, T. Toyoda, T. Miyama, H. Igarashi, and Y. Ishikawa. Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations. *Journal of Geophysical Research: Oceans*, 113(C10), 2008.

[115] R. Todling. A complementary note to 'A lag-1 smoother approach to system-error estimation': the intrinsic limitations of residual diagnostics. *Quarterly Journal of the Royal Meteorological Society*, 141(692):2917–2922, 2015.

[116] R. Todling. A lag-1 smoother approach to system-error estimation: sequential method. *Quarterly Journal of the Royal Meteorological Society*, 141(690):1502–1513, 2015.

[117] Y. Trémolet. Accounting for an imperfect model in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132(621):2483–2504, 2006.

[118] Y. Trémolet. Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 133(626):1267–1280, 2007.

[119] E. Tziperman and W.C. Thacker. An optimal-control/adjoint-equations approach to studying the oceanic general circulation. *Journal of Physical Oceanography*, 19(10):1471–1485, 1989.

[120] B. Vannière, E. Guilyardi, T. Toniazzo, G. Madec, and S. Woolnough. A systematic approach to identify the sources of tropical SST errors in coupled models using the adjustment of initialised experiments. *Climate Dynamics*, 43(7-8):2261–2282, 2014.

[121] D.N. Walters, K.D. Williams, I.A. Boutle, A.C. Bushell, J.M. Edwards, P.R. Field, A.P. Lock, C.J. Morcrette, R.A. Stratton, J.M. Wilkinson, et al. The Met Office unified model global atmosphere 4.0 and JULES global land 4.0 configurations. *Geoscientific Model Development*, 7(1):361–386, 2014.

[122] T.T. Warner. *Numerical Weather and Climate prediction.* Cambridge University Press, 2010.

[123] J. Waters, D.J. Lea, M.J. Martin, D. Storkey, and J. While. Describing the development of the new FOAM-NEMOVar system in the global 1/4 degree configuration. Technical report, 578, Met Office, 2013.

[124] A.T. Weaver, J. Vialard, and D.L.T. Anderson. Three-and four-dimensional variational assimilation with a general circulation model of the tropical pacific ocean. Part I: Formulation, internal diagnostics, and consistency checks. *Monthly Weather Review*, 131(7):1360–1378, 2003.

[125] R.O. Weber and P. Talkner. Some remarks on spatial correlation function models. *Monthly Weather Review*, 121(9):2611–2617, 1993.

[126] P. Weston. Progress towards the implementation of correlated observation errors in 4D-Var. Technical report, 560, Met Office, 2011.

[127] P.P. Weston, W. Bel, and J.R. Eyre. Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429, 2014.

[128] P.W. White. *IFS Documentation: Part III: Dynamics and Numerical Procedures (CY21R4)*. European Centre for Medium-Range Weather Forecasts, 2000.

[129] A.T.A. Wood. When is a truncated covariance function on the line a covariance function on the circle? *Statistics & probability letters*, 24(2):157–164, 1995.

[130] S. Zhang. Impact of observation-optimized model parameters on decadal predictions: Simulation with a simple pycnocline prediction model. *Geophysical Research Letters*, 38(2), 2011.

[131] S. Zhang. A study of impacts of coupled model initial shocks and state-parameter optimization on climate predictions using a simple pycnocline prediction model. *Journal of Climate*, 24(23), 2011.

[132] J. Zhu, G. Zhou, C. Yan, W. Fu, and X. You. A three-dimensional variational ocean data assimilation system: Scheme and preliminary results. *Science in China Series D: Earth Sciences*, 49(11):1212–1222, 2006.