THE UNIVERSITY OF READING
DEPARTMENT OF MATHEMATICS

Data Assimilation in Ocean Circulation Models with Systematic Errors

by

Matthew J. Martin

Thesis submitted for the degree of
Doctor of Philosophy

December 2000

Abstract

Data assimilation is a means of estimating the true state of a dynamical system by combining the output from a numerical model with the available observations. Most data assimilation methods assume that the model error is serially uncorrelated. We describe a method for taking systematic model errors into account by augmenting the state vector with a vector of model error variables. The stability and convergence of this augmented state are discussed with respect to a general sequential data assimilation method.

Systematic model error correction is examined in a simple model of the tropical Pacific ocean in the context of the two-dimensional Shallow Water equations. A new method is proposed which can account for systematic errors in the wind forcing of the model by correcting the pressure field, using differences between the observations and the model fields. An analysis of this pressure correction method shows that the pressure and vertical velocities will converge with time to their true values.

We apply the pressure correction method to a three-dimensional Primitive equation model of the ocean. Experiments performed with the operational system used at the Met. Office, FOAM, show a much improved balance at the end of a two year integration with the pressure correction method included in comparison to the original data assimilation scheme. These results show that the method produces better initial conditions for a forecast. Also, diagnostics from the method may help to indicate how to improve the model.

Acknowledgements

I would like to thank my academic supervisor, Prof. Nancy Nichols, for her invaluable help and encouragement during my research. I also thank my industrial supervisor, Dr. Mike Bell of the Met. Office, for the interest he has shown and the assistance he has given me, especially on the work at the Met. Office.

I would like to thank my family and friends, both inside and outside of the University, for helping me to enjoy my time as a PhD student.

Finally, I acknowledge financial support from the Natural Environment Research Council and from the Met. Office.

Contents

Chapter 1

Introduction

Data assimilation is a means of estimating the true state of a system by combining observations with a numerical model of a dynamical system. These techniques provide initial conditions for forecasts and can also be used for diagnosing past events and for model verification. However, systematic errors in the numerical model can cause the data assimilation to produce spurious results which will affect our ability to obtain good initial conditions for forecasts. An aim of this thesis is to investigate methods for assimilation of observations into models which contain systematic errors.

A motivation for the work in this thesis is an example of where systematic model errors can significantly affect our ability to estimate the state of a system using data assimilation. An example is presented in Figure 1.0.1 which shows large vertical velocities in the tropical Pacific ocean due to the assimilation of temperature observations into a biased model of the ocean. This region is shown in the example because the problem of model biases is worst here. The tropical Pacific is an important area because a large

fraction of the variation in the energy of the world's oceans occurs in this region. Also, exchanges between the atmosphere and ocean in the tropical Pacific are important for seasonal prediction and climate research. These interactions are especially important when predicting El Niño, a period of anomalously warm sea surface temperature (SST) in the eastern equatorial Pacific associated with ceasing trade winds in the atmosphere.

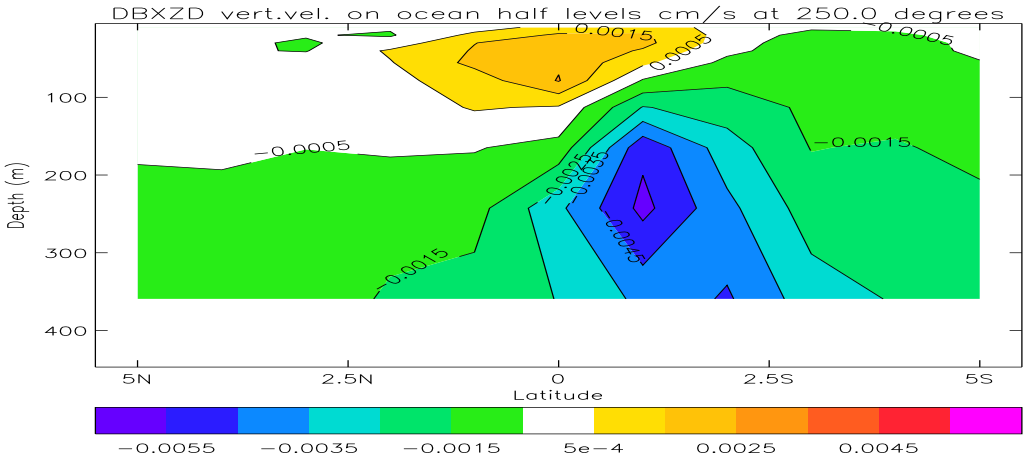

Figure 1.0.1: Spurious vertical velocities (cm/s) associated with assimilation of temperature observations into a biased ocean model.

Many geophysical modelling centres around the world have computer models of the ocean and use these to simulate the state of the ocean, i.e. currents, temperature, salinity and density. In Chapter 2, we describe the basic equations used in these models. The types of observations which are available for use in the estimation of the state of the ocean are discussed. We also give some background information on the types of flow expected in the tropical Pacific and the problems associated with predicting them. One

of the main topics of interest is the way in which the atmosphere forces the ocean circulation in equatorial oceans, so some background on this topic is given.

When attempting to predict the state of the ocean and/or atmosphere, an accurate estimate of the initial conditions is essential. To obtain the best possible estimate, the output from a numerical model should be combined with the available observations. This data assimilation can be done in many different ways. The more common methods are very similar to the observers of Control Theory. In Chapter 3, we describe some relevant results from the Control Theory literature and show what they imply about the convergence and stability of sequential data assimilation methods.

In Chapter 4, the objective function used in most of the sequential data assimilation methods is derived in a statistical framework. The more common ways to estimate the solution of this objective function are also described, together with other popular sequential data assimilation methods. Some examples of how these methods perform are shown using two simple ordinary differential equation (ODE) models. An overview of the four-dimensional data assimilation methods is also given for completeness.

A problem with these methods is that it is assumed that no systematic errors are present in the model or observations. Observational bias is not usually a problem as it is reasonably easy to detect and correct. In ocean models however, there are often systematic errors, mainly due to the number of parameterisations required. A general method for accounting for systematic model error is given in Chapter 5. Some of the important advantages and disadvantages of this bias correction are illustrated using some simple

examples. These include experiments performed with the two ODE models introduced in Chapter 4, and also experiments with a simple partial differential equation (PDE) model.

In Chapter 6, we begin to explore how the bias correction method accounts for a specific type of model error. This is done using a simplified model of the equatorial Pacific ocean. Using the results of some analysis of the linear two-dimensional shallow water equations, a modification of the bias correction method is proposed. Some experiments are described which compare the results of using normal data assimilation to the results obtained when the bias correction method and the new pressure correction method are used.

To test the effectiveness of the pressure correction method in a full primitive equation global ocean model, we implement it in the Forecasting Ocean Atmosphere Model (FOAM) used operationally at the Met. Office. These experiments are described in Chapter 7. Results are presented which show improvements in the balance of the system in the tropical Pacific and hence in the estimate of the state of the ocean there.

In Chapter 8, some conclusions are given, together with some possible extensions of the work.

Chapter 2

Oceanography

2.1 Introduction

As we are dealing with data assimilation in ocean circulation models, we devote a chapter to describing the equations of motion, the numerical solution of these equations, the observations available to oceanographers and the types of circulation we will be dealing with later in the thesis.

The main area of interest for the purposes of this study is the prediction of the state of the equatorial Pacific over seasonal time scales. This involves using an ocean circulation model, with forcing from the atmosphere coming from wind stresses, evaporation and precipitation. We will begin by describing the approximations made to the Navier-Stokes equations to obtain the so-called Primitive equations. The discretisations used in a numerical model will then be discussed. The types of observations which are available for use by oceanographers will also be described. To examine the types of motion we expect to find in the tropical Pacific ocean, the linear shallow water equations

are derived and some wave solutions are found on the β-plane. The effects of wind forcing are then described.

2.2 Ocean general circulation models

2.2.1 Equations of motion

The equations of motion for large-scale ocean dynamics are obtained from the Navier-Stokes equations, [19]. A number of approximations are made however which simplify these equations. The first of these is the Boussinesq approximation, where the density is taken to be constant in computing rates of change of momentum, and density variations are taken into account only when they give rise to buoyancy forces, [13]. This is a valid approximation since changes in the density of the ocean with time are much smaller than the average magnitudes of density. As the velocity of a fluid element and the phase speed of disturbances are much less than the speed of sound, we can assume that the flow is incompressible. This means that changes in density with pressure are negligible. We also assume the turbulent mixing hypothesis, where stresses exerted by scales of motion too small to be resolved by the computational grid are represented as an enhanced molecular mixing.

The momentum equations are then given by

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} + 2(\Omega_y w - \Omega_z v) = -\rho_0^{-1}\frac{\partial p}{\partial x} + F^u, \quad (2.2.1)$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} + 2(\Omega_z u - \Omega_x w) = -\rho_0^{-1}\frac{\partial p}{\partial y} + F^v, \quad (2.2.2)$$

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + v\frac{\partial w}{\partial y} + w\frac{\partial w}{\partial z} + 2(\Omega_x v - \Omega_y u) = -\rho_0^{-1}\frac{\partial p}{\partial z} - \frac{\rho}{\rho_0}g, \quad (2.2.3)$$

where x, y and z are the local cartesian coordinates, p is the pressure of a fluid element, ρ is the density of a fluid element, $\underline{u} = (u, v, w)$ is the velocity vector, $\underline{\Omega} = (\Omega_x, \Omega_y, \Omega_z)$ is the angular momentum, g is the acceleration due to gravity and F^u and F^v represent forcing terms. The angular momentum of the earth is taken to be $\underline{\Omega} = (0, 0, \frac{1}{2}f)$, where $f = 2\Omega sin\phi$ is the Coriolis parameter, Ω is the earth's angular velocity and ϕ is the latitude.

A scale analysis shows that, in the ocean, local acceleration and terms of equal order can be eliminated from the equation of vertical momentum. This is a valid approximation if the vertical scale of motion is much less than the scale height, which is always true in the ocean, [13]. This results in the hydrostatic equation

$$\frac{\partial p}{\partial z} = -\rho g. \tag{2.2.4}$$

The continuity equation

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{2.2.5}$$

shows that the velocity field is nondivergent. Conservative tracers such as the potential temperature, θ and the salinity S satisfy the advection equations

$$\frac{\partial \theta}{\partial t} + u\frac{\partial \theta}{\partial x} + v\frac{\partial \theta}{\partial y} + w\frac{\partial \theta}{\partial z} = 0, \tag{2.2.6}$$

$$\frac{\partial S}{\partial t} + u\frac{\partial S}{\partial x} + v\frac{\partial S}{\partial y} + w\frac{\partial S}{\partial z} = 0. \tag{2.2.7}$$

The last equation needed is the equation of state which relates salinity, potential temperature and pressure to the density, i.e.

$$\rho = \rho(p, S, \theta). \tag{2.2.8}$$

We therefore have seven equations (2.2.1), (2.2.2), (2.2.4), (2.2.5), (2.2.6), (2.2.7) and (2.2.8), in seven unknowns $\underline{u} = (u, v, w)$, p, ρ, θ and S.

2.2.2 Numerical solution

There are many possibilities for the representation of the oceanic circulation using computers. Quasi-geostrophic models, which discretise a simplified version of the equations presented in the previous section, have been used to model the mesoscale eddy fields in the ocean, [43]. However, most of the major oceanographic institutions now use primitive equation models which discretise the equations in the previous section. The most common of these models is the Bryan-Cox model, [9], [10], [19], which uses a finite difference approximation to the equations. More detail on this model is given in Section 7.2, as it is used in the Forecasting Ocean Atmosphere Model (FOAM) system at the Met. Office. Other methods for the discretisation of the equations include more sophisticated finite difference schemes, finite element methods and spectral models, although these are rarely used in operational models at present, [51].

2.3 Observations

As well as having a computer model, an important source of information on the state of the ocean is from observations. These observations are available from both in situ and satellite instruments. The number of observations available to oceanographers has increased dramatically over the last decade due to the introduction of satellite observing instruments. There are also plans to increase the number of in situ data through the deployment of a large number of drifters.

The in situ instruments which provide information about the state of the

ocean come from various sources. Sea surface temperature (SST) data is routinely available and is observed from ships, drifting buoys and moored buoys. Observations of the temperature structure beneath the surface are more sparse. These come mainly from expendable bathythermographs (XBTs) which measure the temperature structure of the top few hundred metres of the ocean, together with conductivity temperature and depth (CTD) instruments. Some information is also available about the ocean currents from a number of buoys and also ships, which use acoustic Doppler current profilers (ADCPs) to measure velocities. The salinity of the oceans is not currently well observed but some information is available and there are plans to increase the number of these observations.

An important source of observations of the tropical Pacific ocean is the tropical atmosphere ocean (TAO) array which is a number of moored buoys spread over this region, [41]. There are nearly 70 buoys in this array which measure winds, SST, relative humidity, air temperature and subsurface temperature down to about 500m. Some of these buoys also measure the velocity of the ocean.

Various satellites are in orbit with instruments which observe the state of the ocean. These include instruments which give information about the height of the sea surface and also instruments which measure SST. The sea surface height (SSH) data comes from altimeters which are accurate to about 4cm rms. SSTs are available from advanced very high resolution radiometers (AVHRR) which have an accuracy of about 0.5^0C and give the information at a high resolution of about 8km.

There are plans to significantly increase the number of drifters as part of

the ARGO proposal. Here the aim is to deploy approximately 3000 floats globally by 2003. These floats will stay at depth for most of the time but periodically (about every 10 days) rise to the surface, measuring temperature and salinity profiles. It is hoped that these floats will be spread evenly over the global ocean which will provide very useful information about the dynamics of the ocean.

2.4 Tropical dynamics

The rotation of the earth has an important effect on the way the ocean responds to imposed changes. The Coriolis parameter, $f = 2\Omega sin\phi$, depends on the latitude ϕ, and is close to zero in the tropics. This has important consequences on the ocean circulation near the equator, as will be discussed in this section.

To study the type of motions we would expect to see near the equator, we will examine the linear shallow water equations on an equatorial β-plane. These can be derived from the horizontal momentum equations (2.2.1-2.2.2), the hydrostatic equation (2.2.4), the continuity equation (2.2.5) and a mass conservation equation based on equations (2.2.6)-(2.2.8), given by

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z} = 0. \tag{2.4.1}$$

If we linearise these about a vertically stratified state of rest $\rho_0(z)$ where $u_0 \equiv 0$, $v_0 \equiv 0$, $w_0 \equiv 0$, we obtain

$$\rho_0(\frac{\partial u}{\partial t} - fv) = -\frac{\partial p}{\partial x}, \tag{2.4.2}$$

$$\rho_0(\frac{\partial v}{\partial t} + fu) = -\frac{\partial p}{\partial y}, \tag{2.4.3}$$

$$\frac{\partial p}{\partial z} = -\rho g, \tag{2.4.4}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{2.4.5}$$

$$\frac{\partial \rho}{\partial t} + w\frac{d\rho_0}{dz} = 0. \tag{2.4.6}$$

Since these equations are linear, we can split the solution into a depth dependent part and a part which varies in the horizontal and with time, so that

$$u(x,y,z,t) = \frac{A(z)}{g\rho_0}\hat{u}(x,y,t),\ v(x,y,z,t) = \frac{A(z)}{g\rho_0}\hat{v}(x,y,t), \tag{2.4.7}$$

$$p(x,y,z,t) = A(z)\hat{p}(x,y,t),\ w(x,y,z,t) = B(z)\frac{\partial \hat{p}}{\partial t}(x,y,t). \tag{2.4.8}$$

The horizontal structure then satisfies the equations

$$\frac{\partial \hat{u}}{\partial t} - f\hat{v} = -g\frac{\partial \hat{p}}{\partial x}, \tag{2.4.9}$$

$$\frac{\partial \hat{v}}{\partial t} + f\hat{u} = -g\frac{\partial \hat{p}}{\partial y}, \tag{2.4.10}$$

$$\frac{\partial \hat{p}}{\partial t} + H_e\left(\frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y}\right) = 0, \tag{2.4.11}$$

and the vertical structure is given by

$$\frac{d}{dz}\left(\frac{dB}{dz}\rho_0\right)H_e - B\frac{d\rho_0}{dz} = 0, \tag{2.4.12}$$

where H_e is the separation constant, often referred to as the equivalent depth.

The equatorial β-plane approximation says that the Coriolis parameter can be written as some constant, β, multiplied by the northward distance away from the equator, i.e. $f = \beta y$, where $\beta = 2.3 \times 10^{-11} m^{-1} s^{-1}$. With this approximation, we can write equations (2.4.9-2.4.11) as one equation in terms of $\hat{v}$,

$$\frac{\partial}{\partial t}\left\{\frac{1}{c^2}\left(\frac{\partial^2 \hat{v}}{\partial t^2} + f^2\hat{v}\right) - \left(\frac{\partial^2 \hat{v}}{\partial x^2} + \frac{\partial^2 \hat{v}}{\partial y^2}\right)\right\} - \beta\frac{\partial \hat{v}}{\partial x} = 0, \tag{2.4.13}$$

where $c^2 = gH_e$.

2.4.1 Equatorial waves

We look for equatorially trapped waves by finding solutions of equation (2.4.13) of the form $\hat{v}(x,y,t) = v(y)exp(ikx - i\omega t)$. This yields the ordinary differential equation

$$\frac{d^2v}{dy^2} + (\frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega} - \frac{\beta^2 y^2}{c^2})v = 0. \tag{2.4.14}$$

If we transform variable to $\zeta = (\beta/c)^{1/2}y$, equation (2.4.14) becomes

$$\frac{d^2v}{d\zeta^2} - \zeta^2 v = \frac{c}{\beta}(\frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega})v. \tag{2.4.15}$$

This has solutions of the form $v = v_0 exp(-\zeta^2/2)H_n(\zeta)$, $n = 0, 1, 2, \ldots$, where H_n is an nth order Hermite polynomial, as described in Appendix B. These solutions vanish as $y \to \pm\infty$ which is what we require for equatorially trapped waves. The dispersion relation is then given by

$$\frac{\omega^2}{c^2} - k^2 - \frac{\beta k}{\omega} = \frac{\beta}{c}(2n+1), \qquad n = 0, 1, 2, \ldots. \tag{2.4.16}$$

There are three main types of waves arising from this expression, [51]:

1. **Planetary waves**

   For these waves, w^2/c^2 is negligible, giving $\omega = -\beta k/(k^2 + \frac{\beta}{c}(2n+1))$. Long planetary waves are fairly slowly propagating westward waves, the first baroclinic mode having a wave speed of approximately $0.9ms^{-1}$. For shorter waves, the group velocity is eastward but they only carry information at an eighth of the speed at which the long waves carry information westwards.

2. **Gravity waves**

   Here, $\beta k/\omega$ is negligible, giving $\omega^2 = k^2c^2 + \beta c(2n+1)$. These waves can propagate energy eastward or westward.

3. **Kelvin waves**

   The Kelvin wave satisfies the dispersion relation (2.4.16) when $n = -1$, but can also be obtained by setting $v = 0$ in the shallow water equations. It has dispersion relation $\omega = kc$. This wave therefore propagates eastward along the equator without dispersion at speed c. For the first baroclinic mode, c is approximately $2.8ms^{-1}$.

For more information on these waves, see [35].

2.4.2 Wind forcing

If the wind forcing is included in the nonlinear model, the linear shallow water equations (2.4.9-2.4.11) become

$$\frac{\partial \hat{u}}{\partial t} - \beta y\hat{v} = -g\frac{\partial \hat{p}}{\partial x} + \tau^x, \tag{2.4.17}$$

$$\frac{\partial \hat{v}}{\partial t} + \beta y\hat{u} = -g\frac{\partial \hat{p}}{\partial y} + \tau^y, \tag{2.4.18}$$

$$\frac{\partial \hat{p}}{\partial t} + H_e(\frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y}) = 0, \tag{2.4.19}$$

where $\underline{\tau} = (\tau^x, \tau^y)$ is the wind stress. We can then write this as one equation in $\hat{v}$ to obtain

$$\frac{\partial}{\partial t}\{\frac{1}{c^2}(\frac{\partial^2 \hat{v}}{\partial t^2} + f^2\hat{v}) + (\frac{\partial^2 \hat{v}}{\partial x^2} + \frac{\partial^2 \hat{v}}{\partial y^2})\} - \beta\frac{\partial \hat{v}}{\partial x} = \frac{1}{c^2}\frac{\partial}{\partial t}(\frac{\partial \tau^y}{\partial t} - \beta y\tau^x) - \frac{\partial}{\partial x}(\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y}). \tag{2.4.20}$$

The response to a uniform wind, parallel to the equator, should be independent of x. In this case, the steady state equation for $\hat{v}$ is

$$-\beta y \hat{v} = \tau^x. \tag{2.4.21}$$

Easterly winds ($\tau^x < 0$) will therefore produce transport away from the equator to the North and South. This implies that upwelling must occur at the equator. This upwelling raises the thermocline, which is a region of strong vertical temperature gradient at the bottom of the mixed layer, and a current develops in the direction of the wind. Boundary effects then become important because of the east-west pressure gradient associated with the eastward propagation of the Kelvin waves. There is a westward current at the surface but an eastward undercurrent in the thermocline, [81]. This current is called the Equatorial Undercurrent (EUC) and is an important feature of the equatorial ocean circulation, [82].

This discussion is meant to give a broad outline of the response of the tropical ocean to wind forcing. It is also meant to indicate the complexity of this response to a given forcing. In [48] the large number of currents which are present in the tropical Pacific are described and [36] shows that it is in fact very difficult to estimate the wind stresses themselves very accurately as there is a large amount of time and space variability in the tropical Pacific. The inaccuracy in the supplied winds together with the complexity of the response to these winds makes the prediction, and even the estimation, of flow in the tropical Pacific a hard task.

2.5 Conclusions

This chapter has given a brief description of the equations of motion which are used in computer models of the global ocean. These equations contain a number of approximations which, whilst being justifiable, will reduce the accuracy of these models. Of more significance is the resolution of the model grid, which in the Met. Office's FOAM model is $1^0 \times 1^0$ in the horizontal. The problem with having such a coarse resolution is that important mesoscale motions are not resolved by the grid. These therefore have to be parameterised which reduces the accuracy of the model.

An outline of the dynamics in the tropical oceans has been given. The wind stresses are an important driving force in this region. The way these stresses are distributed in the vertical in an ocean model has to be parameterised and it is thought, at least in the FOAM model, that this parameterisation is not always accurate. This fact, together with the idea that the supplied surface wind stresses are not always accurate, leads to uncertainties in the resulting circulation.

When attempting to predict the state of the tropical Pacific ocean over seasonal time scales, especially in the presence of El Niño, the coupling between the atmosphere and ocean is very important. The manner in which the ocean responds to atmospheric forcing is therefore also important. An advantage of the tropical Pacific is that we have a fairly good observational coverage, [12]. The TAO array is a number of fixed buoys which take observations of temperature with depth, [41]. There are also observations of sea surface temperature (SST) and temperature profiles made available from ships. These observations should therefore be used to keep the model close

to the true state of the ocean, as will be discussed in the next few chapters.

Chapter 3

Control Theory and Data Assimilation

3.1 Introduction

Systems such as the ocean and atmosphere are complicated, infinite dimensional systems. Much work has been done on understanding the underlying physical laws which govern these fluids. The equations that are thought to describe the evolution of the ocean state, and the methods used to approximate these equations so as to enable computational modelling of the ocean, are described in the previous chapter. These approximations, both in the equations themselves and in the numerical scheme used to solve them, will necessarily produce uncertainties in the model that we use to forecast the ocean state, [33]. We do, however, have observations that we can use to constrain the model. We would therefore like to combine the available observations with our model to produce the best initial conditions for a forecast.

In Section 3.2, we introduce some statistical concepts and notation which we will use throughout this thesis. We then introduce the general data assimilation problem in Section 3.3, together with some of the assumptions which are usually made when trying to solve the problem. In Section 3.4, we investigate the solution of the linear data assimilation problem and give conditions on the stability and convergence of the data assimilation process using some ideas and theorems from the Control Theory literature. The extension of these ideas to the nonlinear case is investigated in Section 3.5 with conclusions given in Section 3.6.

3.2 Statistical concepts

We begin by defining some statistical concepts which we will use in this and subsequent chapters.

Definition 3.1 *The* **expected value** *(or* **mean***) of a random variable x, denoted $E\{x\}$, which has probability density function (PDF) p_x, is defined as*

$$E\{x\} = \int_{-\infty}^{\infty} x p_x dx. \tag{3.2.1}$$

The expected value of a random n-vector $\underline{x} \in \mathbb{R}^n$, where $\underline{x} = [x_1, x_2, \ldots, x_n]^T$, is defined as

$$E\{\underline{x}\} = \begin{bmatrix} E\{x_1\} \\ E\{x_2\} \\ \vdots \\ E\{x_n\} \end{bmatrix}. \tag{3.2.2}$$

Definition 3.2 *The* **variance** *of a random variable x is defined as*

$$Var\{x\} = E\{(x - E\{x\})^2\}. \tag{3.2.3}$$

Definition 3.3 *The* **covariance** *of two random variables x and y is defined as*

$$Cov\{x, y\} = E\{(x - E\{x\})(y - E\{y\})\}. \tag{3.2.4}$$

The **covariance matrix** *of a random n-vector $\underline{x}$, is written as*

$$\begin{aligned} Cov\{\underline{x}\} &= Cov\{\underline{x}, \underline{x}\} \\ &= \begin{bmatrix} Var\{x_1\} & Cov\{x_1, x_2\} & \dots & Cov\{x_1, x_n\} \\ Cov\{x_1, x_2\} & Var\{x_2\} & \dots & Cov\{x_2, x_n\} \\ \vdots & \vdots & \ddots & \vdots \\ Cov\{x_1, x_n\} & Cov\{x_2, x_n\} & \dots & Var\{x_n\} \end{bmatrix} \end{aligned} \tag{3.2.5}$$

which is symmetric.

Definition 3.4 *The* **correlation** *between two random variables x and y is defined as*

$$Cor\{x, y\} = E\{xy\}. \tag{3.2.6}$$

The **correlation coefficient** *of x and y is defined as*

$$\chi(x, y) = \frac{Cov\{x, y\}}{(Var\{x\}Var\{y\})^{1/2}}. \tag{3.2.7}$$

Definition 3.5 *The* **Gaussian** *(or* **Normal***) PDF for a random n-vector $\underline{x}$ with nonsingular covariance matrix is*

$$\begin{aligned} p_{\underline{x}}(\underline{x}) &= \left[(2\pi)^{n/2}(det(Cov\{\underline{x}\}))^{1/2}\right]^{-1} \\ &\quad exp\left[-1/2(\underline{x} - E\{\underline{x}\})^T Cov^{-1}\{\underline{x}\}(\underline{x} - E\{\underline{x}\})\right], \end{aligned} \tag{3.2.8}$$

where $Cov^{-1}\{\underline{x}\}$ denotes the matrix inverse of the covariance matrix.

3.3 The data assimilation problem

In this section, we state the general data assimilation problem, together with the assumptions that are usually made to enable its solution.

3.3.1 True system equation

We assume that the true state of the system to be studied can be represented on a model grid by the solution to the stochastic vector difference equation

$$\underline{x}_{k+1}^t = \underline{f}_k(\underline{x}_k^t, \underline{u}_k^t) + \underline{\zeta}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{3.3.1}$$

with given initial conditions, $E\{\underline{x}_0^t\}$. Here, $\underline{x}_k^t \in \mathbb{R}^n$ is the true state vector at time t_k, $\underline{u}_k^t \in \mathbb{R}^m$ is the true input vector, $\underline{f}_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is assumed to be continuously differentiable and $\underline{\zeta}_k^t \in \mathbb{R}^n$ is a vector of random disturbances. We suppose that $\underline{\zeta}_k^t$ is a white Gaussian sequence.

If the random disturbances $\{\underline{\zeta}_k^t, k = 0, 1, \ldots, N-1\}$ are absent from equation (3.3.1), then we have a deterministic system with solution $\underline{x}_k^t, k = 0, 1, \ldots, N$. When the random disturbances are present, however, we are interested in the PDF of $\underline{x}_k^t$, from which $E\{\underline{x}_k^t\}$, $Cov\{\underline{x}_k^t\}$ and other useful information can be obtained. We therefore require $Cov\{\underline{x}_0^t\}$ to be given to determine the PDF at the initial and subsequent times. The vector of random disturbances is included in equation (3.3.1) due to the fact that, in the ocean context, we are representing a continuous, infinite dimensional system on a discrete, finite model grid.

3.3.2 Model system equations

We now write the model equations which consist of a stochastic vector difference equation approximating the true system equation, given by

$$\underline{x}_{k+1} = \underline{g}_k(\underline{x}_k, \underline{u}_k) + \underline{\zeta}_k, \qquad k = 0, 1, \ldots, N-1, \tag{3.3.2}$$

a set of observations given by

$$\underline{y}_k = \underline{h}_k(\underline{x}_k^t) + \underline{\delta}_k, \qquad k = 0, 1, \ldots, N, \tag{3.3.3}$$

and initial conditions, $E\{\underline{x}_0\}$. We also require $Cov\{\underline{x}_0\}$ to be given. Here, $\underline{x}_k \in \mathbb{R}^n$ is the model state vector at time t_k, $\underline{u}_k \in \mathbb{R}^m$ is the model input vector, $\underline{g}_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is assumed to be continuously differentiable, $\underline{\zeta}_k \in \mathbb{R}^n$ is a vector of random model disturbances, $\underline{y}_k \in \mathbb{R}^{p_k}$ is the observation vector at time t_k, $\underline{h}_k : \mathbb{R}^n \to \mathbb{R}^{p_k}$ is the observation operator which is assumed to be continuously differentiable and $\underline{\delta}_k \in \mathbb{R}^{p_k}$ is a vector of random observation errors. We suppose that $\{\underline{\zeta}_k, k = 0, 1, \ldots, N-1\}$ and $\{\underline{\delta}_k, k = 0, 1, \ldots, N\}$ are white Gaussian sequences.

The observation errors, $\underline{\delta}_k$, include contributions from the error in the observations themselves, but also from errors arising from the interpolation from the model grid to the observational grid and from conversion from state to observed variables. The following assumptions are made about the errors in the model system equations:

1.

$$E\{\underline{\zeta}_k\} = 0, \quad \forall t_k, \tag{3.3.4}$$

$$Cov\{\underline{\zeta}_k, \underline{\zeta}_j\} = \begin{cases} 0, & \text{if } t_j \neq t_k, \\ Q_k, & \text{if } t_j = t_k, \end{cases} \tag{3.3.5}$$

where $Q_k \in \mathbb{R}^{n \times n}$ is the known non-singular error covariance matrix of the model at time t_k.

2.

$$E\{\underline{\delta}_k\} = 0, \quad \forall t_k, \tag{3.3.6}$$

$$Cov\{\underline{\delta}_k, \underline{\delta}_j\} = \begin{cases} 0, & \text{if } t_j \neq t_k, \\ R_k, & \text{if } t_j = t_k, \end{cases} \tag{3.3.7}$$

where $R_k \in \mathbb{R}^{p_k \times p_k}$ is the known non-singular error covariance matrix of the observations at time t_k.

3. $Cov\{\underline{\zeta}_k, \underline{\delta}_k\} = 0, \quad \forall t_k, t_j.$

4. $Cov\{\underline{\zeta}_k, \underline{x}_0\} = Cov\{\underline{\delta}_k, \underline{x}_0\} = 0, \quad \forall t_k.$

We now state the general data assimilation problem.

General data assimilation problem

Given the model system equations, (3.3.2), (3.3.3), with initial conditions $E\{\underline{x}_0\}$, produce some estimate of the true state of the system at the final time, t_N, which is determined by the solution to the true system equation, (3.3.1), with true initial conditions, $E\{\underline{x}_0^t\}$.

The model system equations include an operator $\underline{g}_k$ which is different from the operator $\underline{f}_k$ used in the true system equations. This is included due to the fact that we will not usually know the exact representation of the true state of the system on the model grid, and approximations to the true

operator are often made when modelling systems such as the ocean. The fact that these two operators are different makes it difficult to say anything about the solution to the data assimilation problem. Also, the inputs to the model system equations, $\underline{u}_k$, are not the true inputs $\underline{u}_k^t$ as these are not always known accurately. To make the data assimilation problem more tractable, it is often assumed that the operator $\underline{f}_k$ and the true inputs $\underline{u}_k^t$ are used in the model system equations. In this case, the model errors $\underline{\zeta}_k$ are still assumed to form a white Gaussian sequence. This case, which we call the correct stochastic model case, is summarised below.

Definition 3.6 *Correct Stochastic Model Case*

1. *The model system equations are given by*

$$\underline{x}_{k+1} = \underline{f}_k(\underline{x}_k, \underline{u}_k^t) + \underline{\zeta}_k, \qquad k = 0, 1, \ldots, N-1, \tag{3.3.8}$$

$$\underline{y}_k = \underline{h}_k(\underline{x}_k^t) + \underline{\delta}_k, \qquad k = 0, 1, \ldots, N, \tag{3.3.9}$$

   with $E\{\underline{x}_0\}$ *given.*

2. *The assumptions about the statistics of* $\underline{\zeta}_k$, $\underline{\delta}_k$ *and* $\underline{x}_0$ *are the same as for the previous case.*

With the correct stochastic model case, the data assimilation problem involves the estimation of the PDF of $\underline{x}_N$ from which we can extract information about the expected value and covariance of the state at the final time. We now make the further assumption that the stochastic forcing is omitted from the equation evolving the state variables in the model system. Observational error is still included in the model system. We call this the perfect deterministic model case which is summarised below.

Definition 3.7 *Perfect Deterministic Model Case*

1. *The model system equations are given by*

$$\underline{x}_{k+1} = \underline{f}_k(\underline{x}_k, \underline{u}_k^t), \qquad k = 0, 1, \ldots, N-1, \tag{3.3.10}$$

$$\underline{y}_k = \underline{h}_k(\underline{x}_k^t) + \underline{\delta}_k, \qquad k = 0, 1, \ldots, N, \tag{3.3.11}$$

with $\underline{x}_0$ *given.*

2.

$$E\{\underline{\delta}_k\} = 0, \qquad \forall t_k, \tag{3.3.12}$$

$$Cov\{\underline{\delta}_k, \underline{\delta}_j\} = \begin{cases} 0, & if\, t_j \neq t_k, \\ R_k, & if\, t_j = t_k, \end{cases} \tag{3.3.13}$$

$$Cov\{\underline{\delta}_k, \underline{x}_0\} = 0, \qquad \forall t_k. \tag{3.3.14}$$

Note that we do not know the true initial conditions in either the correct stochastic or perfect deterministic cases. The data assimilation problem in both situations therefore involves accounting for these errors, as well as other errors in the system.

There are two main types of method which are used to solve the data assimilation problem. Sequential or three-dimensional methods combine a forecast and observational data at one time to produce an analysis. Starting from this analysis, the model is used to produce a forecast at the next analysis time when the forecast and observations are again combined, and so on throughout the time period. In four-dimensional data assimilation methods, the model trajectory is compared with the observations over a period of time to produce some kind of best fit over the entire time period. In this chap-

ter, we concentrate on sequential data assimilation, but the four-dimensional methods will be briefly discussed in the following chapter.

3.4 Linear systems

There is much theory for solving the data assimilation problem in the correct stochastic and perfect deterministic model cases, when the system to be estimated is described by linear equations. When we are dealing with nonlinear systems, it is often possible to linearise the system about some reference state, as will be described in the following section. It is possible to apply results from linear systems to these linearisations of nonlinear systems, after making certain assumptions. We therefore examine the linear data assimilation problem here and leave its extension to nonlinear systems to the next section.

3.4.1 Time invariant system

Much of the basic theory associated with control theory and data assimilation is based on the assumption that the forecast model and the observation operator are linear, [45], [49]. The further assumption that the model is time invariant is also useful when developing the theory, [2]. In the perfect deterministic model case, we write the linear, time-invariant true system equation as

$$\underline{x}_{k+1}^t = A\underline{x}_k^t + B\underline{u}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{3.4.1}$$

with $E\{\underline{x}_0^t\}$ given, and the model system equations as

$$\underline{x}_{k+1} = A\underline{x}_k + B\underline{u}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{3.4.2}$$

$$\underline{y}_k = H\underline{x}_k^t + \underline{\delta}_k, \qquad k = 0, 1, \ldots, N, \tag{3.4.3}$$

with $\underline{x}_0$ given. Here, $\underline{x}_k$, $\underline{u}_k^t$, and $\underline{\delta}_k$ have the same dimensions and assumptions as in Definition 3.7, $\underline{y}_k \in \mathbb{R}^p$, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $H \in \mathbb{R}^{p\times n}$.

We now define the concept of complete observability, which is an important one when trying to solve a problem of the type given.

Definition 3.8 *The model system (3.4.2), (3.4.3) is completely observable if there exists a finite time, $t_N > t_0$ such that knowledge of $\underline{u}_k$ and $\underline{y}_k$ for $t_k \in [t_0, t_N]$ suffices to determine $\underline{x}_0$ uniquely.*

As will be shown, complete observability often implies that the time invariant problem can be solved using methods from control theory. We therefore give an important condition on the complete observability of linear, time invariant systems, [2].

Theorem 3.1 *The model system (3.4.2), (3.4.3) is completely observable if and only if the observability matrix, C, has rank n, where*

$$C = \begin{bmatrix} H \\ HA \\ \vdots \\ HA^{n-1} \end{bmatrix}. \tag{3.4.4}$$

We now look at a general sequential data assimilation method, otherwise known as an observer or filter, which attempts to reconstruct the true state

of the system using some form of feedback. The usual form of feedback used for discrete systems of the type given, [45], is

1. Starting from some guess at the true state of the system at time t_k, $\underline{x}_k^a$, evolve the state variables to the next time step using the forecast model to obtain a forecast, $\underline{x}_{k+1}^f$, at time t_{k+1},

$$\underline{x}_{k+1}^f = A\underline{x}_k^a + B\underline{u}_k^t. \tag{3.4.5}$$

2. Combine the observations and the forecast to produce an analysis at time t_{k+1} as follows:

$$\underline{x}_{k+1}^a = \underline{x}_{k+1}^f + K[\underline{y}_{k+1} - H\underline{x}_{k+1}^f], \tag{3.4.6}$$

   where $K \in \mathbb{R}^{p\times n}$ is some gain matrix, also known as a weighting matrix.

Combining these two steps into one equation for the evolution of the analysis, we obtain

$$\underline{x}_{k+1}^a = (I - KH)(A\underline{x}_k^a + B\underline{u}_k^t) + K\underline{y}_{k+1}, \tag{3.4.7}$$

where $\underline{x}_0^a$ is our estimate of the initial conditions. The difference between our estimate of the state and the expected value of the true state at time t_{k+1} is given as

$$\hat{\underline{\epsilon}}_{k+1} = (I - KH)A\hat{\underline{\epsilon}}_k, \tag{3.4.8}$$

where $\hat{\underline{\epsilon}}_k = \hat{\underline{x}}_k^t - \underline{x}_k^a$ and $\hat{\underline{\mu}}_k$ denotes $E\{\underline{\mu}_k\}$. This last equation shows that, as long as we can choose the gain matrix K correctly, we should be able to ensure that the error will tend to zero as time increases, i.e. our analysis will converge to the true solution. We now give a condition under which this is possible.

Let $\Lambda_n = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be an arbitrary set of n complex numbers such that any which are not purely real occur in conjugate pairs. Then, from [2], [68], we have the following theorem.

Theorem 3.2 *The time invariant model system (3.4.2), (3.4.3) is completely observable if and only if there exists a real matrix K such that the characteristic roots of $(I - KH)A$ are the set Λ_n, for any set Λ_n.*

This theorem implies that we can choose the roots of $(I - KH)A$ to lie inside the unit circle, therefore ensuring that $\underline{x}_k^a \to \underline{\hat{x}}_k^t$ as $t_k \to \infty$, provided the model system (3.4.2), (3.4.3) is completely observable. In fact, it may be possible to stabilise the feedback system even if complete observability does not hold. In this case, the poles of the system which cannot be moved must already be stable, [52].

3.4.2 Time varying system

We write the time varying linear system, in the perfect deterministic model case as

$$\underline{x}_{k+1}^t = A_k \underline{x}_k^t + B_k \underline{u}_k^t + \underline{\zeta}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{3.4.9}$$

with $E\{\underline{x}_0^t\}$ given for the true system equations, and

$$\underline{x}_{k+1} = A_k \underline{x}_k + B_k \underline{u}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{3.4.10}$$

$$\underline{y}_k = H_k \underline{x}_k^t + \underline{\delta}_k, \qquad k = 0, 1, \ldots, N, \tag{3.4.11}$$

with $\underline{x}_0$ given for the model system equations. Here, $\underline{x}_k$, $\underline{u}_k^t$, and $\underline{\delta}_k$ have the same dimensions and assumptions as in Definition 3.7, $\underline{y}_k \in \mathbb{R}^{p_k}$, $A_k \in \mathbb{R}^{n \times n}$, $B_k \in \mathbb{R}^{n \times m}$ and $H_k \in \mathbb{R}^{p_k \times n}$.

The following definition gives an equivalent expression to Definition 3.8 for complete observability in the time varying case, [83].

Definition 3.9 *The model system (3.4.10), (3.4.11) is completely N-step observable at time t_k if and only if there exist a positive integer N such that knowledge of $\underline{y}_k, \underline{y}_{k+1}, \ldots, \underline{y}_{k+N-1}$ and $\underline{u}^t_k, \underline{u}^t_{k+1}, \ldots, \underline{u}^t_{k+N-2}$ is sufficient to determine the state $\underline{x}_k$.*

If the system is completely N-step observable for any time t_k, it is completely N-step observable. Furthermore, if the system is completely N-step observable for some N, then we say that it is completely observable, [83], [74].

We now introduce some notation to keep equations as simple as possible and define the state transition matrix, Φ, as follows:

$$\underline{x}_k = \Phi(k,j)\underline{x}_j, \qquad \forall k \geq j. \tag{3.4.12}$$

This matrix relates the state at time t_k to the state at some earlier time t_j and is given for the unforced model as

$$\Phi(k,j) = \prod_{i=j}^{k-1} A_i. \tag{3.4.13}$$

It has the following properties, [2]:

$$\Phi(j,j) = I, \qquad \forall j, \tag{3.4.14}$$

$$\Phi(j,k) = \Phi^{-1}(k,j), \qquad \forall k \geq j, \tag{3.4.15}$$

$$\Phi(l,j) = \Phi(l,k)\Phi(k,j), \qquad \forall j \leq k \leq l. \tag{3.4.16}$$

It is assumed in the above properties that the state transition matrix has an inverse which is not always true for irreversible processes.

It is now possible to write the model system (3.4.10), (3.4.11) in terms of one equation at each time as

$$\begin{aligned}
\hat{\underline{y}}_k &= H_k \hat{\underline{x}}_k, \\
\hat{\underline{y}}_{k+1} &= H_{k+1}\Phi(k+1,k)\hat{\underline{x}}_k + H_{k+1}B_k\underline{u}_k, \\
&\vdots \qquad\quad \vdots \\
\hat{\underline{y}}_{k+N-1} &= H_{k+N-1}\Phi(k+N-1,k)\hat{\underline{x}}_k + H_{k+N-1}\Big(\sum_{j=k}^{k+N-2}\Phi(k+N-1,j+1)B_j\underline{u}_j\Big).
\end{aligned}$$

If we put the known terms of these equations, i.e. those involving $\hat{\underline{y}}_j$, $\underline{u}_j$, on the left hand side, then we can write this as

$$\tilde{\underline{y}}_k = \begin{bmatrix} H_k \\ H_{k+1}\Phi(k+1,k) \\ \vdots \\ H_{k+N-1}\Phi(k+N-1,k) \end{bmatrix} \hat{\underline{x}}_k \equiv C_k \hat{\underline{x}}_k, \qquad (3.4.17)$$

where

$$\tilde{\underline{y}}_k = \begin{bmatrix} \hat{\underline{y}}_k \\ \hat{\underline{y}}_{k+1} - H_{k+1}B_k\underline{u}_k \\ \vdots \\ \hat{\underline{y}}_{k+N-1} - H_{k+1}(\sum_{j=k}^{k+N-2}\Phi(k+N-1,j+1)B_j\underline{u}_j) \end{bmatrix}. \qquad (3.4.18)$$

We can see from equation (3.4.17) that the system can be solved uniquely to determine $\hat{\underline{x}}_k$, provided the matrix C_k has certain properties.

Theorem 3.3 *The model system (3.4.10), (3.4.11) is completely N-step observable at time t_k if and only if $Rank(C_k) = n$.*

For a proof of this result see [83]. Note that the matrix C_k only involves A_j and H_j and that the condition on the rank of C_k therefore implies some

restriction on the relationship between the number of model variables and the number of observations. If we can show that a particular system is completely N-step observable, then we know it is possible to determine the state vector from a knowledge of the observations and inputs.

We now construct an observer of the same form as that for the time invariant system, that is

$$\underline{x}^f_{k+1} = A_k \underline{x}^a_k + B_k \underline{u}^t_k, \tag{3.4.19}$$

$$\underline{x}^a_{k+1} = \underline{x}^f_{k+1} + K_{k+1}[\underline{y}_{k+1} - H_{k+1}\underline{x}^f_{k+1}], \tag{3.4.20}$$

where $H_k \in \mathbb{R}^{p_k \times n}$. This leads to an equation for the evolution of our analysis as

$$\underline{x}^a_{k+1} = (I - K_{k+1}H_{k+1})(A_k \underline{x}^a_k + B_k \underline{u}^t_k) + K_{k+1}\underline{y}_{k+1}, \tag{3.4.21}$$

where $\underline{x}^a_0$ is our estimate of the initial conditions. The error in our estimate of the true state at time t_{k+1} is then given by

$$\underline{\hat{\epsilon}}_{k+1} = (I - K_{k+1}H_{k+1})A_k \underline{\hat{\epsilon}}_k, \tag{3.4.22}$$

where $\underline{\hat{\epsilon}}_k = \underline{\hat{x}}^t_k - \underline{x}^a_k$. Writing this as

$$\underline{\hat{\epsilon}}_k = \Big(\prod_{i=k-1}^{0} (I - K_{i+1}H_{i+1})A_i\Big)\underline{\hat{\epsilon}}_0 \equiv S_k \underline{\hat{\epsilon}}_0, \tag{3.4.23}$$

we see that a condition on the convergence of our observer is that $S_k \to 0$ as $k \to \infty$. We therefore have that a necessary condition for convergence, [79], [40], is that

$$\rho(S_k) \to 0 \qquad \text{as} \qquad t_k \to \infty, \tag{3.4.24}$$

where ρ denotes the spectral radius of a matrix.

The theory described so far for the time invariant and time varying cases has been for the perfect deterministic model case. The conditions on the stability and convergence of the observers also apply in the correct stochastic model case. This can be seen by taking the expectation of the true and model equations in the time varying case, say. This implies that the mean $\underline{\hat{x}}_k^t = E\{\underline{x}_k^t\}$ satisfies

$$\underline{\hat{x}}_{k+1}^t = A_k \underline{\hat{x}}_k^t + B_k \underline{u}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{3.4.25}$$

with $\underline{\hat{x}}_0^t$ given, and the mean model state $\underline{\hat{x}}_k^a = E\{\underline{x}_k^a\}$ satisfies

$$\underline{\hat{x}}_{k+1} = A_k \underline{\hat{x}}_k + B_k \underline{u}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{3.4.26}$$

$$\underline{\hat{y}}_k = H_k \underline{\hat{x}^t}_k, \qquad k = 0, 1, \ldots, N, \tag{3.4.27}$$

with $\underline{\hat{x}}_0$ given. Applying the observer to this system and finding the equation for the evolution of the error gives

$$\underline{\hat{\epsilon}}_{k+1} = (I - K_{k+1} H_{k+1}) A_k \underline{\hat{\epsilon}}_k, \tag{3.4.28}$$

where $\underline{\hat{\epsilon}}_k = \underline{\hat{x}}_k^t - \underline{\hat{x}}_k^a$. This means that the expectation of our analysis is the quantity which will converge to the expectation of the true system as time increases, provided the conditions given previously on the matrix S_k hold.

Much of the theory of observers for the time varying system was developed by Kalman, [49], [50]. A discussion of the Kalman filter will be given in the following chapter but we include some of the theory and results here. Kalman, [49], derives a method for choosing $K_k, k = 0, 1, \ldots, N$, based on statistical arguments for the correct stochastic model case, such that the resulting observer is optimal in the least squares sense. Conditions on the stability of the filter are also given.

Definition 3.10 *The observer (3.4.21) is said to be* **uniformly asymptotically stable** *if*

$$||S_k|| \leq c_1 e^{-c_2(t_k - t_0)}, \qquad \forall t_k \geq t_0, \tag{3.4.29}$$

where c_1, c_2 are positive constants and $||.||$ denotes the 2-norm, say.

This definition is taken from [45] and [14].

Theorem 3.4 *If the model system (3.4.10), (3.4.11) is uniformly completely observable and uniformly completely controllable and $P_0^a \geq 0$, then the Kalman filter is uniformly asymptotically stable.*

Here, $P_0^a = Cov\{\underline{\epsilon}_0\}$. For definitions of uniform complete observability and controllability, and a proof of this theorem, see [45], [75].

We have shown that, in the linear case, it is possible to construct an observer, under certain conditions, such that the feedback will ensure the stability and convergence of both the correct stochastic and perfect deterministic data assimilation problems. We now look at the nonlinear case.

3.5 Nonlinear systems

The model system equations in the correct stochastic model case are

$$\underline{x}_{k+1} = \underline{f}_k(\underline{x}_k, \underline{u}_k^t) + \underline{\zeta}_k, \qquad k = 0, 1, \ldots, N-1, \tag{3.5.1}$$

$$\underline{y}_k = \underline{h}_k(\underline{x}_k) + \underline{\delta}_k, \qquad k = 0, 1, \ldots, N. \tag{3.5.2}$$

We suppose that we generate some reference state at time t_k using the equation

$$\underline{\bar{x}}_{k+1} = \underline{f}_k(\underline{\bar{x}}_k, \underline{u}_k^t). \tag{3.5.3}$$

We also define a perturbation to this reference state as

$$\Delta \underline{x}_{k+1} \equiv \underline{x}_{k+1} - \bar{\underline{x}}_{k+1}. \tag{3.5.4}$$

These perturbations then satisfy the equation

$$\Delta \underline{x}_{k+1} = \underline{f}_k(\underline{x}_k, \underline{u}_k^t) - \underline{f}_k(\bar{\underline{x}}_k, \underline{u}_k^t) + \underline{\zeta}_k. \tag{3.5.5}$$

Assuming the perturbations are small in some sense, we can expand using a Taylor series to obtain

$$\Delta \underline{x}_{k+1} \approx F_k \Delta \underline{x}_k + \underline{\zeta}_k, \tag{3.5.6}$$

where $F_k \equiv \partial \underline{f}_k / \partial \underline{x}_k |_{\bar{\underline{x}}_k, \underline{u}_k^t}$ is the Jacobian of the forecast model, sometimes known as the tangent linear model.

We define a reference observation as

$$\bar{\underline{y}}_k \equiv \underline{h}_k(\bar{\underline{x}}_k), \tag{3.5.7}$$

and a perturbation to this as

$$\Delta \underline{y}_k \equiv \underline{y}_k - \bar{\underline{y}}_k. \tag{3.5.8}$$

These perturbations then satisfy the equation

$$\Delta \underline{y}_k = \underline{h}_k(\underline{x}_k) - \underline{h}_k(\bar{\underline{x}}_k) + \underline{\delta}_k. \tag{3.5.9}$$

Expanding $\underline{h}_k(\underline{x}_k)$ about the reference state using a Taylor series, we obtain

$$\Delta \underline{y}_k \approx H_k \Delta \underline{x}_k + \underline{\delta}_k, \tag{3.5.10}$$

where $H_k \equiv \partial \underline{h}_k / \partial \underline{x}_k |_{\bar{\underline{x}}_k, \underline{u}_k^t}$ is the Jacobian of the observation operator.

We now have a linearised system given by

$$\Delta \underline{x}_{k+1} = F_k \Delta \underline{x}_k + \underline{\zeta}_k, \tag{3.5.11}$$

$$\Delta \underline{y}_k = H_k \Delta \underline{x}_k + \underline{\delta}_k. \tag{3.5.12}$$

It is possible to apply the observers of the previous section to this linearised system but it is important to recognise that the system is only valid for small perturbations to the reference state.

The method usually used in the nonlinear case, [45], is as follows:

1. Starting from some estimate of the state $\underline{x}_k^a$ at time t_k, produce some reference state, now written as $\underline{x}_{k+1}^f$, at time t_{k+1} using the equation

$$\underline{x}_{k+1}^f = \underline{f}_k(\underline{x}_k^a, \underline{u}_k^t). \tag{3.5.13}$$

2. Produce an analysis at time t_{k+1} using the linear filter on the perturbations to this reference state,

$$\Delta \underline{x}_{k+1}^a = \Delta \underline{x}_{k+1}^f + K_{k+1}[\Delta \underline{y}_{k+1} - H_{k+1} \Delta \underline{x}_{k+1}^f]. \tag{3.5.14}$$

   The quantity $\Delta \underline{x}_{k+1}^f \equiv 0$, [45], and so this equation can be written, using equations (3.5.4) and (3.5.8), as

$$\underline{x}_{k+1}^a - \underline{x}_{k+1}^f = K_{k+1}[\underline{y}_{k+1} - \underline{h}_{k+1}(\underline{x}_{k+1}^f)]. \tag{3.5.15}$$

It is difficult to say anything about the stability or convergence of this process, especially in the stochastic model case. In the deterministic model case, we can write an equation for the evolution of the error as

$$\begin{aligned} \underline{\epsilon}_{k+1} &= \{\underline{f}_k(\underline{x}_k^t, \underline{u}_k^t) - K_{k+1}\underline{h}_{k+1}(\underline{x}_{k+1}^t)\} \\ &- \{\underline{f}_k(\underline{x}_k^a, \underline{u}_k^t) - K_{k+1}\underline{h}_{k+1}(\underline{x}_{k+1}^f)\} + \underline{\delta}_{k+1}, \end{aligned} \tag{3.5.16}$$

where $\underline{\epsilon}_k = \underline{x}^t_k - \underline{x}^a_k$. We now expand the function $\underline{f}_k(\underline{x}^t_k, \underline{u}^t_k)$ about the analysed state to obtain

$$\begin{aligned}
\underline{\epsilon}_{k+1} &= \{\underline{f}_k(\underline{x}^a_k + \underline{\epsilon}_k, \underline{u}^t_k) - K_{k+1}\underline{h}_{k+1}[\underline{f}_k(\underline{x}^a_k + \underline{\epsilon}_k, \underline{u}^t_k)]\} \\
&- \{\underline{f}_k(\underline{x}^a_k, \underline{u}^t_k) - K_{k+1}\underline{h}_{k+1}[\underline{f}_k(\underline{x}^a_k, \underline{u}^t_k)]\} + \underline{\delta}_{k+1} \qquad (3.5.17) \\
&\approx \frac{\partial \underline{f}_k}{\partial \underline{x}_k}\underline{\epsilon}_k - K_{k+1}\underline{h}_{k+1}[\underline{f}_k(\underline{x}^a_k, \underline{u}^t_k) + \frac{\partial \underline{f}_k}{\partial \underline{x}_k}\underline{\epsilon}_k] \\
&+ K_{k+1}\underline{h}_{k+1}[\underline{f}_k(\underline{x}^a_k, \underline{u}^t_k)] + \underline{\delta}_{k+1}, \qquad (3.5.18)
\end{aligned}$$

where the Jacobians are evaluated at $\underline{x}^a_k, \underline{u}^t_k$. Expanding $\underline{h}_{k+1}$ about $\underline{f}_k(\underline{x}^a_k, \underline{u}^t_k)$, we obtain

$$\begin{aligned}
\underline{\epsilon}_{k+1} &\approx \frac{\partial \underline{f}_k}{\partial \underline{x}_k}\underline{\epsilon}_k - K_{k+1}\underline{h}_{k+1}[\underline{f}_k(\underline{x}^a_k, \underline{u}^t_k)] \\
&- K_{k+1}\frac{\partial \underline{h}_{k+1}}{\partial \underline{f}_k}\frac{\partial \underline{f}_k}{\partial \underline{x}_k}\underline{\epsilon}_k + K_{k+1}\underline{h}_{k+1}[\underline{f}_k(\underline{x}^a_k, \underline{u}^t_k)] + \underline{\delta}_{k+1} \qquad (3.5.19) \\
&= (I - K_{k+1}H_{k+1})\frac{\partial \underline{f}_k}{\partial \underline{x}_k}\underline{\epsilon}_k + \underline{\delta}_{k+1}, \qquad (3.5.20)
\end{aligned}$$

where H_{k+1} is the Jacobian of the observation operator we defined previously. Taking the expectation operator gives

$$\hat{\underline{\epsilon}}_{k+1} \approx (I - K_{k+1}H_{k+1})\frac{\partial \underline{f}_k}{\partial \underline{x}_k}\hat{\underline{\epsilon}}_k. \qquad (3.5.21)$$

This equation is of the same form as that for the error evolution of the linear time varying case in equation (3.4.22). We can therefore say that, provided the Taylor series expansions are valid and the linearised system has the required properties, the data assimilation process should converge in the nonlinear perfect deterministic model case. The same method is also often applied to the correct stochastic model case.

3.6 Conclusions

In this chapter we have described the general data assimilation problem whereby some estimate of the state is required using the available observations and inputs. Theory and methods from control theory were discussed which showed that, under certain conditions, it is possible to reconstruct the state variables from a model and observations using some form of feedback. These observers, both in the time invariant and time varying cases, can be used to estimate the evolution of the state variables as long as conditions on the complete observability of the system hold.

The nonlinearity of most practical applications of the data assimilation problem is a problematic issue. It has been shown that under certain conditions, the data assimilation method will provide useful results in the nonlinear context, but the optimality of data assimilation methods is lost, as will be discussed in the next chapter. The size of the problem in the ocean or atmospheric context has not been discussed here but is a major issue when implementing data assimilation methods, as the computational cost of some of the methods is too high for existing computers. Also, the best model possible is often required to resolve flows adequately, so the data assimilation has to be computationally inexpensive. The effects this has on the type of methods used in practice will be seen in the next chapter.

Chapter 4

Data Assimilation Methods

4.1 Introduction

Many types of data assimilation have been proposed since the need to include observations to constrain a numerical model was first identified. To begin with subjective methods were used. These then progressed to simple empirical objective methods such as the Successive Corrections method, [7]. More recently, statistical methods have been derived where the errors in the observations and in the model are explicitly taken into account. These more recent methods, such as the Kalman filter, [49], [50], and variational methods, [71], were derived so that the optimal state of the system should be obtained, given certain assumptions.

One of the problems of using data assimilation in ocean models is that the size of the system is very large and there are large numbers of observations. This is also true of the numerical weather prediction problem, which is the field in which most of the data assimilation methods were originally

developed. However, the types of methods used in operational oceanography are not as sophisticated as those in some atmospheric models at present. We therefore include some of the more simple data assimilation methods in Section 4.2 and compare them with the statistically optimal method of the Kalman filter. In Section 4.3 we include an overview of the so-called four-dimensional data assimilation methods for completeness. The results of some experiments which intercompare some of the sequential data assimilation methods are given in Section 4.4. Conclusions are given in Section 4.5.

4.2 Sequential data assimilation methods

In this section, a derivation of the objective function which is used in most of the sequential data assimilation methods is given. We then give a review of some of those methods.

At time t_k, we assume that we have some prior estimate of the random n-vector $\underline{X}_k$, given by $\underline{x}_k^f$, and that this estimate contains errors,

$$\underline{X}_k = \underline{x}_k^f + \underline{\epsilon}_k, \tag{4.2.1}$$

where $\underline{\epsilon}_k$ is assumed to be a white Gaussian vector of errors with covariance matrix $P_k \equiv Cov\{\underline{\epsilon}_k\} \in \mathbb{R}^{n \times n}$. From Definition 3.5 in the previous chapter, we have that the PDF of $\underline{X}_k$ is given by

$$\underline{p}_{\underline{X}_k}(\underline{x}_k) = c_1 exp[-1/2(\underline{x}_k - \underline{x}_k^f)^T (P_k^f)^{-1}(\underline{x}_k - \underline{x}_k^f)], \tag{4.2.2}$$

where c_1 is a positive constant. We also assume that the random p_k-vector

$\underline{Y}_k$ is related to $\underline{X}_k$ by the equation

$$\underline{Y}_k = H_k \underline{X}_k + \underline{\delta}_k, \tag{4.2.3}$$

where $\underline{\delta}_k$ is assumed to be a white Gaussian vector of errors with covariance matrix $R_k \equiv Cov\{\underline{\delta}_k\} \in \mathbb{R}^{p_k \times p_k}$. The conditional PDF for $\underline{Y}_k$, given that $\underline{X}_k = \underline{x}_k$, is

$$\underline{p}_{\underline{Y}_k|\underline{X}_k=\underline{x}_k}(\underline{y}_k) = c_2 exp[-1/2(H_k\underline{x}_k - \underline{y}_k)^T R_k^{-1}(H_k\underline{x}_k - \underline{y}_k)], \tag{4.2.4}$$

where c_2 is a positive constant.

We would like to find the most likely estimate of $\underline{X}_k$, denoted $\underline{x}_k^a$, given that $\underline{Y}_k = \underline{y}_k$ where $\underline{y}_k$ is a particular realisation of $\underline{Y}_k$. From Bayes' theorem, [31], this conditional probability can be given as

$$\underline{p}_{\underline{X}_k|\underline{Y}_k=\underline{y}_k}(\underline{x}_k) = \frac{\underline{p}_{\underline{Y}_k|\underline{X}_k=\underline{x}_k}(\underline{y}_k)\underline{p}_{\underline{X}_k}(\underline{x}_k)}{\underline{p}_{\underline{Y}_k}(\underline{y}_k)} \tag{4.2.5}$$

$$= \frac{c_1 c_2}{c_3} exp[-1/2(H_k\underline{x}_k - \underline{y}_k)^T R_k^{-1}(H_k\underline{x}_k - \underline{y}_k) - 1/2(\underline{x}_k - \underline{x}_k^f)^T (P_k^f)^{-1}(\underline{x}_k - \underline{x}_k^f)], \tag{4.2.6}$$

where $\underline{p}_{\underline{Y}_k}(\underline{y}_k) = c_3$ because $\underline{y}_k$ is given.

The most likely estimate could be either the minimum variance or maximum likelihood estimates (mean or mode) which, in this linear case, are identical, [59]. For the maximum likelihood estimate, we maximise equation (4.2.6), which is equivalent to minimising $-ln\left(\frac{c_3}{c_1 c_2}\underline{p}_{\underline{X}_k|\underline{Y}_k=\underline{y}_k}(\underline{x}_k)\right)$. Therefore, given that $\underline{Y}_k = \underline{y}_k$ and the prior estimate $\underline{X}_k = \underline{x}_k^f$, the most likely estimate of $\underline{X}_k$ is given by the value $\underline{x}_k^a$ which minimises the variational objective function

$$\mathcal{J} = \frac{1}{2}(H_k\underline{x}_k - \underline{y}_k)^T R_k^{-1}(H_k\underline{x}_k - \underline{y}_k) + \frac{1}{2}(\underline{x}_k - \underline{x}_k^f)^T (P_k^f)^{-1}(\underline{x}_k - \underline{x}_k^f). \tag{4.2.7}$$

If we minimise this cost function directly, we obtain what is known as the 3D-Var solution. The Physical-space Statistical Analysis System (PSAS) is another way of minimising this objective function but differs from 3D-Var in that the minimisation is done in 'observation space', [18].

4.2.1 The Successive Corrections method

The Successive Corrections method (SCM) was one of the first data assimilation techniques to be implemented in practical problems. Bergthorsson and Doos [7] were the first to introduce the method in 1955, followed shortly after by Cressman [20]. This method assumes that we have a perfect deterministic model and that there are no errors in the observations. Corrections are made to a first guess or background state by adding a weighted difference between the observations and the background. This background state can be an output from a model or the climatology of the system. The SCM algorithm can be written as

$$\underline{x}_k^{j+1} = \underline{x}_k^j + K^{(j)}[\underline{y}_k - \underline{h}_k(\underline{x}_k^j)], \tag{4.2.8}$$

where $\underline{x}_k^0 = \underline{x}_k^b$ is the background state, $\underline{x}_k^i = \underline{x}_k^a$ is the analysis after i corrections and $K^{(i)}$ is a weighting matrix.

The main point of the method, as discussed in the previous chapter, is how we choose the weighting matrix. At first, this matrix was chosen empirically. However, over the years since the method was introduced there have been many suggestions as to the best and most efficient choices. In [20] the weights are chosen to smooth the observations into the analysis so that there are no sharp jumps in the solution.

In [21], Daley shows that if the weighting matrix is chosen so that the SCM

converges, then it will always converge to the observations as the number of corrections increases. In reality, observational errors exist and so the corrections are usually stopped after only a few iterations in practice.

4.2.2 Optimal Interpolation

Optimal Interpolation, sometimes known as Statistical Interpolation, [58], was first introduced in [70] and is also derived in [59], [60]. It is assumed that the observations are related to the model states by a linear operator H_k. Differentiating equation (4.2.7) with respect to the components of $\underline{x}_k$ and setting to zero gives

$$0 = H_k^T R_k^{-1}(\underline{y}_k - H_k \underline{x}_k^a) + P_k^{f\,-1}(\underline{x}_k^f - \underline{x}_k^a). \tag{4.2.9}$$

We can rearrange this equation to give an explicit expression for the analysis in terms of the observations and forecast, [59], as

$$\underline{x}_k^a = \underline{x}_k^f + P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}(\underline{y}_k - H_k \underline{x}_k^f), \tag{4.2.10}$$

which is in the same form as the usual sequential data assimilation with the gain matrix specified in terms of the error covariance matrices and observation operator as

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}. \tag{4.2.11}$$

The error covariance matrix of the analysis can also be obtained, [59], and is given by

$$\begin{aligned} P_k^a &= (I - P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1} H_k) P_k^f & (4.2.12)\\ &= (I - K_k H_k) P_k^f, & (4.2.13) \end{aligned}$$

where $P_k^a \equiv Cov\{\underline{x}_k^t - \underline{x}_k^a\} \in \mathbb{R}^{n \times n}$. At the next assimilation time, the forecast is obtained using the model equations.

In the Optimal Interpolation method, the underlying assumptions are those for the perfect deterministic model. The difficulty with this method is that the forecast error covariance matrix is not known accurately in practice. This matrix is usually estimated once and then held constant. The OI method is therefore not optimal over a period of time because the error covariance matrix P_k^f is not updated, or at least is not updated in a way which takes into account the observations already assimilated. This is one of the main differences between this method and the Kalman filter. If the observation operator is nonlinear, then it is still possible to use this method by linearising $\underline{h}_k$ about the forecast state, as described in Section 3.5.

4.2.3 The Analysis Correction scheme

The Analysis Correction (AC) scheme is a modified SCM. Here, corrections are made to the observations as well as the background state to take into account the error in the observations. In [8] a particular weighting matrix is derived using the method of OI as a reference to obtain the optimal weightings. Lorenc, Bell and MacPherson [61] also give a derivation of the method and show that it converges to the OI solution as the number of iterations increases. The assumptions for the AC scheme are the same as those for OI, i.e. perfect deterministic model. The AC algorithm can be written in the case of a linear observation operator, as

$$\underline{x}_k^{j+1} = \underline{x}_k^j + W_k V_k [\underline{y}_k^j - H_k \underline{x}_k^j], \tag{4.2.14}$$

$$\underline{y}_k^{j+1} = \underline{y}_k^j - V_k[\underline{y}_k^j - H_k\underline{x}_k^j], \tag{4.2.15}$$

where $\underline{y}_k^0$ is the vector of observations, $W_k = P_k^f H_k^T R_k^{-1}$, $V_k = (H_k W_k + I)^{-1}$ and $\underline{x}_k^i, \underline{y}_k^i \to \underline{x}_k^a, H_k\underline{x}_k^a$ as $i \to \infty$. If the observation operator is nonlinear, then it should be linearised about a background state. This matrix is then used in the formulation of W_k and V_k. The matrix V_k is usually denoted Q_k but we alter this to avoid confusion with the model error covariance matrix.

4.2.4 The Kalman filter

The Kalman filter is the optimal method over a period of time for time varying linear systems satisfying the assumptions of the correct stochastic model, [32], [33], [50]. The main distinctions between this and the other sequential methods is that the error covariance matrices are evolved with the analysis and random model error is taken into account. In other words, we are solving a minimisation similar to equation (4.2.7) but with time as an additional dimension.

With the assumptions of Section 3.3.2, and the error covariance matrix of our initial estimate of the state $\underline{x}_0^f$ given by $P_0^f \equiv Cov\{\underline{\epsilon}_0\} \in \mathbb{R}^{n\times n}$, the Kalman filter algorithm can be written as:

1. Combine the forecast and observations at time t_k as follows

$$\underline{x}_k^a = \underline{x}_k^f + K_k[\underline{y}_k - H_k\underline{x}_k^f], \tag{4.2.16}$$

where the gain matrix is given by $K_k = P_k^f H_k^T[H_k P_k^f H_k^T + R_k]^{-1}$. The error covariance of this analysis is

$$P_k^a = [I - K_k H_k]P_k^f. \tag{4.2.17}$$

2. Evolve the state variables to the next time step using the forecast model

$$\underline{x}^f_{k+1} = A_k \underline{x}^a_k + B_k \underline{u}^t_k. \tag{4.2.18}$$

The error covariance of the forecast is

$$P^f_{k+1} = A_k P^a_k A^T_k + Q_{k+1}. \tag{4.2.19}$$

Note that the cost of propagating the error covariance in equation (4.2.19)is very expensive when compared to the other steps in the Kalman filter procedure.

The Extended Kalman Filter

For a nonlinear problem, the Kalman filter can be altered to give the extended Kalman filter (EKF). To do this, the forecast model and observation operator must be linearised in some way, so that the error covariance matrices can be evolved from one time step to the next. These linearisations are done in the manner described in Section 3.5 to produce the tangent linear model, $F_k \equiv \partial \underline{f}_k / \partial \underline{x}_k |_{\underline{x}^f_k, \underline{u}^t_k}$, and the linearised observation operator $H_k \equiv \partial \underline{h}_k / \partial \underline{x}^f_k |_{\underline{x}^f_k, \underline{u}^t_k}$.

The only changes that are made to the algorithm for the usual Kalman filter are that the tangent linear model is now used to evolve the error covariance matrices rather than the full model as before and the linearised observation operator must be used when computing the gain matrix and analysis error covariance matrix. These approximations have the effect of destroying the optimality of the EKF, [11].

Simplified Kalman filtering

Both the Kalman filter and its nonlinear extension are very expensive data assimilation methods. Practical implementation of these methods on a global ocean or atmospheric model is not possible because of the size of the problem, [16]. It has therefore been an active area of research recently to produce some suboptimal approximations to the Kalman filter, [80]. The main computational cost of the Kalman filter is the propagation of the error covariance matrices at each time step. The simplifications usually made therefore attempt to reduce the size of these error covariance matrices by various methods.

One method for reducing the cost of propagating the error covariance matrices is by using a coarser grid for the calculation of these matrices. A second method, used by some weather centres, is to reduce the order of the covariance matrices by choosing those modes of the flow which move fastest. There are two main ways of doing this. One is by performing a singular value decomposition on the tangent linear model, $\underline{F}_k$, and choosing some number, L, of the leading singular values, [16]. The part of the tangent linear model associated with these leading singular values is then used to evolve the error covariance matrices. The second method is done by decomposing the matrix $\underline{F}_k P_k^a \underline{F}_k^T$ into its eigenvalues and eigenvectors and approximating it by the part which contains the L largest eigenvalues, [16]. The cost of both of these methods is proportional to L/n of the cost of the standard Kalman filter, [16].

Another method which reduces the cost of the full Kalman filter is the ensemble Kalman filter, [26]. Here, a large number of forecasts are made at each time step. This is done by perturbing the initial conditions of the

forecasts according to the analysis error covariance calculated there. The distribution of the forecasts then gives information about the errors associated with the forecast. An alternative is to assimilate into each ensemble member. The need to propagate the error covariance matrices is therefore eliminated, although a large ensemble, of the order of 100 members, is needed to produce statistically meaningful results.

4.3 Four-dimensional data assimilation methods

4.3.1 Four-dimensional variational assimilation

In four-dimensional variational data assimilation methods, we want to minimise some variational objective function over a period of time in order to fit the data over the entire period, subject to certain constraints. This idea is different from the sequential data assimilation in that the resulting trajectory will be smooth in some sense, due to the fact that we minimise an approximation of an integral over time, [38]. The variational methods were first introduced to meteorology in [71] and to physical oceanography in [69].

In the discrete case, the objective function to be minimised is

$$\mathcal{J} = \frac{1}{2}(\underline{x}_0 - \underline{x}_0^f)^T P_0^{f\,-1}(\underline{x}_0 - \underline{x}_0^f) + \frac{1}{2}\sum_{k=0}^{N-1}(\underline{h}_k(\underline{x}_k) - \underline{y}_k)^T R_k^{-1}(\underline{h}_k(\underline{x}_k) - \underline{y}_k), \tag{4.3.1}$$

with respect to $\underline{x}_0, \ldots, \underline{x}_N$, subject to certain constraints. There are two main approaches to this minimisation problem, [72].

Strong constraint

The strong constraint approach ensures the model equations, given by

$$\underline{x}_{k+1} = \underline{f}_k(\underline{x}_k, \underline{u}_k^t), \qquad k = 0, \ldots, N-1, \tag{4.3.2}$$

are satisfied exactly by the analysis, $\underline{x}_k^a$. The assumptions here are therefore those of the perfect deterministic model. In this case, the problem can be reduced to finding the initial conditions only, for which the objective function is minimised subject to the model equations, [54], [56], [59].

The constrained minimisation of the objective function (4.3.1) is equivalent to the unconstrained optimisation of the Lagrange function

$$\mathcal{L} = \mathcal{J} + \sum_{k=0}^{N-1} \underline{\lambda}_{k+1}^T (\underline{x}_{k+1} - \underline{f}_k(\underline{x}_k, \underline{u}_k^t)), \tag{4.3.3}$$

with respect to $\underline{x}_0, \ldots, \underline{x}_N, \underline{\lambda}_1, \ldots, \underline{\lambda}_N$, where $\underline{\lambda}_j \in \mathbb{R}^n$, $j = 1, \ldots, N$, are vectors of Lagrange multipliers. A necessary condition for finding the extremal of $\mathcal{L}$ is that its gradient with respect to $\underline{x}_j, \underline{\lambda}_j$, $j = 0, \ldots, N-1$ vanishes. It is therefore possible to write a set of equations for the adjoint vectors, $\underline{\lambda}_j$, which ensures that this condition is met. The equations are given by

$$\underline{\lambda}_N = 0, \tag{4.3.4}$$

$$\underline{\lambda}_k = \underline{F}_k^T \underline{\lambda}_{k+1} - H_k^T R_k^{-1} (\underline{h}_k(\underline{x}_k) - \underline{y}_k), \qquad k = 1, \ldots, N-1, \tag{4.3.5}$$

where $\underline{F}_k$ is the Jacobian of $\underline{f}_k(\underline{x}_k, \underline{u}_k^t)$ with respect to $\underline{x}_k$ and H_k is the Jacobian of $\underline{h}_k(\underline{x}_k)$ with respect to $\underline{x}_k$. A derivation of these equations is given in [38].

Weak constraint

In the weak constraint method, the model equations are satisfied approximately, allowing for random model error, [25]. The correct stochastic model assumptions are therefore made in this method. In this case, the objective function changes, and we now want to minimise

$$\begin{aligned} \mathcal{J} &= \frac{1}{2}(\underline{x}_0 - \underline{x}_0^f)^T P_0^{f\,-1}(\underline{x}_0 - \underline{x}_0^f) + \frac{1}{2}\sum_{k=0}^{N-1}(\underline{h}_k(\underline{x}_k) - \underline{y}_k)^T R_k^{-1}(\underline{h}_k(\underline{x}_k) - \underline{y}_k) \\ &+ \frac{1}{2}\sum_{k=0}^{N-1}\underline{\zeta}_k^T Q_k^{-1}\underline{\zeta}_k, \end{aligned} \tag{4.3.6}$$

with respect to $\underline{x}_0, \ldots, \underline{x}_N, \underline{\zeta}_0, \ldots, \underline{\zeta}_N$, subject to

$$\underline{x}_{k+1} = \underline{f}_k(\underline{x}_k, \underline{u}_k^t) + \underline{\zeta}_k, \qquad k = 0, \ldots, N-1. \tag{4.3.7}$$

In the linear case, the estimate of the state at the final time t_N, $\underline{x}_N$, is the same as the estimate given by the Kalman filter at that time.

In the variational methods, it is possible to estimate the errors in the analysis by calculating the Hessian or second derivative of the cost function with respect to $\underline{x}_k$. This is given by

$$P_k^a = \mathcal{J}''_{\underline{x}_k}. \tag{4.3.8}$$

The computational cost of performing four-dimensional variational assimilation is less than implementing the full Kalman filter. Calculating the analysis error covariance matrix from the Hessian, however, significantly increases the cost of the method to be similar to that of the Kalman filter.

4.3.2 Kalman smoother

The Kalman smoother is a data assimilation method which uses future observations as well as those past and present. For this reason, it is thought of as a four-dimensional method. There are various types of smoothers, [45], but here we discuss briefly the fixed-lag Kalman smoother (FLKS). In this case, the problem is to estimate the state of a system at all times, t_k, $k = 0, 1, \ldots$, using observations available from time t_0 to time t_{k+L}, where L is the fixed lag, [15]. The FLKS gives the optimal solution to the fixed-lag smoothing problem for linear problems with known model and observational error covariances. A derivation and discussion of the FLKS is given in [15]. The objective is similar to that of 4D-Var, but the perfect model assumption is not made and extensions to nonlinear dynamics and observation operators can be made in similar ways to the nonlinear extensions of the Kalman filter.

As was discussed in Section 4.2.4, the Kalman filter is too expensive to be fully implemented in oceanographic problems at present. The FLKS is $(M+2)/2$ times as expensive as the Kalman filter, where M is the number of analyses calculated at each observation time, [15]. However, approximations similar to those for the Kalman filter discussed in Section 4.2.4 could be applied to the FLKS to reduce its cost.

4.4 Experiments and results

To investigate the similarities and differences between the four sequential data assimilation techniques described in Section 4.2, we carry out some numerical experiments on two models described by ordinary differential equa-

tions. The first is a simple linear oscillating system for which we expect the data assimilation methods to perform well. The second is the chaotic, nonlinear Lorenz equations which will provide an insight into how well these methods are likely to perform when applied to more complicated models.

We describe the methods for the numerical solution of the two systems. A description of the experiments is also given, with the relevant results presented. The figures displayed in this section show the true solution (dashed lines), the solution given by the data assimilation scheme (solid lines) and the observations (crosses). The error in the data assimilation solution when compared to the true solution is also displayed.

4.4.1 Damped oscillating system

The damped oscillating system is given by the ordinary differential equation

$$\frac{d^2y}{dt^2} = -l\frac{dy}{dt} - ny, \tag{4.4.1}$$

where the damping, l, and the square of the frequency, n, are given values of 0.1 and 1.0 respectively. The differential equation (4.4.1) can be expressed as the first order system

$$\begin{pmatrix} \frac{dy}{dt} \\ \frac{dx}{dt} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -n & -l \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix}. \tag{4.4.2}$$

We discretise this system using a second order Runge-Kutta method which results in the following set of discrete equations,

$$y_{k+1} = (1 - n(\Delta t)^2/2)y_k + (\Delta t - l(\Delta t)^2/2)x_k, \tag{4.4.3}$$

$$\begin{aligned} x_{k+1} &= (nl(\Delta t)^2/2 - n\Delta t)y_k \\ &\quad +(1 - l\Delta t + l^2(\Delta t)^2/2 - n(\Delta t)^2/2)x_k, \end{aligned} \tag{4.4.4}$$

where $\Delta t = 0.1$ and $t_k = k\Delta t$, $k = 0, 1, \ldots, 500$. We define the 'true solution' to be given by the solution of these discretised equations with initial conditions of $y_0 = 1$ and $x_0 = 0$. We also define a background solution which is obtained using initial conditions which are equal to those for the true solution but with some random noise added. This noise has a variance of 0.1 and zero mean. This is the solution we would obtain, starting from the incorrect initial conditions, if no data assimilation was performed. The true solution is shown for the variable y in Figure 4.4.1.

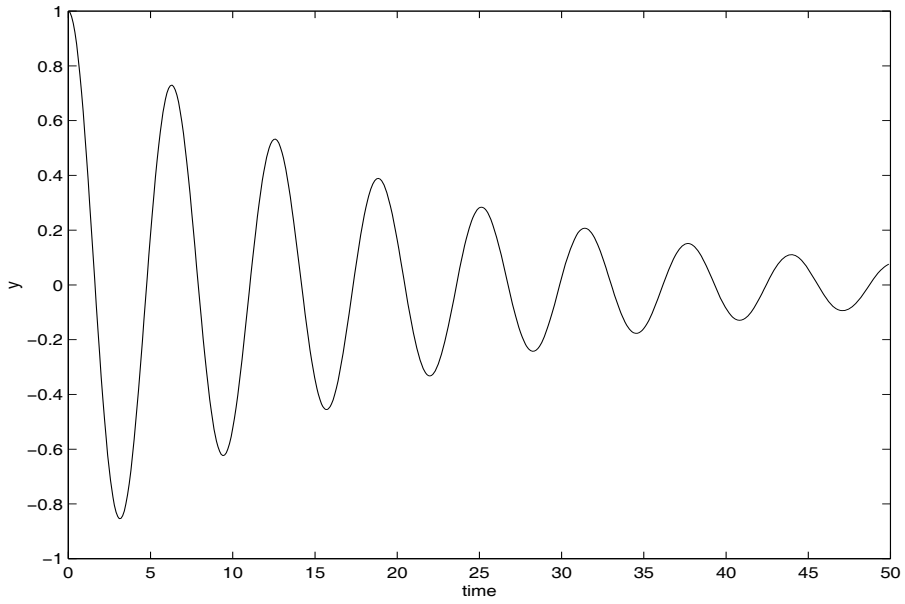

Figure 4.4.1: The damped oscillating system: true solution

In the experiments presented in this section, observations of both y and x are taken from the true solution at regular intervals over the first 25 time units. In one set of experiments the observations are taken every time unit and in the other they are taken every 2.5 time units. In both experiments,

noise was added to the observations which had a Gaussian distribution with variance of 0.1 and zero mean. The data assimilation is performed over the first 25 time units and the analysis at the end of this period is used as initial conditions for a forecast which lasts for another 25 time units.

When implementing the Successive Corrections method, we set the weighting matrix to be of a very simple form, i.e. $W = 0.5 \times I$. We only perform two iterations at each assimilation step. The results for the two experiments are shown in Figures 4.4.2 and 4.4.3. As has been said before, the SCM converges to the observations as the number of iterations increases so alteration of the number of corrections gives different results. After two corrections, both experiments give fairly poor results both during the assimilation period and for the forecast. The results for the experiment with observations every time unit shows a similar amount of error in the analysis to the one with observations every 2.5 time units.

For the Optimal Interpolation and Analysis Correction methods, the forecast error covariance matrix was calculated by averaging the statistics of differences in the background and true solutions and the observation error covariance matrix was calculated by averaging statistics of differences between the observations and the true solution over the assimilation interval. The results of these experiments are shown in Figures 4.4.4, 4.4.5, 4.4.6 and 4.4.7. Both these methods seem to do quite poorly both during the assimilation and in the forecast, the results being similar to those of the SCM. The AC scheme converges to the OI solution after only two iterations.

For the implementation of the Kalman filter, the model error covariance matrix, Q_k, was set to be zero. The KF performed the best out of all the

schemes as can be seen in Figures 4.4.8 and 4.4.9. When observations are available every time unit, the KF converges to the true solution after only 5 time units and follows the true solution throughout the rest of the assimilation and forecast. With observations every 2.5 time units, it takes the KF 7.5 time units to get close to the true solution but from then on tracks it well. These simple experiments have shown that in a linear system, the KF does indeed perform the best out of all the methods, as we would expect from the theory. We would now like to observe how well the methods perform when used on a nonlinear system.

Key for Figures 4.4.2 - 4.4.9: true solution (dashed line), analysed solution (solid line) and observations (crosses).

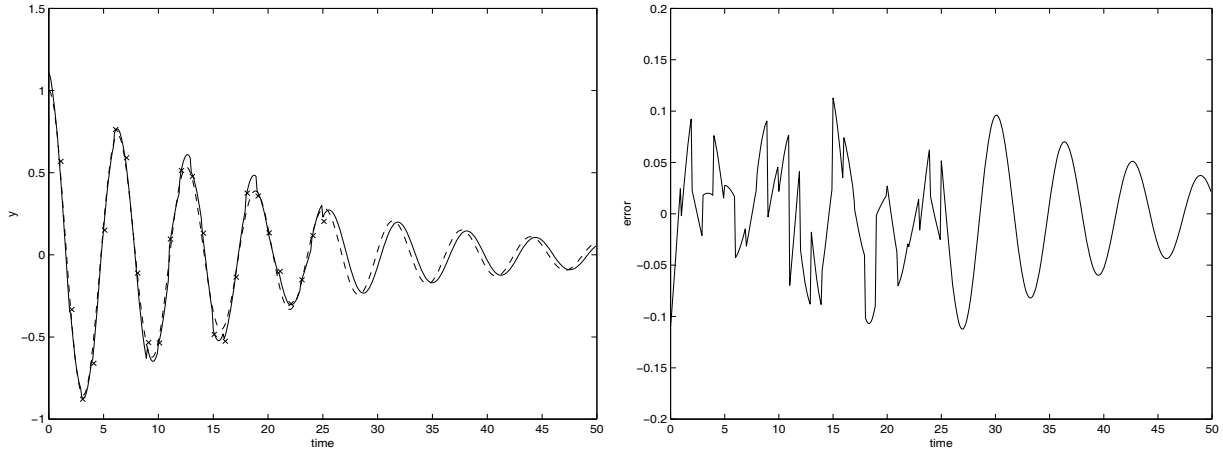

Figure 4.4.2: SCM on the oscillating system with observations every time unit after 2 iterations: (i) y variable. (ii) Error in y variable.

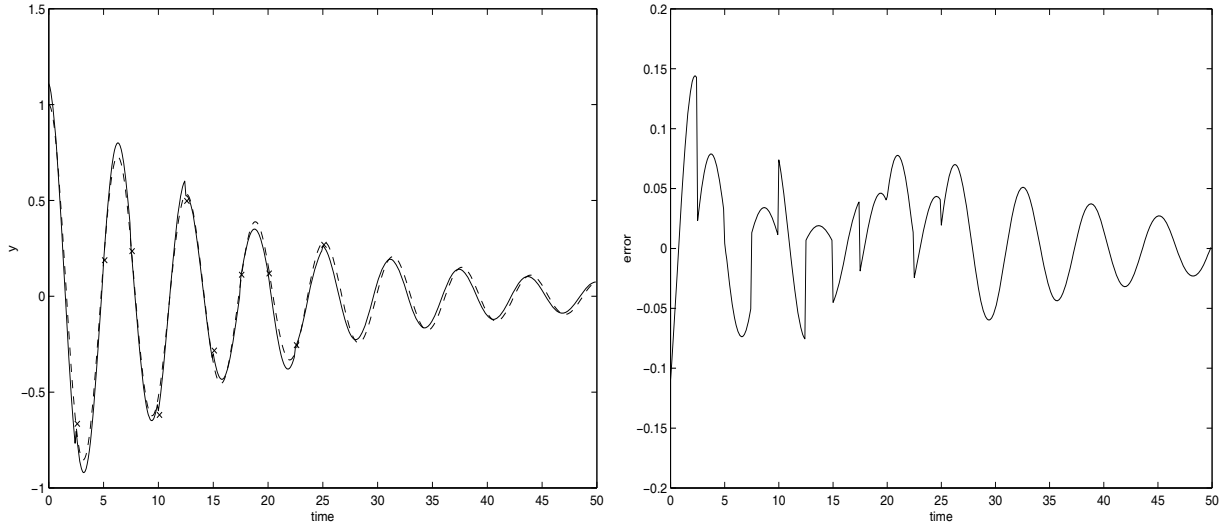

Figure 4.4.3: SCM on the oscillating system with observations every 2.5 time units after 2 iterations: (i) y variable. (ii) Error in y variable.

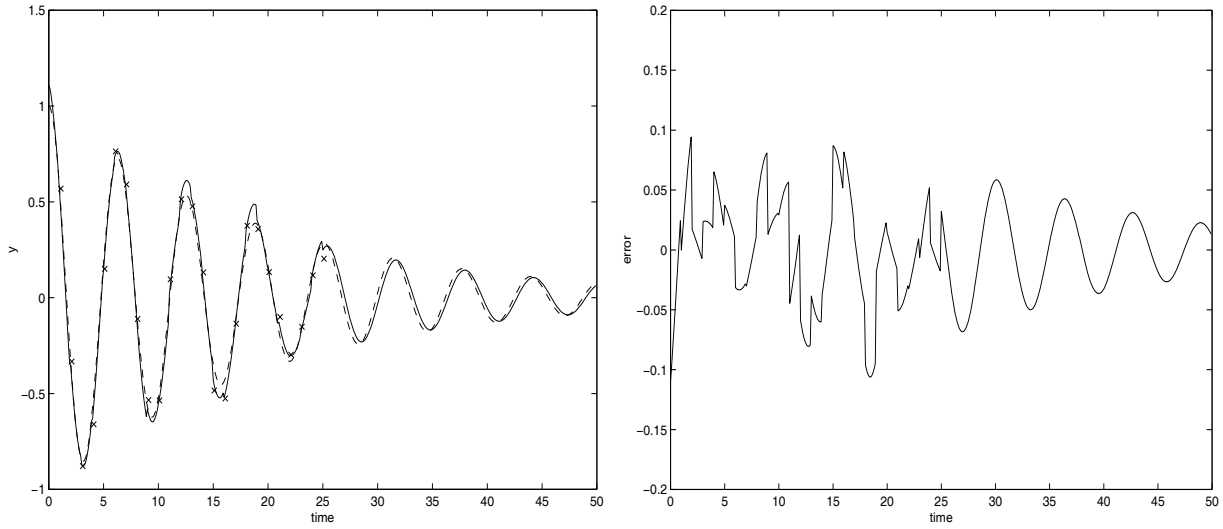

Figure 4.4.4: AC on the oscillating system with observations every time unit after 2 iterations: (i) y variable. (ii) Error in y variable.

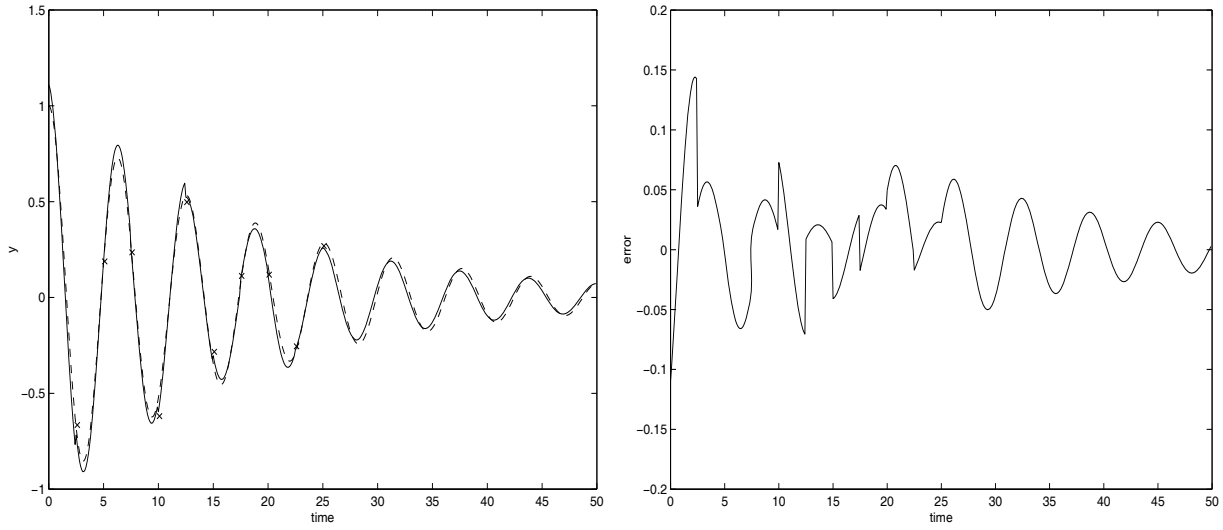

Figure 4.4.5: AC on the oscillating system with observations every 2.5 time units after 2 iterations: (i) y variable. (ii) Error in y variable.

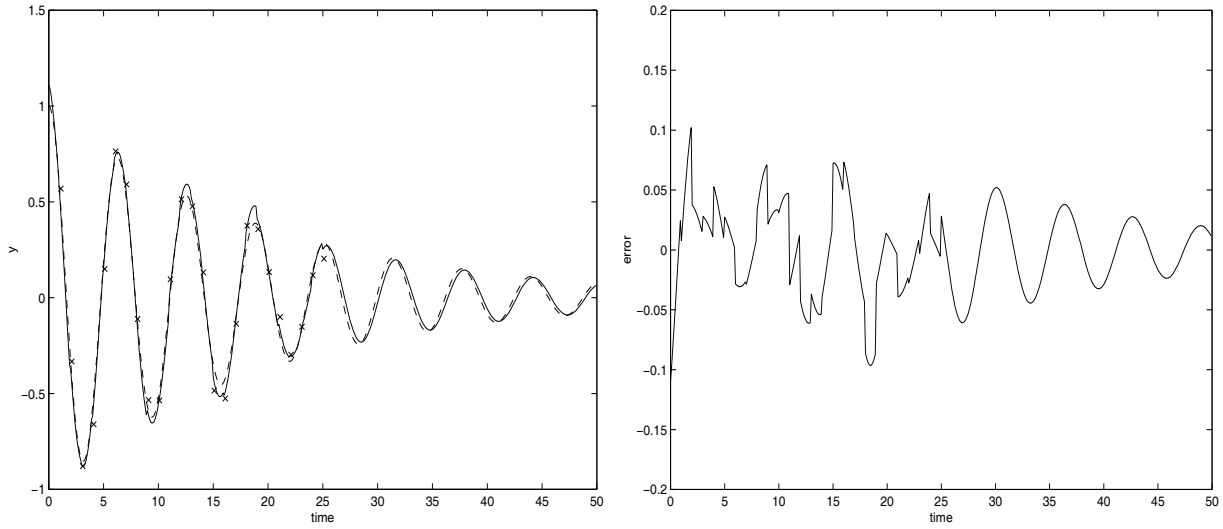

Figure 4.4.6: OI on the oscillating system with observations every time unit: (i) y variable. (ii) Error in y variable.

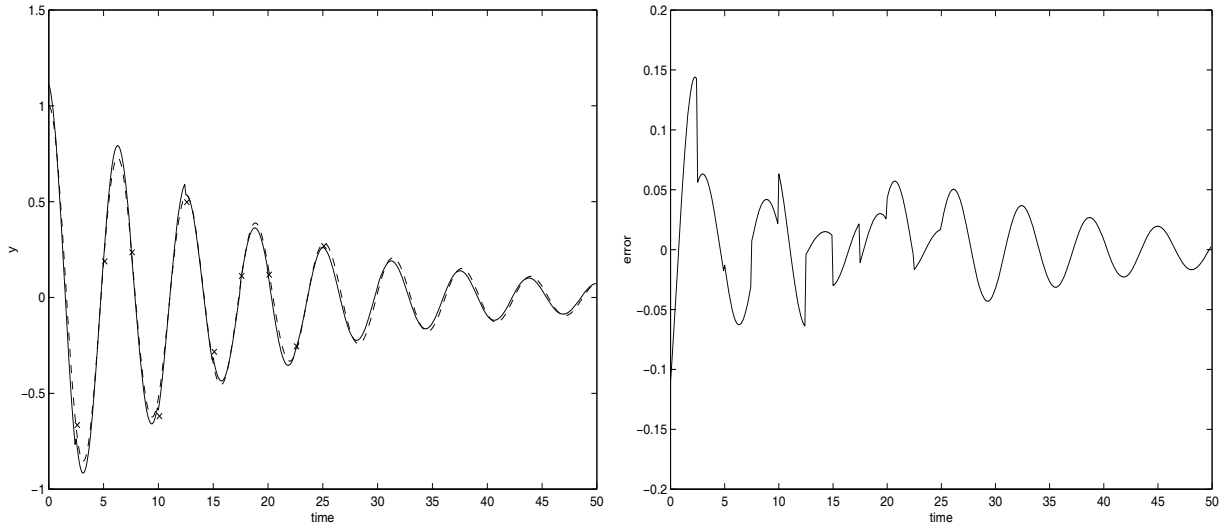

Figure 4.4.7: OI on the oscillating system with observations every 2.5 time units: (i) y variable. (ii) Error in y variable.

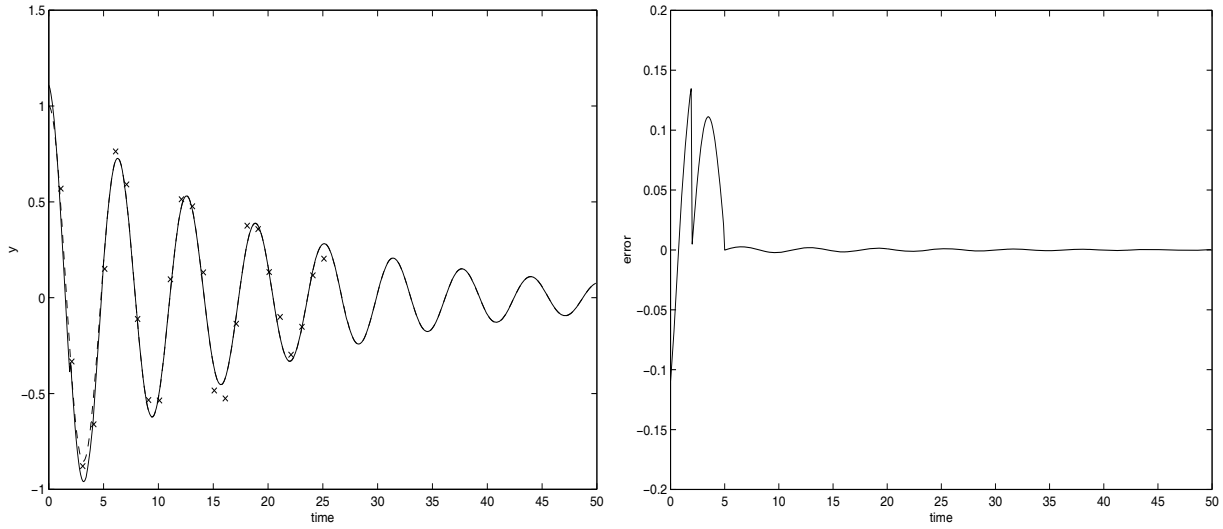

Figure 4.4.8: KF on the oscillating system with observations every time unit: (i) y variable. (ii) Error in y variable.

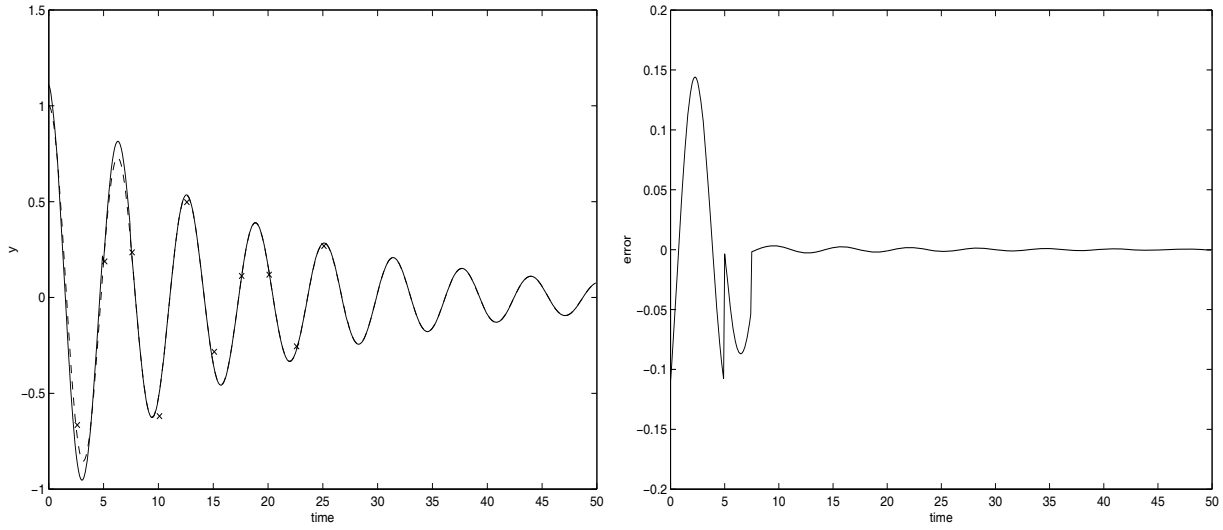

Figure 4.4.9: KF on the oscillating system with observations every 2.5 time units: (i) y variable. (ii) Error in y variable.

4.4.2 Lorenz equations

The Lorenz equations are a nonlinear system of three ordinary differential equations. They were originally obtained from the first terms in a Fourier truncation of the flow equations governing thermal convection, [65], and are often used in the testing of data assimilation methods, [30], [65], [27]. We write them as

$$\dot{x} = -\sigma(x-y), \tag{4.4.5}$$

$$\dot{y} = \rho x - y - xz, \tag{4.4.6}$$

$$\dot{z} = xy - \beta z. \tag{4.4.7}$$

where the parameters σ, ρ and β are chosen to have the values first used by Lorenz, [62], that is 10, 28 and 8/3 respectively. This set of parameter values gives chaotic solutions to the system, [76]. The system has three equilibrium points, one at the origin and two at the coordinates

$$(\pm\sqrt{\beta(\rho-1)}, \pm\sqrt{\beta(\rho-1)}, \rho-1). \tag{4.4.8}$$

All these equilibrium points are unstable for the choice of parameters given. The origin is an unstable saddle point and the other two equilibria are unstable spiral points.

We discretise this system using a second order Runge-Kutta method which gives us the following discrete equations,

$$\begin{aligned} x_{k+1} &= x_k + \sigma\Delta t/2[2(y_k - x_k) + \Delta t(\rho x_k - y_k - x_k y_k) \\ &- \sigma\Delta t(y_k - x_k)], \qquad (4.4.9) \\ y_{k+1} &= y_k + \Delta t/2[\rho x_k - y_k - x_k z_k + \rho(x_k + \sigma\Delta t(y_k - x_k)) - y_k \end{aligned}$$

$$
\begin{aligned}
& - \Delta t(\rho x_k - y_k - x_k z_k) - (x_k + \sigma \Delta t(y_k - x_k)) \\
& (z_k + \Delta t(x_k y_k - \beta z_k))], \qquad (4.4.10) \\
z_{k+1} &= z_k + \Delta t/2[x_k y_k - \beta z_k + (x_k + \Delta t \sigma(y_k - x_k)) \\
& (y_k + \Delta t(\rho x_k - y_k - x_k z_k)) - \beta z_k - \Delta t(x_k y_k - \beta z_k)], (4.4.11)
\end{aligned}
$$

where we choose $\Delta t = 0.01$, with $t_k = k\Delta t$, $k = 0, 1, \ldots, 3000$. Here, we define the 'true solution' to be the solution to these discrete equations with initial conditions given by $x_0 = y_0 = z_0 = 1.5$. The background solution is obtained using initial conditions which are equal to those for the true solution but with some random noise added. This noise has a variance of two and a mean of zero. The solutions for variables x and z are shown in Figures 4.4.10 and 4.4.11.

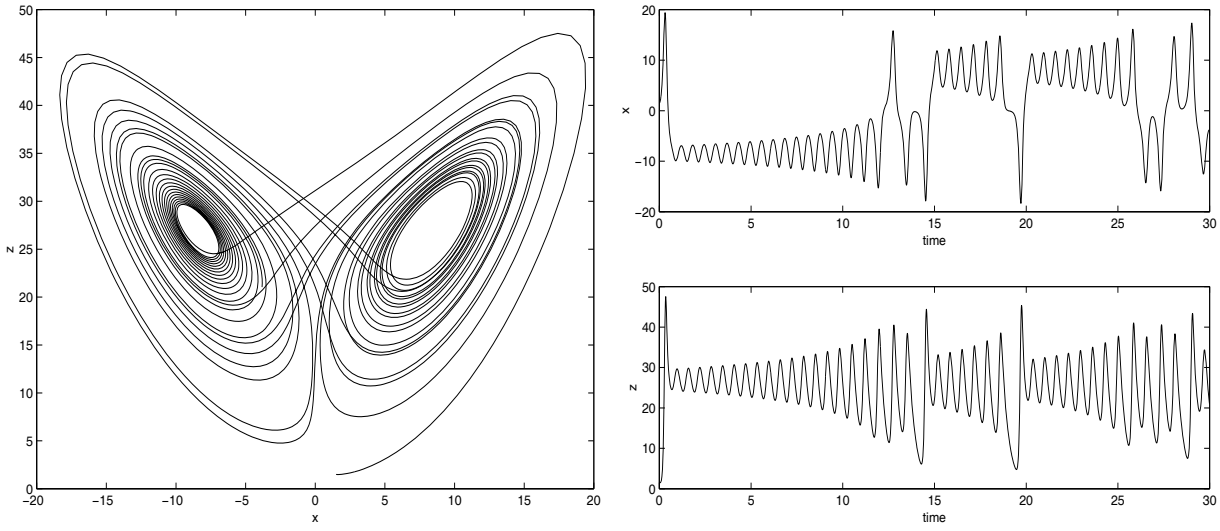

Figure 4.4.10: Lorenz equations: true solution

In the following experiments, we obtain observations of x, y and z from the true solution at regular intervals. In one set of experiments we have

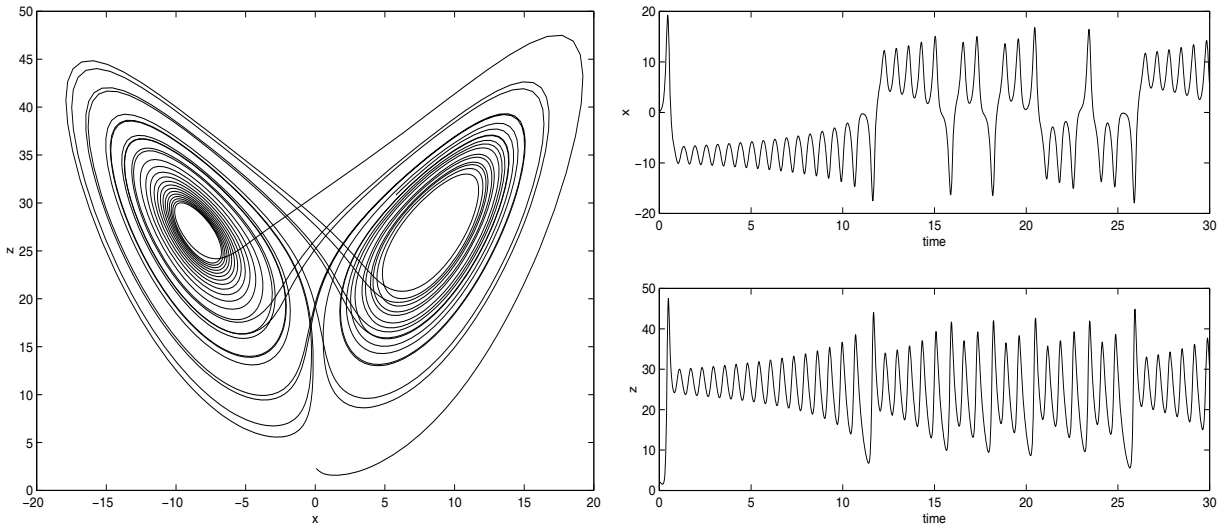

Figure 4.4.11: Lorenz equations: background solution

observations every 0.25 time units and in the other we have them every 0.5 time units. In both experiments we add random noise to the observations which has a Gaussian distribution with variance of two and a mean of zero. The assimilation is performed over the first 20 time units and the analysis at the end of this period is used as initial conditions for a forecast which lasts for another 10 time units.

The results of performing the SCM on the Lorenz equations are shown in Figures 4.4.12 and 4.4.13. When observations are available every 0.25 time units, the SCM performs reasonably well during the assimilation but the errors increase rapidly as soon as the forecast begins. With observations every 0.5 time units, the errors in the SCM begin to be significant after 12 time units of assimilation and continue to be large during the rest of the time interval.

For the Optimal Interpolation and Analysis Correction methods, the fore-

cast error covariance matrix was calculated by averaging the statistics of differences in the background and true solutions and the observation error covariance matrix was calculated by averaging statistics of differences between the observations and the true solution over the assimilation interval. The AC scheme performs well when observations are available every 0.25 time units and the forecast follows the true solution for about 5 time units, as can be seen in Figure 4.4.14. When observations are only available every 0.5 time units, large errors appear in the solution during the assimilation and are present throughout the rest of the assimilation and forecast as shown in Figure 4.4.15. The OI solution in both of these experiments is shown in Figures 4.4.16 and 4.4.17. These give qualitatively the same results as the AC scheme although there are slight differences in the magnitudes of the errors.

The extended Kalman filter has also been implemented on the Lorenz equations and the results are shown in Figures 4.4.18 and 4.4.19. Here we specify the model error covariance matrix, as suggested in [27], to be

$$Q_k = \begin{pmatrix} 0.1491 & 0.1505 & 0.0007 \\ 0.1505 & 0.9048 & 0.0014 \\ 0.0007 & 0.0014 & 0.9180 \end{pmatrix}. \tag{4.4.12}$$

The results here are not as good as the results when using the KF on the linear system and the effects of the loss of optimality when using it on a nonlinear system are clearly seen. However, the results are still at least as good as the other methods when there are observations every 0.25 time units. With observations every 0.5 time units, the EKF performs the best out of all the methods as it keeps close to the true solution for most of the assimilation period and tracks the forecast for about 5 time units.

Key for Figures 4.4.12 - 4.4.19: true solution (dashed line), analysed solution (solid line) and observations (crosses).

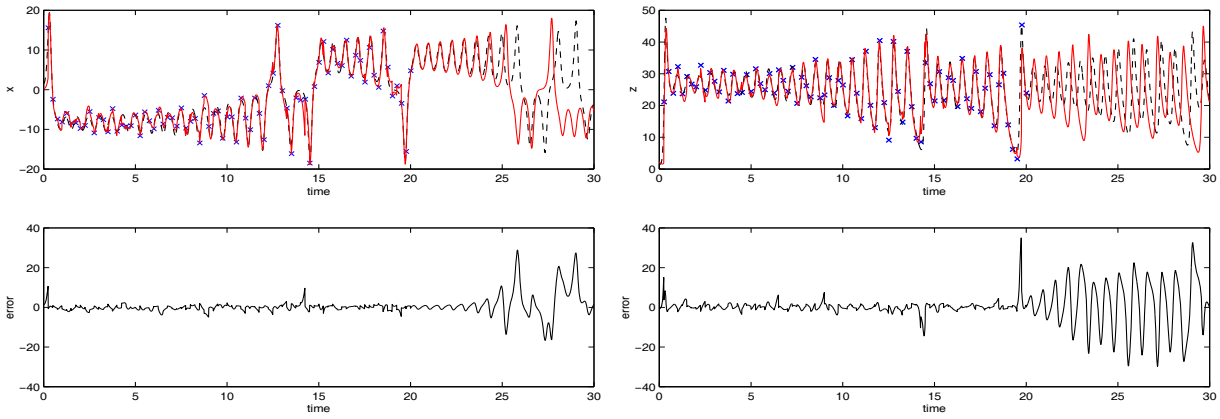

Figure 4.4.12: SCM on Lorenz equations with observations every 0.25 time units after 2 iterations: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

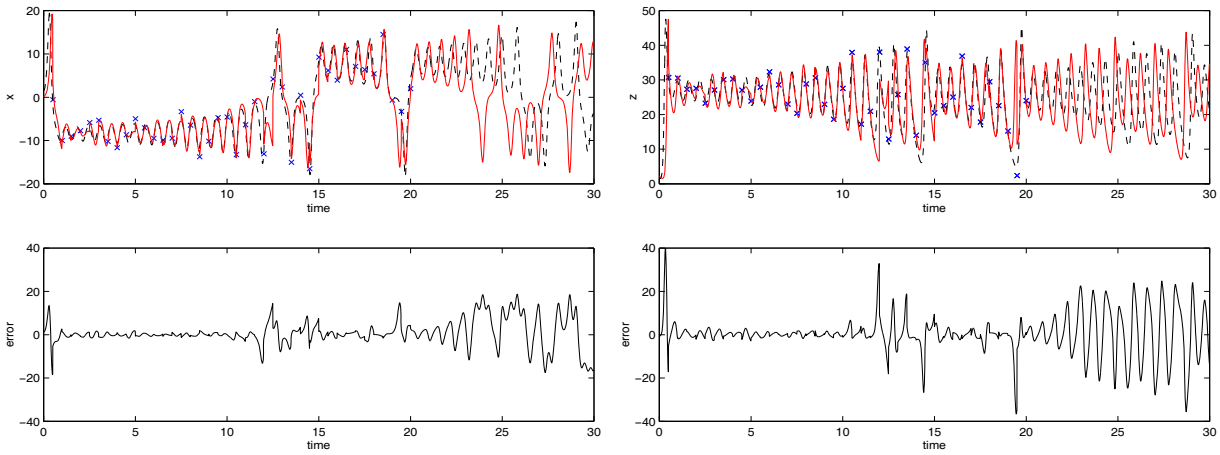

Figure 4.4.13: SCM on Lorenz equations with observations every 0.5 time units after 2 iterations: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

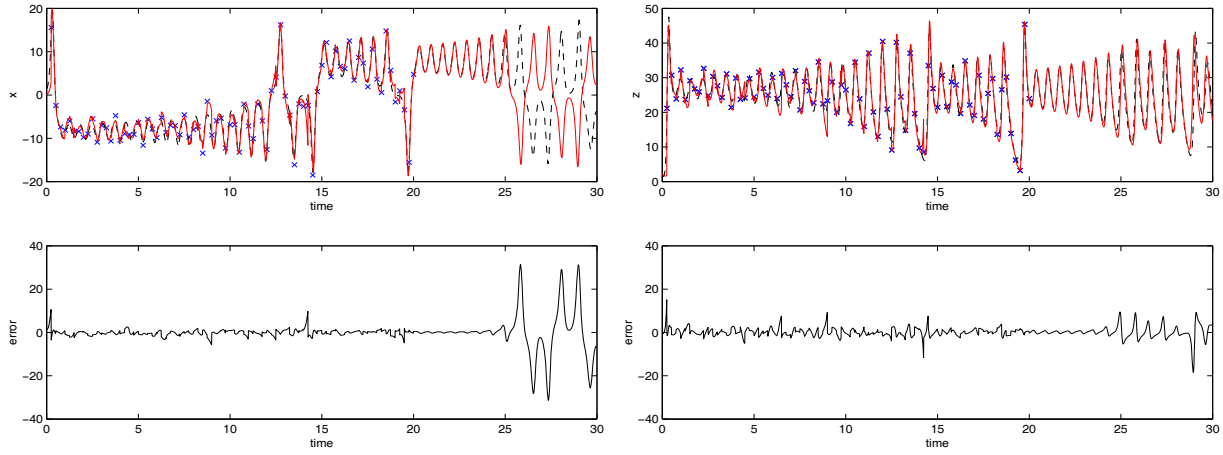

Figure 4.4.14: AC on Lorenz equations with observations every 0.25 time units after 2 iterations: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

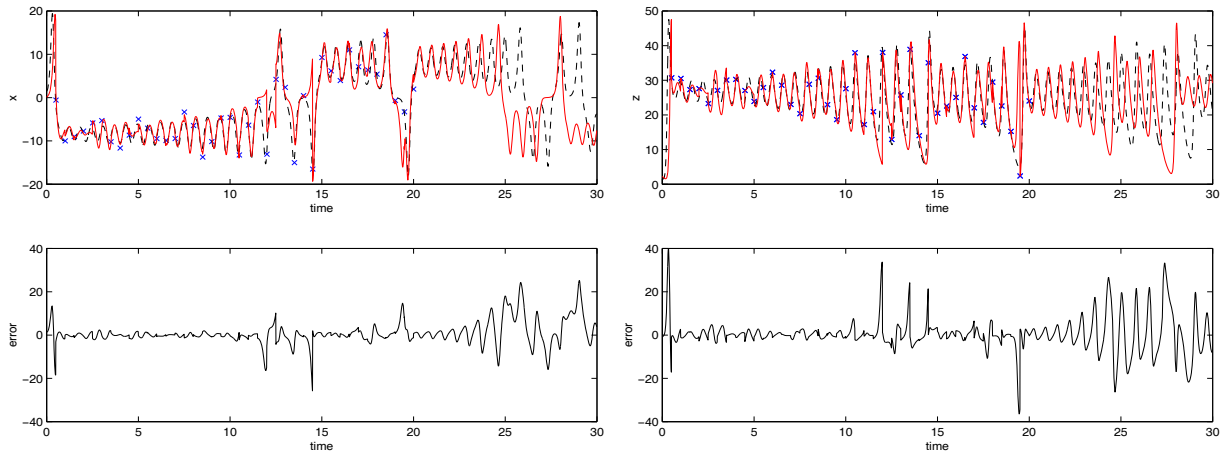

Figure 4.4.15: AC on Lorenz equations with observations every 0.5 time units after 2 iterations: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

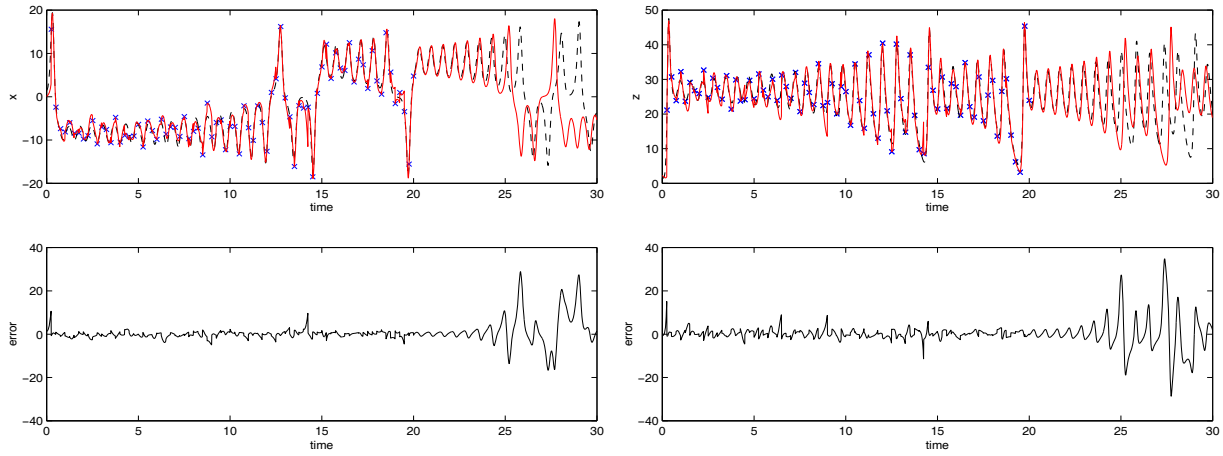

Figure 4.4.16: OI on Lorenz equations with observations every 0.25 time units: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

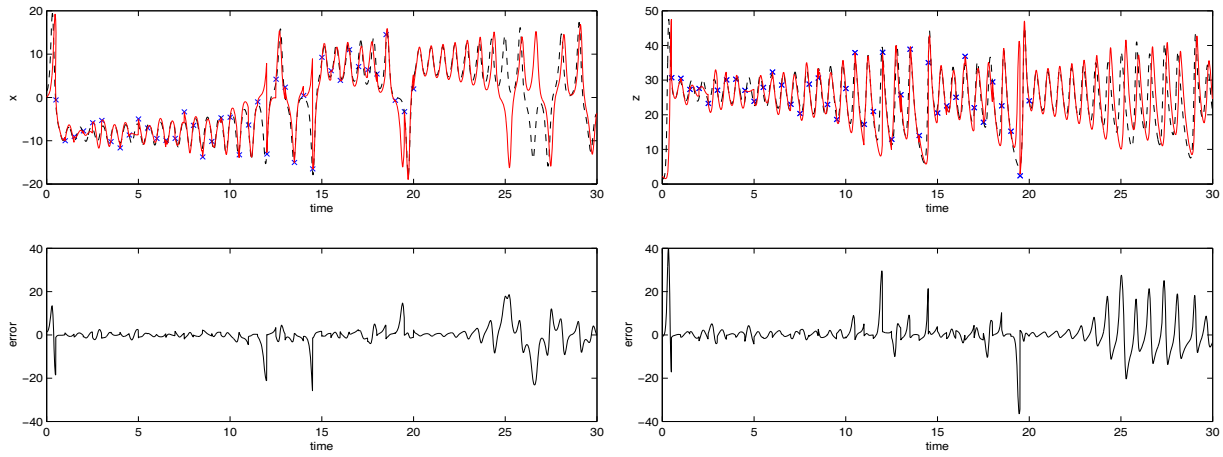

Figure 4.4.17: OI on Lorenz equations with observations every 0.5 time units: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

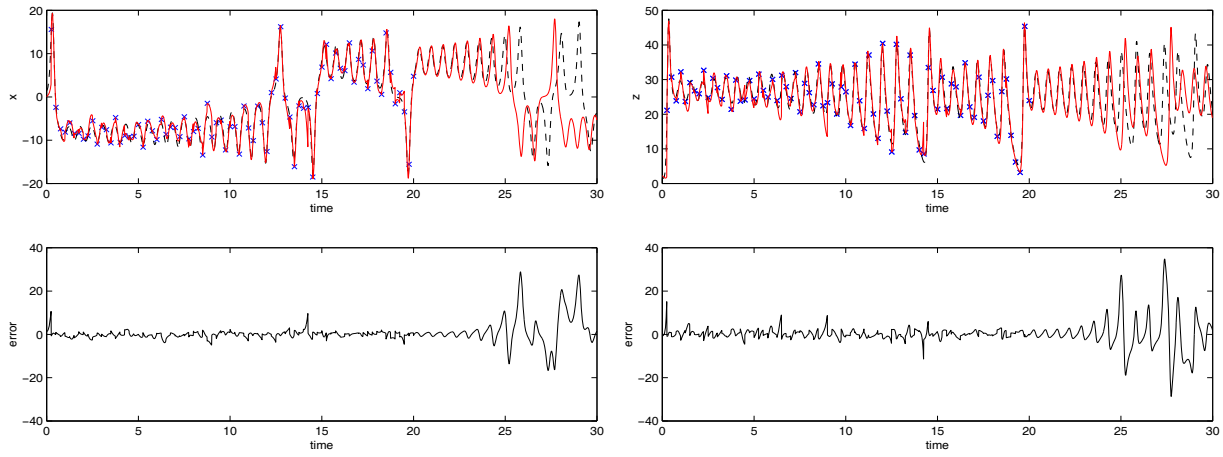

Figure 4.4.18: EKF on Lorenz equations with observations every 0.25 time units: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

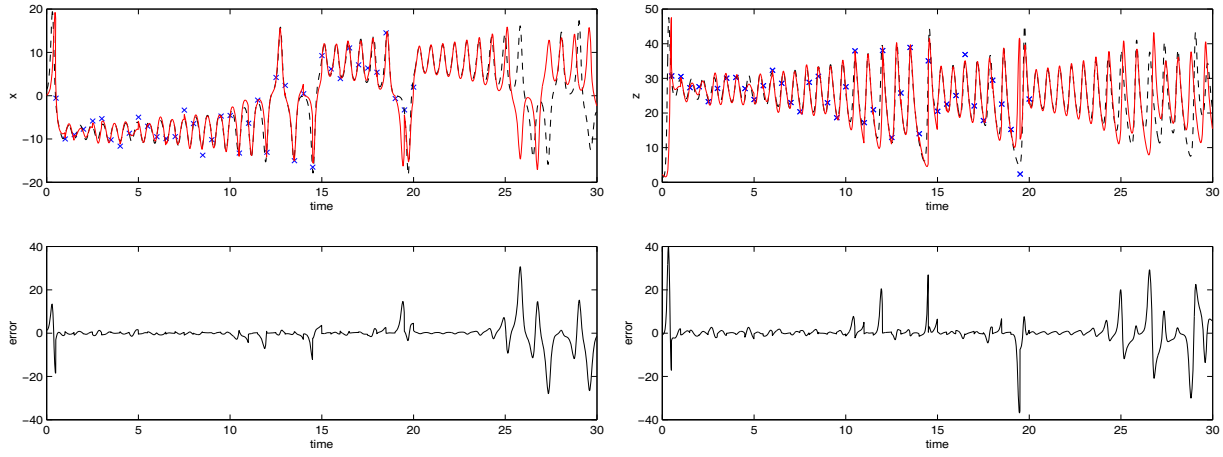

Figure 4.4.19: EKF on Lorenz equations with observations every 0.5 time units: (i) x variable (ii) z variable (iii) error in x variable (iv) error in z variable.

4.5 Conclusions

In this chapter, we first of all derived the objective function which most of the sequential data assimilation methods attempt to minimise. We then gave some information about the most common of the data assimilation methods, including the optimal method, known as the Kalman filter. We also gave a brief description of some of the four-dimensional methods for comparison. The size of the problem in most oceanographical settings will be too large for the Kalman filter to be implemented, even on the most powerful computers available today. Most operational centres therefore use some approximation, whether it is by using a different method or by using some simplification of the KF.

The last section in this chapter contained some experiments applying some of the sequential data assimilation methods to two ordinary differential equation models. One was a linear oscillating system on which most of the methods performed well. This illustrated that the KF is indeed the best method when applied to linear systems. The second model was the nonlinear, chaotic Lorenz equations. The EKF produced slightly better results than the other methods but the difference was not as great as in the linear system. For further examples of experiments on these models, see [63].

Chapter 5

Systematic Errors

5.1 Introduction

The data assimilation schemes described in the previous chapter assume that errors in the observations and model are random and Gaussian with zero mean. In operational forecasting of the ocean or atmosphere however, there will often be errors which are correlated in time. For instance, observing instruments may have a bias which would cause systematic errors in their data, [22]. Ocean models are also likely to contain some form of systematic error, [28]. These could arise from many different sources, such as incorrect specification of model parameters, truncation errors due to the numerical scheme used, and inaccurate forcing fields and boundary conditions. These violations of the assumptions made by the data assimilation schemes are likely to make the resulting analyses suboptimal, or worse. It is therefore desirable to examine methods which will account for these types of systematic errors.

In the engineering literature, a method for accounting for a constant model bias in a linear system is described in [29] with respect to the Kalman filter. Here, a vector of bias variables, which are specified to be constant in time, is added to the model forecast equation and the state vector is augmented with these bias variables. The matrices of the filter are transformed to give it a decoupled structure. This method has been extended to the cases where the bias contains noise, [44], and where the bias is time-varying, [78]. It has also been extended for use in nonlinear systems, [64], [85]. A similar method has also been proposed for use in meteorological data assimilation, [24]. The augmentation of the state vector by a vector of model error variables has also been developed in the variational assimilation framework, [25],[37], [38], [39].

A method for accounting for systematic model errors using data assimilation, based on [29], is described in Section 5.2. The application of this method to various data assimilation procedures is then described in Section 5.3. In Section 5.4, the problem of how to evolve the systematic model error variables is discussed. Some experiments applying the method to some models with different types of systematic model errors and the results of these experiments are described in Section 5.5, with conclusions given in Section 5.6.

5.2 Systematic errors

We assume that the true system equation is given, as in Chapter 3, by the stochastic vector difference equation

$$\underline{x}_{k+1}^t = \underline{f}_k(\underline{x}_k^t, \underline{u}_k^t) + \underline{\zeta}_k^t, \qquad k = 0, 1, \ldots, N-1, \tag{5.2.1}$$

with $E\{\underline{x}_0^t\}$ given, where $\{\underline{\zeta}_k^t, k = 0, 1, \ldots, N-1\}$ is a white Gaussian sequence. We also assume that we have observations of the form

$$\underline{y}_k = \underline{h}_k(\underline{x}_k^t) + \underline{\delta}_k, \qquad k = 0, 1, \ldots, N, \tag{5.2.2}$$

where $\{\underline{\delta}_k, k = 0, 1, \ldots, N\}$ is a white Gaussian sequence.

In Chapter 3 we gave the general model system equation as

$$\underline{x}_{k+1} = \underline{g}_k(\underline{x}_k, \underline{u}_k) + \underline{\zeta}_k, \qquad k = 0, 1, \ldots, N-1, \tag{5.2.3}$$

with $E\{\underline{x}_0\}$ given, where $\{\underline{\zeta}_k, k = 0, 1, \ldots, N-1\}$ is a white Gaussian sequence. Here, the model $\underline{g}_k$ is not the same as the true model $\underline{f}_k$, and the inputs to the model system are not necessarily the true inputs. In Chapter 3 we assumed that the model and inputs were correct, which ensured that the results on the convergence and stability of the data assimilation process held. We now relax this assumption.

We assume that the inputs to the model system are correct, but that the model we use to propagate the state variables and inputs contains systematic errors which we denote $\underline{b}_k \in \mathbb{R}^q$. We write this assumption as

$$\underline{g}_k(\underline{x}_k, \underline{u}_k) = \underline{f}_k(\underline{x}_k, \underline{u}_k) + T_k \underline{b}_k, \qquad k = 0, 1, \ldots, N-1, \tag{5.2.4}$$

where $T_k \in \mathbb{R}^{n \times q}$ is some operator which is known. This operator relates the systematic model error vector to the state vector, and is included because

it is possible that only certain parts of the model will contain systematic errors. In this case we will have $q < n$. We also assume that the evolution of the systematic model errors is governed by the stochastic vector difference equation

$$\underline{b}_{k+1} = \underline{m}_k(\underline{b}_k, \underline{x}_k) + \underline{\mu}_k, \qquad k = 0, 1, \ldots, N-1, \tag{5.2.5}$$

with $E\{\underline{b}_0\}$ given, where $\underline{m}_k : \mathbb{R}^q \times \mathbb{R}^n \to \mathbb{R}^q$ is assumed to be continuously differentiable and $\{\underline{\mu}_k \in \mathbb{R}^q, k = 0, 1, \ldots, N-1\}$ is a white Gaussian sequence. The following assumptions are made about the errors in the model system equations:

1.

$$E\{\underline{\zeta}_k\} = 0, \quad \forall t_k, \tag{5.2.6}$$

$$Cov\{\underline{\zeta}_k, \underline{\zeta}_j\} = \begin{cases} 0 & \text{if } t_j \neq t_k \\ Q_k^x & \text{if } t_j = t_k, \end{cases} \tag{5.2.7}$$

where Q_k^x is the known non-singular error covariance matrix of the model at time t_k.

2.

$$E\{\underline{\mu}_k\} = 0, \quad \forall t_k, \tag{5.2.8}$$

$$Cov\{\underline{\mu}_k, \underline{\mu}_j\} = \begin{cases} 0 & \text{if } t_j \neq t_k \\ Q_k^b & \text{if } t_j = t_k, \end{cases} \tag{5.2.9}$$

where Q_k^b is the known non-singular error covariance matrix of the stochastic errors in the systematic model error variables at time t_k.

3.

$$Cov\{\underline{\zeta}_k, \underline{\mu}_k\} = \begin{cases} 0 & \text{if } t_j \neq t_k \\ Q_k^{xb} & \text{if } t_j = t_k, \end{cases} \tag{5.2.10}$$

where Q_k^{xb} is the known non-singular cross-covariance matrix of errors in the state and systematic model error variables at time t_k.

4.

$$Cov\{\underline{\zeta}_k, \underline{\delta}_k\} = Cov\{\underline{\mu}_k, \underline{\delta}_k\} = 0, \quad \forall t_k, t_j. \tag{5.2.11}$$

5.

$$\begin{aligned} Cov\{\underline{\zeta}_k, \underline{x}_0\} &= Cov\{\underline{\zeta}_k, \underline{b}_0\} = 0, \quad \forall t_k, t_j, & (5.2.12)\\ Cov\{\underline{\mu}_k, \underline{x}_0\} &= Cov\{\underline{\mu}_k, \underline{b}_0\} = 0, \quad \forall t_k, t_j, & (5.2.13)\\ Cov\{\underline{\delta}_k, \underline{x}_0\} &= Cov\{\underline{\delta}_k, \underline{b}_0\} = 0, \quad \forall t_k, t_j. & (5.2.14) \end{aligned}$$

Under these assumptions, the true system is given by

$$\underline{x}_{k+1}^t = \underline{f}_k(\underline{x}_k^t, \underline{u}_k^t) + \underline{\zeta}_k^t, \quad k = 0, 1, \ldots, N-1, \tag{5.2.15}$$

with $E\{\underline{x}_0^t\}$ given. The model system is given by

$$\underline{x}_{k+1} = \underline{g}_k(\underline{x}_k, \underline{u}_k^t) + \underline{\zeta}_k, \quad k = 0, 1, \ldots, N-1, \tag{5.2.16}$$

with $E\{\underline{x}_0\}$ given, where $\underline{g}_k$ and $\underline{f}_k$ are given by equation (5.2.4) and the systematic errors evolve according to

$$\underline{b}_{k+1} = \underline{m}_k(\underline{b}_k, \underline{x}_k) + \underline{\mu}_k, \quad k = 0, 1, \ldots, N-1, \tag{5.2.17}$$

with $E\{\underline{b}_0\}$ given.

We now examine the effects of this serially correlated model error in the linear deterministic case, where we write the true system as

$$\underline{x}_{k+1}^t = A_k \underline{x}_k^t + B_k \underline{u}_k^t, \quad k = 0, 1, \ldots, N-1, \tag{5.2.18}$$

with $\underline{x}_0^t$ given. The observations are available from the true state of the system,

$$\underline{y}_k = H_k \underline{x}_k^t + \underline{\delta}_k, \quad k = 0, 1, \ldots, N, \tag{5.2.19}$$

and the model system equation is given by

$$x_{k+1} = A_k \underline{x}_k + B_k \underline{u}_k^t + T_k \underline{b}_k, \quad k = 0, 1, \ldots, N-1, \tag{5.2.20}$$

with $\underline{x}_0$ given. This form of bias correction is quite general in that it is possible to choose $T_k = A_k$ if the systematic error is on the state variables, $T_k = B_k$ if the systematic error is in the inputs or $T_k = I$ if the model equations contain systematic errors.

If we now perform the normal data assimilation procedure, as described in Chapter 3, on this system, that is

$$\underline{x}_{k+1}^f = A_k \underline{x}_k^a + B_k \underline{u}_k^t + T_k \underline{b}_k, \tag{5.2.21}$$

$$\underline{x}_{k+1}^a = \underline{x}_{k+1}^f + K_{k+1}[\underline{y}_{k+1} - H_{k+1} \underline{x}_{k+1}^f], \tag{5.2.22}$$

then we obtain an equation for the evolution of the analysis, given by

$$\underline{x}_{k+1}^a = (I - K_{k+1} H_{k+1})(A_k \underline{x}_k^a + B_k \underline{u}_k^t + T_k \underline{b}_k) + K_{k+1} \underline{y}_{k+1}. \tag{5.2.23}$$

The expected value of the difference between this analysed state and the true state is determined by

$$\begin{aligned} \hat{\underline{\epsilon}}_{k+1} &= \hat{\underline{x}}_{k+1}^t - \hat{\underline{x}}_{k+1}^a \\ &= (I - K_{k+1} H_{k+1}) A_k \hat{\underline{\epsilon}}_k - (I - K_{k+1} H_{k+1}) T_k \underline{b}_k, \end{aligned} \tag{5.2.24}$$

where $\hat{\underline{\epsilon}}_k \equiv E\{\underline{\epsilon}_k\}$. This shows that if our model contains systematic errors, then the difference between the analysed and true states will be forced by the systematic errors and so will not converge to zero as we would like.

5.2.1 State augmentation

We now return to the nonlinear stochastic case, where we augment the state vector with the systematic model error vector, that is

$$\underline{z}_k = \begin{pmatrix} \underline{x}_k \\ \underline{b}_k \end{pmatrix}, \tag{5.2.25}$$

where $\underline{z}_k \in \mathbb{R}^{n+q}$.

We now write the model system equations in terms of the augmented state vector as

$$\underline{z}_{k+1} = \tilde{\underline{g}}_k(\underline{z}_k, \underline{u}_k^t) + \tilde{\underline{\zeta}}_k, \quad k = 0, 1, \ldots, N-1, \tag{5.2.26}$$

with $E\{\underline{z}_0\}$ given, where $\tilde{\underline{g}}_k \equiv \begin{pmatrix} \underline{g}_k \\ \underline{m}_k \end{pmatrix} : \mathbb{R}^{n+q} \times \mathbb{R}^m \to \mathbb{R}^{n+q}$ is continuously differentiable and $\{\tilde{\underline{\zeta}}_k \equiv \begin{pmatrix} \underline{\zeta}_k \\ \underline{\mu}_k \end{pmatrix} \in \mathbb{R}^{n+q}, k = 0, 1, \ldots, N-1\}$ is a white Gaussian sequence.

The observations can be written in terms of the augmented state vector as

$$\underline{y}_k = \tilde{\underline{h}}_k(\underline{z}_k^t) + \underline{\delta}_k, \quad k = 0, 1, \ldots, N, \tag{5.2.27}$$

where $\tilde{\underline{h}}_k \equiv (\underline{h}_k(\underline{x}_k^t), 0) : \mathbb{R}^{n+q} \to \mathbb{R}^{p_k}$ is continuously differentiable.

In the linear deterministic case, we write the true system equations as

$$\underline{x}_{k+1}^t = A_k \underline{x}_k^t + B_k \underline{u}_k^t, \quad k = 0, 1, \ldots, N-1, \tag{5.2.28}$$

with $\underline{x}_0^t$ given. The observations are obtained from

$$\underline{y}_k = \tilde{H}_k \underline{z}_k^t + \underline{\delta}_k, \quad k = 0, 1, \ldots, N, \tag{5.2.29}$$

where $\tilde{H}_k \equiv (H_k, 0) \in \mathbb{R}^{p_k \times (n+q)}$. The augmented model equations are written as

$$\underline{z}_{k+1} = \tilde{A}_k \underline{z}_k + \tilde{B}_k \underline{u}_k^t, \quad k = 0, 1, \ldots, N-1, \tag{5.2.30}$$

with $\underline{z}_0$ given, where $\tilde{A}_k \equiv \begin{pmatrix} A_k & T_k \\ N_k & M_k \end{pmatrix} \in \mathbb{R}^{(n+q)\times(n+q)}$ and $\tilde{B}_k \equiv \begin{pmatrix} B_k \\ 0 \end{pmatrix} \in \mathbb{R}^{m\times(n+q)}$ which we can write separately as

$$\underline{x}_{k+1} = A_k \underline{x}_k + B_k \underline{u}_k^t + T_k \underline{b}_k, \quad k = 0, 1, \ldots, N-1, \tag{5.2.31}$$

$$\underline{b}_{k+1} = M_k \underline{b}_k + N_k \underline{x}_k, \quad k = 0, 1, \ldots, N-1, \tag{5.2.32}$$

with $\underline{x}_0$, $\underline{b}_0$ given.

We can now perform the normal data assimilation process on this augmented state to obtain

$$\underline{z}_{k+1}^f = \tilde{A}_k \underline{z}_k^a + \tilde{B}_k \underline{u}_k^t, \tag{5.2.33}$$

$$\underline{z}_{k+1}^a = \underline{z}_{k+1}^f + \tilde{K}_{k+1}[\underline{y}_{k+1} - \tilde{H}_{k+1} \underline{z}_{k+1}^f], \tag{5.2.34}$$

where $\tilde{K}_k \equiv \begin{pmatrix} K_k^x \\ K_k^b \end{pmatrix} \in \mathbb{R}^{p_k \times (n+q)}$ is some gain matrix to be determined by a particular data assimilation method.

If we assume that the model used to evolve the systematic model error variables is the correct model, that is

$$\underline{b}_{k+1}^t = M_k \underline{b}_k^t + N_k \underline{x}_k^t, \quad k = 0, 1, \ldots, N-1, \tag{5.2.35}$$

then the expected value of the difference between the analysed state and the true state, and between the analysed and true systematic model error variables is determined by

$$\tilde{\underline{\epsilon}}_{k+1} = \begin{pmatrix} (I - K_{k+1}^x H_{k+1}) A_k & (I - K_{k+1}^x H_{k+1}) T_k \\ N_k - K_{k+1}^b H_{k+1} A_k & M_k - K_{k+1}^b H_{k+1} T_k \end{pmatrix} \tilde{\underline{\epsilon}}_k, \tag{5.2.36}$$

where $\tilde{\underline{\epsilon}}_k = \begin{pmatrix} \hat{\underline{x}}_k^t \\ \hat{\underline{b}}_k^t \end{pmatrix} - \begin{pmatrix} \hat{\underline{x}}_k^a \\ \hat{\underline{b}}_k^a \end{pmatrix}$.

In this case, provided we can choose the gain matrix so that

$$\rho(\tilde{S}_k) \to 0 \quad \text{as} \quad t_k \to \infty, \tag{5.2.37}$$

where $\tilde{S}_k \equiv \prod_{i=k-1}^{0}(I - \tilde{K}_{i+1}\tilde{H}_{i+1})\tilde{A}_i$, the analysis will converge to the true solution as time increases, that is

$$\hat{\underline{x}}_k^a \to \hat{\underline{x}}_k^t, \quad \hat{\underline{b}}_k^a \to \hat{\underline{b}}_k^t, \quad \text{as} \quad t_k \to \infty. \tag{5.2.38}$$

This is an important result as it shows that we can account for systematic model errors by augmenting the state vector with a vector of systematic model error variables and performing the data assimilation process on this augmented state. This also shows that as well as converging to the true state of the system, we will obtain an estimate of the systematic model errors themselves. This could be very useful information when trying to improve the model.

It may be that we do not know exactly how the systematic model errors evolve. If we can make a reasonable estimate, however, the analysis of the augmented state should provide a better analysis than without the systematic model error correction.

5.2.2 Observability of the augmented state

In Section 3.4 we showed that complete observability of the system implies that we can choose the eigenvalues of the matrix evolving the error to ensure that the analysis will converge. We therefore look for similar conditions for

the augmented state. The conditions on the complete observability of time invariant systems, found in Section 3.4, follow directly for the augmented state. We concentrate here on time varying systems.

To keep the following simple, we assume that the evolution of the bias variables does not depend on the state variables, i.e. $N_k \equiv 0$, $\forall k$. The state transition matrix for the augmented system can now be written as

$$\tilde{\Phi}(k,j) = \begin{pmatrix} \Phi(k,j) & \sum_{i=j}^{k-1} \Phi(k,i+1) T_i \Psi(i,j) \\ 0 & \Psi(k,j) \end{pmatrix}, \tag{5.2.39}$$

where $\Psi(k,j) = \prod_{i=j}^{k-1} M_i$ is the transition matrix for the systematic model error variables and Φ is the transition matrix for the state variables, as introduced in Section 3.4. We can now write the augmented system in terms of one equation at each time as,

$$\begin{aligned} \underline{\hat{y}}_k &= \tilde{H}_k \underline{\hat{z}}_k, \\ \underline{\hat{y}}_{k+1} &= \tilde{H}_{k+1} \tilde{\Phi}(k+1,k) \underline{\hat{z}}_k + \tilde{H}_{k+1} \tilde{B}_k \underline{u}_k, \\ &\vdots \qquad\qquad \vdots \\ \underline{\hat{y}}_{k+N} &= \tilde{H}_{k+N} \tilde{\Phi}(k+N,k) \underline{\hat{z}}_k \\ &\quad + \tilde{H}_{k+N} \Big(\sum_{j=k}^{k+N-1} \tilde{\Phi}(k+N,j+1) \tilde{B}_j \underline{u}_j \Big). \end{aligned}$$

Again, we put the known terms on the left hand side and write this as

$$\underline{\tilde{y}}_k = \begin{bmatrix} \tilde{H}_k \\ \tilde{H}_{k+1} \tilde{\Phi}(k+1,k) \\ \vdots \\ \tilde{H}_{k+N} \tilde{\Phi}(k+N,k) \end{bmatrix} \underline{\hat{z}}_k \equiv \tilde{C}_k \underline{\hat{z}}_k, \tag{5.2.40}$$

where

$$\tilde{\underline{y}}_k = \begin{bmatrix} \hat{\underline{y}}_k \\ \hat{\underline{y}}_{k+1} - \tilde{H}_{k+1}\tilde{B}_k\underline{u}_k \\ \vdots \\ \hat{\underline{y}}_{k+N} - \tilde{H}_{k+N}(\sum_{j=k}^{k+N-1}\tilde{\Phi}(k+N,j+1)\tilde{B}_j\underline{u}_j) \end{bmatrix}. \tag{5.2.41}$$

So, from Theorem 3.3, we get the following theorem.

Theorem 5.1 *The augmented system is completely N-step observable at time t_k if and only if $Rank(\tilde{C}_k) = n + q$.*

It is possible to expand the matrix $\tilde{C}_k$ as

$$\tilde{C}_k = (C_k, D_k), \tag{5.2.42}$$

where

$$C_k = \begin{bmatrix} H_k \\ H_k\Phi(k+1,k) \\ \vdots \\ H_{k+N}\Phi(k+N,k) \end{bmatrix}, \tag{5.2.43}$$

and

$$D_k = \begin{bmatrix} 0 \\ H_kT_k \\ \vdots \\ H_{k+N}(\sum_{j=k}^{k+N-1}\Phi(k+N,j+1)T_j\Psi(j,k)) \end{bmatrix}. \tag{5.2.44}$$

Note that the matrix C_k is the same as the observability matrix obtained in Section 3.4 for the system without systematic model errors. This leads us to the following theorem.

Theorem 5.2 *Necessary conditions for the augmented system to be completely N-step observable at time t_k are:*
(i) the original system is completely N-step observable
(ii) $Rank(D_k) = q$.

A proof of this result can be obtained by an extension of the proof for Theorem 5.7 in [38].

It is possible to apply the Kalman filter to the augmented state directly because the assumptions made about the augmented system are the same as those for the original system. Theorem 3.4 therefore gives us conditions on the stability of the filter.

We have shown that, in the linear deterministic case, it is possible to construct an observer so that the feedback will ensure the augmented system is stable and the analysed state will converge to the true state as time increases. This is also true for the linear stochastic case, which can be seen by taking the expectation of the true and model equations, as done in Section 3.4. The application of the linear observers to nonlinear systems for the augmented state follows from the arguments given in Section 3.5. We will not elaborate on this here.

The theory of this section has concentrated on the case where the model contains systematic errors. This applies equally to the part of the model which evolves the state variables as for the part which includes the model inputs. A method which allows for systematic errors in the model inputs themselves can be obtained by similar arguments.

5.3 Choosing the gain matrix

We now look at the implications of augmenting the state vector on the various ways of determining the gain matrices, $\tilde{K}_k$, $k = 0, 1, \ldots, N$. The Kalman filter is the optimal solution to the state estimation problem so we first of all show how this relates to the augmented state. This method has been described in [29] in the engineering literature, and in [24] in the meteorological literature. The ideas have been extended here to a more general case with a less specific model for the evolution of the systematic model errors, which allow the error to depend on the state variables and may contain random model error itself. The application of the simpler cases of Optimal Interpolation and Analysis Correction to the augmented state are described subsequently.

We partition the forecast and analysis error covariance matrices as follows

$$\tilde{P}_k^f = \begin{pmatrix} P_k^{x\,f} & P_k^{xb\,f} \\ P_k^{xb\,f^T} & P_k^{b\,f} \end{pmatrix}, \tag{5.3.1}$$

$$\tilde{P}_k^a = \begin{pmatrix} P_k^{x\,a} & P_k^{xb\,a} \\ P_k^{xb\,a^T} & P_k^{b\,a} \end{pmatrix}, \tag{5.3.2}$$

where $P_k^{x\,a/f}$ are the analysis/forecast error covariance matrices for the state vector $\underline{x}_k$, $P_k^{b\,a/f}$ are the analysis/forecast error covariance matrices for the systematic model error vector $\underline{b}_k$, and $P_k^{xb\,a/f}$ are the analysis/forecast error covariance matrices for the cross-correlations between errors in the state and systematic model error vectors. We split the analysis into state and systematic model error parts as

$$\underline{x}_k^a = \underline{x}_k^f + K_k^x[\underline{y}_k - H_k\underline{x}_k^f], \tag{5.3.3}$$

$$\underline{b}_k^a = \underline{b}_k^f + K_k^b[\underline{y}_k - H_k\underline{x}_k^f]. \tag{5.3.4}$$

5.3.1 The Kalman filter

The gain matrices for the Kalman filter are given by

$$K_k^x = P_k^{x\,f} H_k^T [H_k P_k^{x\,f} H_k^T + R_k]^{-1}, \tag{5.3.5}$$

$$K_k^b = P_k^{xb\,f^T} H_k^T [H_k P_k^{x\,f} H_k^T + R_k]^{-1}, \tag{5.3.6}$$

where the forecast error covariances are given by

$$\begin{aligned} P_{k+1}^{x\,f} &= A_k P_k^{x\,a} A_k^T + T_k P_k^{xb\,a^T} A_k + A_k P_k^{xb\,a} T_k^T \\ &\quad + T_k P_k^{b\,a} T_k^T + Q_{k+1}^x, \end{aligned} \tag{5.3.7}$$

$$\begin{aligned} P_{k+1}^{xb\,f} &= A_k P_k^{xb\,a} M_k^T + A_k P_k^{x\,a} N_k^T + T_k P_k^{b\,a} M_k^T \\ &\quad + T_k P_k^{xb\,a^T} N_k^T + Q_{k+1}^{xb}, \end{aligned} \tag{5.3.8}$$

$$\begin{aligned} P_{k+1}^{b\,f} &= M_k P_k^{b\,a} M_k^T + N_k P_k^{x\,a} N_k^T + N_k P_k^{xb\,a^T} M_k \\ &\quad + M_k P_k^{xb\,a^T} N_k^T + Q_{k+1}^b, \end{aligned} \tag{5.3.9}$$

and the analysis error covariances are

$$P_k^{x\,a} = [I - K_k^x H_k] P_k^{x\,f}, \tag{5.3.10}$$

$$P_k^{xb\,a} = [I - K_k^x H_k] P_k^{xb\,f}, \tag{5.3.11}$$

$$P_k^{b\,a} = P_k^{b\,f} - K_k^b H_k P_k^{xb\,f}. \tag{5.3.12}$$

The expressions for the forecast error covariance matrices appear to be very complicated here. The cost is significantly reduced, however, if the model for the evolution of the systematic errors does not depend on the states, i.e. $N_k \equiv 0$, $\forall t_k$.

5.3.2 Optimal Interpolation

In the Optimal Interpolation data assimilation method, the gain matrices for the augmented state are similar to those of the Kalman filter. However, the cost of the method is much smaller due to the fact that the error covariance matrices do not have to be propagated at each time step. The Optimal Interpolation gain matrices are given by

$$K_k^x = P_k^{x\,f} H_k^T [H_k P_k^{x\,f} H_k^T + R]^{-1}, \tag{5.3.13}$$

$$K_k^b = P_k^{xb\,f^T} H_k^T [H_k P_k^{x\,f} H_k^T + R]^{-1}. \tag{5.3.14}$$

Note that the matrix inside the inverse is the same for both gain matrices. Inverting this matrix is one of the major costs of the method and so the systematic model error correction does not impose significant extra cost on the computing time of the data assimilation process. Also, we only need to specify the error covariance matrix for the cross-correlation between the state variables and the systematic model error variables, not the systematic model error covariance matrix itself. However, these cross-correlations might be difficult to specify in practice as statistics of the errors in the systematic model error variables might be difficult to ascertain.

5.3.3 Analysis Correction

The gain matrices for the AC scheme are similar to those of OI, as is to be expected due to the similarity of the two schemes. They can be written as:

$$W_k^x = P_k^{x\,f} H_k^T R_k^{-1}, \tag{5.3.15}$$

$$W_k^b = P_k^{xb\,f} H_k^T R_k^{-1}, \tag{5.3.16}$$

$$V_k \;=\; [W_k^x + I]^{-1}. \tag{5.3.17}$$

Here again, the only new covariance matrix required is the one for the cross-correlation between the state variables and the systematic model error variables. Also, the V_k matrix remains the same as before.

5.4 Systematic model error evolution

An important part of the method of systematic model error correction is the choice of model for the propagation of the systematic model errors from one time step to the next. There are many possible choices for our estimate of the operator, $\underline{m}_k$, which evolves the systematic model errors, and it is useful if something is known about the nature of the systematic model errors *a priori*. Here, we give a number of possible models for certain types of systematic model error but the eventual choice will depend on the situation for which it is to be applied.

Perhaps the simplest situation is when the systematic model errors are constant in time. This type of systematic model error might arise if the errors are due to incorrect forcing or boundary conditions. In this constant bias case, it would be sensible to specify that the systematic model errors are not altered from one time step to the next, unless the data assimilation changes them. We would then write the model for the systematic model error variables as

$$\underline{b}_{k+1} = \underline{b}_k, \qquad k = 0, 1, \ldots, N-1. \tag{5.4.1}$$

Another case is when the systematic model errors are due to truncation errors in the model, or mis-specification of model parameters. The systematic

model errors are now likely to evolve in a similar way to the model variables themselves. It would therefore be a good idea to use the model $\underline{g}_k$ for the evolution of the systematic model errors, i.e.

$$\underline{b}_{k+1} = \underline{g}_k(\underline{b}_k, \underline{u}_k), \qquad k = 0, 1, \ldots, N-1. \tag{5.4.2}$$

If we are dealing with a nonlinear system, the tangent linear model, $G_k \equiv \frac{\partial \underline{g}_k}{\partial \underline{x}_k}|_{\underline{x}_k^f, \underline{u}_k^t}$, might be used instead of the full nonlinear model.

In Section 5.2, it was noted that we could decrease the number of systematic model error variables by including the matrix $T_k \in \mathbb{R}^{n \times q}$, $q < n$. It might also be useful in some cases to allow $q > n$ so that we have more choice in the specification of $\underline{m}_k$. One example of where this might be a good idea is where the model error is varying in time. We could then specify a spectral form of the model error, so that

$$\underline{b}_k = \begin{pmatrix} \underline{b}_k^{(1)} \\ \underline{b}_k^{(2)} \\ \underline{b}_k^{(3)} \end{pmatrix}, \tag{5.4.3}$$

$$T_k = \left(T_k^{(1)}, T_k^{(2)} sin(k/N\tau), T_k^{(3)} cos(k/N\tau) \right), \tag{5.4.4}$$

where $\underline{b}_k \in \mathbb{R}^{3n}$, $T_k \in \mathbb{R}^{n \times 3n}$ and τ is some constant to be specified, depending on the timescale of the variations in the systematic model error.

There are many other possible models for the evolution of the systematic model error variables, including combinations of those given above, [37], [38].

5.5 Experiments

Experiments using systematic model error correction on the damped oscillating system and the Lorenz equations have been carried out with the Optimal Interpolation and Kalman filter assimilation schemes. The results are presented in Sections 5.5.1 and 5.5.2 respectively. The numerical methods used for discretising both the oscillating system and the Lorenz equations, as well as the step sizes used, are the same as those in Sections 4.4.1 and 4.4.2. We also keep the same length of assimilation period and forecast period. In the figures shown, we display three plots for each experiment. One of these contains the true solution (dashed lines), the solution given by the data assimilation scheme (solid lines) and the observations (crosses) for one of the variables, the second shows the error between the data assimilation solution and the true solution for that variable, and the third plot displays the estimate of the systematic model error.

We also present the results of some experiments on the Heat equation. We describe the set-up and results of these experiments in Section 5.5.3.

5.5.1 Oscillating system

In the following experiments we take the true model system, $\underline{f}_k$, to be the same as the model that was used in Section 4.4.1, that is equations (4.4.3) and (4.4.4). The observations are taken to be of the same frequency, in one experiment they are taken of y and x every time unit and in the other they are taken every 2.5 time units. Noise was also added to the observations which had a Gaussian distribution with variance of 0.1 and zero mean.

Bias was added to the true model system to obtain the forecast model system, $\underline{g}_k$, by including a constant forcing term in the discrete equations. Equations (4.4.3) and (4.4.4) are now changed to

$$y_{k+1} = (1 - n(\Delta t)^2/2)y_k + (\Delta t - l(\Delta t)^2/2)x_k + \alpha, \qquad (5.5.1)$$

$$x_{k+1} = (nl(\Delta t)^2/2 - n\Delta t)y_k + (1 - l\Delta t + l^2(\Delta t)^2/2 - n(\Delta t)^2/2)x_k + \beta, \qquad (5.5.2)$$

where the systematic error terms are chosen to be $\alpha = \beta = 0.1$. For the OI data assimilation method, the covariance matrices were calculated by averaging the statistics over the assimilation interval using knowledge of the truth and of the background solution. The model error covariance matrices for use in the Kalman filter were taken to be zero. We only show figures of the y variable here, but the results for the x variable are qualitatively similar.

The effect this change has on the OI and KF data assimilation schemes when no special treatment of the bias is made is shown in Figures 5.5.1 and 5.5.2. In these experiments, there are large oscillations in the analyses during the assimilation period because the forecast model is trying to move to a different position from the true state, but the observations bring the analyses back to near the true solution. Once the forecast is started, the error increases as there is nothing to constrain the solution to the unbiased true solution.

The OI solution, with constant bias correction included, is shown in Figures 5.5.3 and 5.5.4 where observations are available at every time unit and at every 2.5 time units respectively. Here we can see that the bias correction has much improved the analysis of the state for both experiments, although

the solution takes longer to approach the true solution than in Section 4.4.1. The estimate of the bias is fluctuating about the true bias due to the noise in the observations. This makes the forecast less accurate because the estimate of the bias at the end of the assimilation period is included in the model for the forecast. To try to overcome this problem, the same experiments were performed but the bias estimate is averaged over a moving time window of 5 assimilation steps. The results are shown in Figures 5.5.5 and 5.5.6. The bias estimate is now slightly more accurate for the case where observations are available every time unit but not for the case where observations are available every 2.5 time units. We have performed this method on the same model in [63] and the results seem to indicate that better results are obtained when the observations are more frequent.

Experiments were also performed with constant bias correction using the KF, the results of which are shown in Figures 5.5.7 and 5.5.8. In both cases, the KF produces the exact value for the bias and therefore gives the perfect forecast. Less frequent observations lead to a longer convergence time but the forecast is the same in both cases. In all of the above experiments, the initial estimate of the bias was quite poor. This has led to large fluctuations in the state variables at the beginning of the assimilation interval. A better estimate of the bias at the initial time would reduce the size of these fluctuations.

To see how the bias correction performs when a different type of systematic model error is introduced, the parameters l and n are altered to 0.3 and 1.2 respectively in the forecast model system. This means that the model system is oscillating at a different frequency and is damped at a different rate to the true system. The true system equations remain the same and so,

therefore, do the observations. However, for these experiments, the noise on the observations is eliminated. For the first of these experiments, the model for the evolution of the systematic model errors is taken to be constant, i.e. $\underline{b}_{k+1} = \underline{b}_k$. With this constant bias correction, the KF produces a good analysis during the assimilation but the forecast error is large, as shown in Figure 5.5.9. This is due to the fact that the bias estimate is kept constant during the forecast. The forecast can be improved if we use a different model for the evolution of the bias. In Figure 5.5.10, we use the state model to propagate the bias, i.e. $\underline{b}_{k+1} = \underline{g}_k(\underline{b}_k)$. This gives a much better estimate of the true solution, although there is a phase difference which leads to fairly large errors.

Key for Figures 5.5.1-5.5.10: true solution (dashed line), analysed solution (solid line) and observations (crosses).

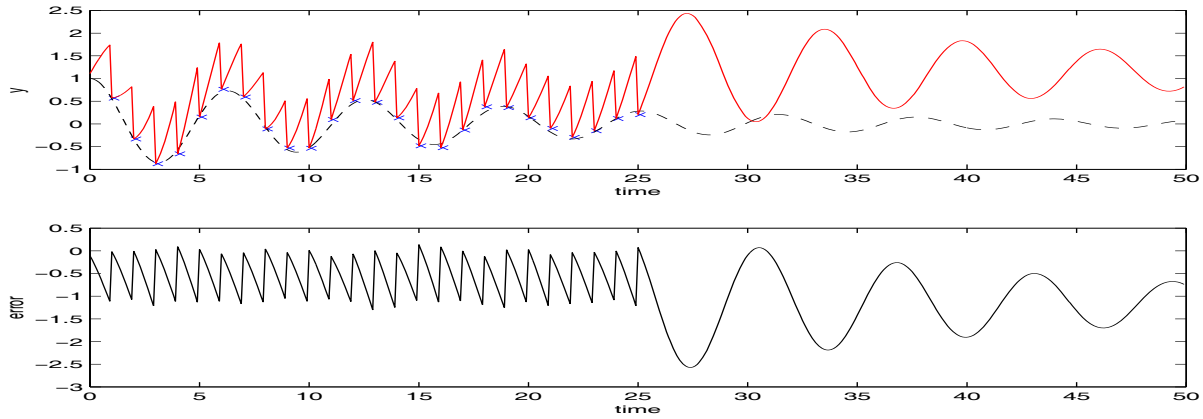

Figure 5.5.1: OI on the oscillating system with observations every time unit. Constant forcing is added to the model and no bias correction is performed. (i) y variable. (ii) Error in y variable.

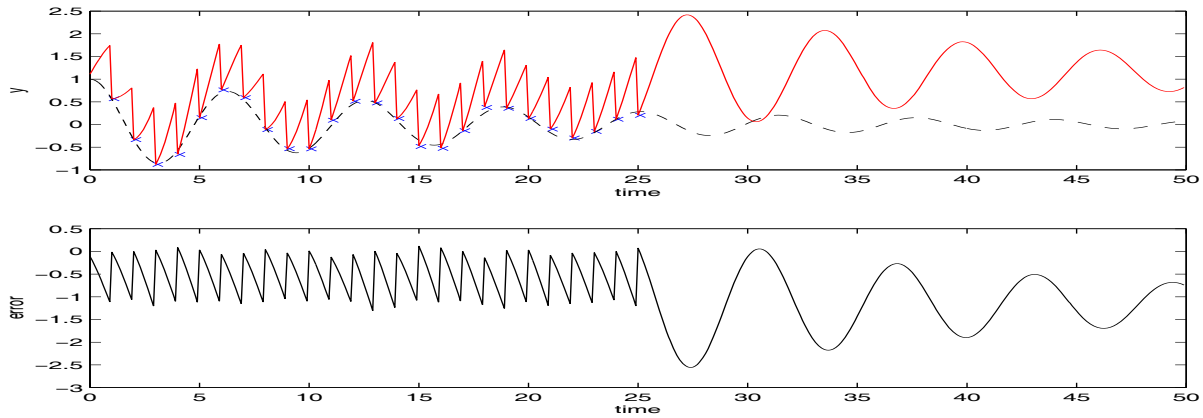

Figure 5.5.2: KF on the oscillating system with observations every time unit. Constant forcing is added to the model and no bias correction is performed. (i) y variable. (ii) Error in y variable.

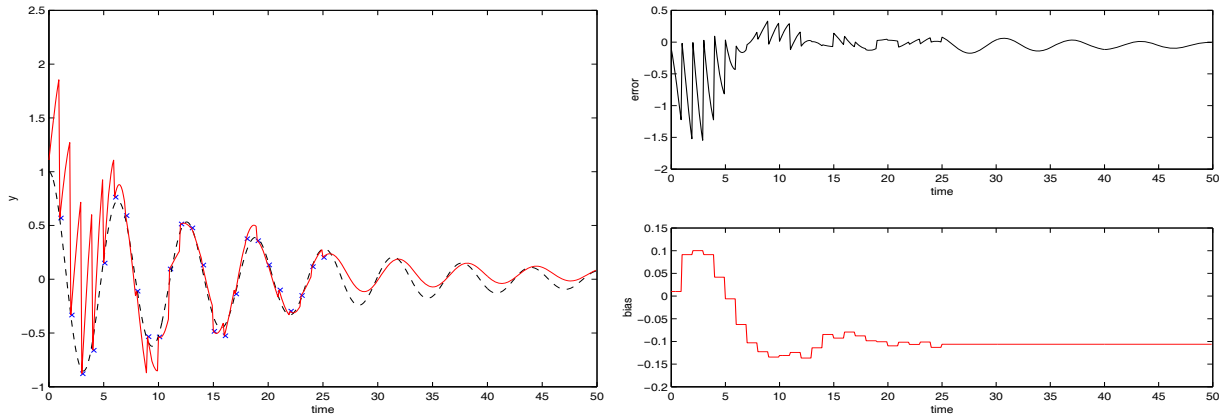

Figure 5.5.3: OI on the oscillating system with observations every time unit. Constant forcing is added to the model and bias correction is performed. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

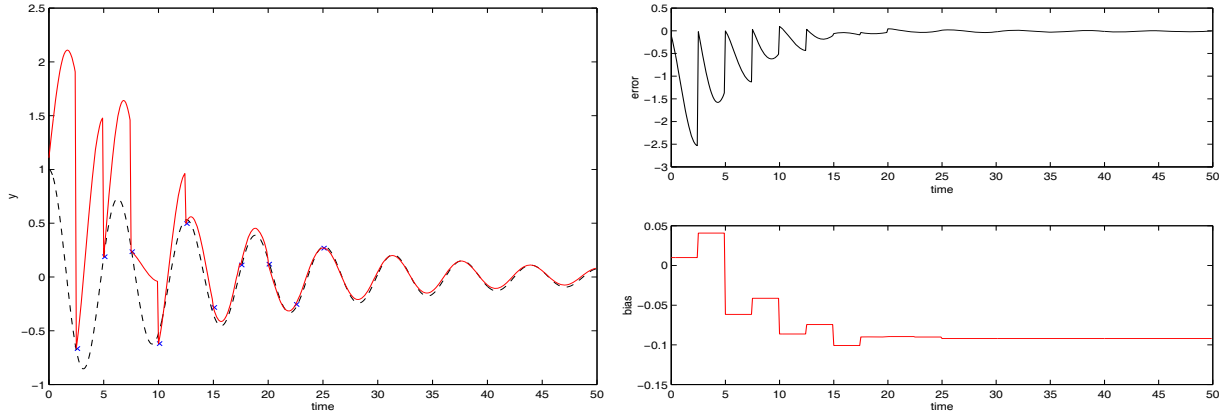

Figure 5.5.4: OI on the oscillating system with observations every 2.5 time units. Constant forcing is added to the model and bias correction is performed. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

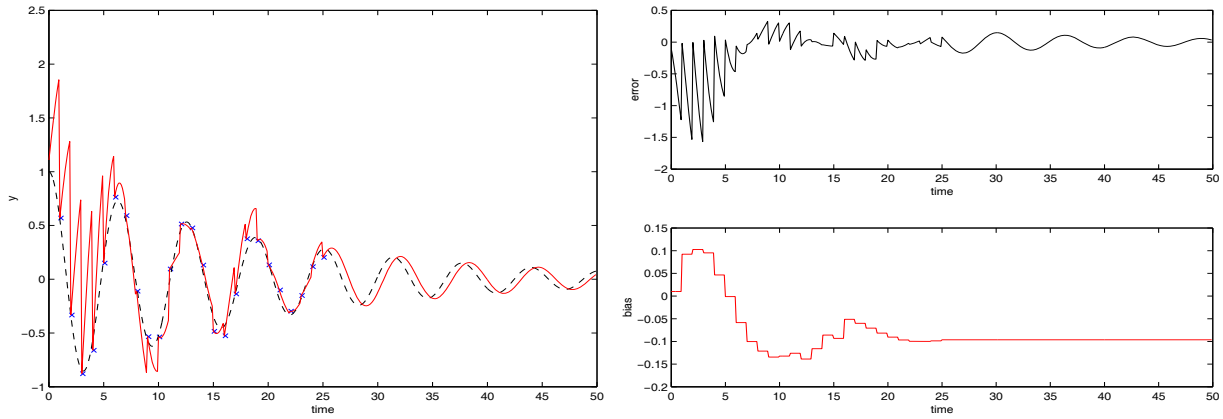

Figure 5.5.5: OI on the oscillating system with observations every time unit. Constant forcing is added to the model and bias correction is performed. Averaging of the bias is done over 5 assimilation steps. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

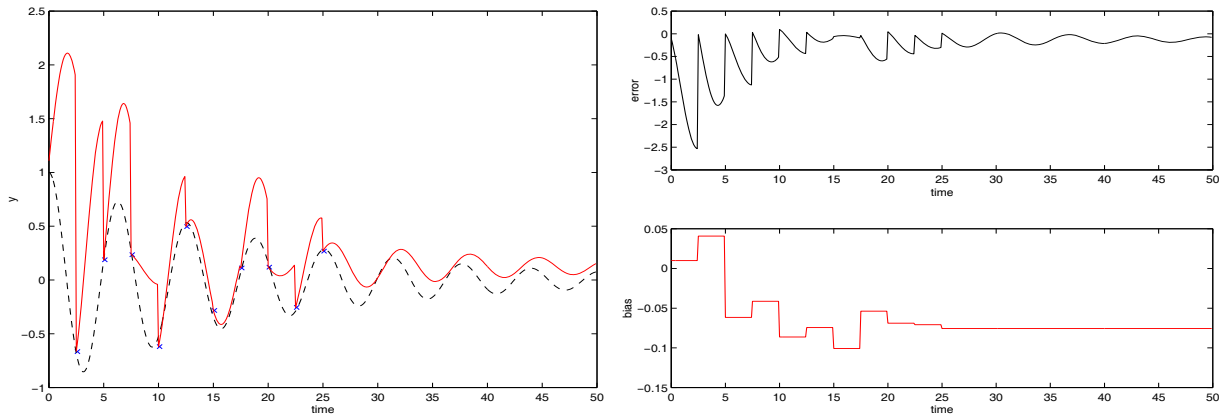

Figure 5.5.6: OI on the oscillating system with observations every 2.5 time units. Constant forcing is added to the model and bias correction is performed. Averaging of the bias is done over 5 assimilation steps. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

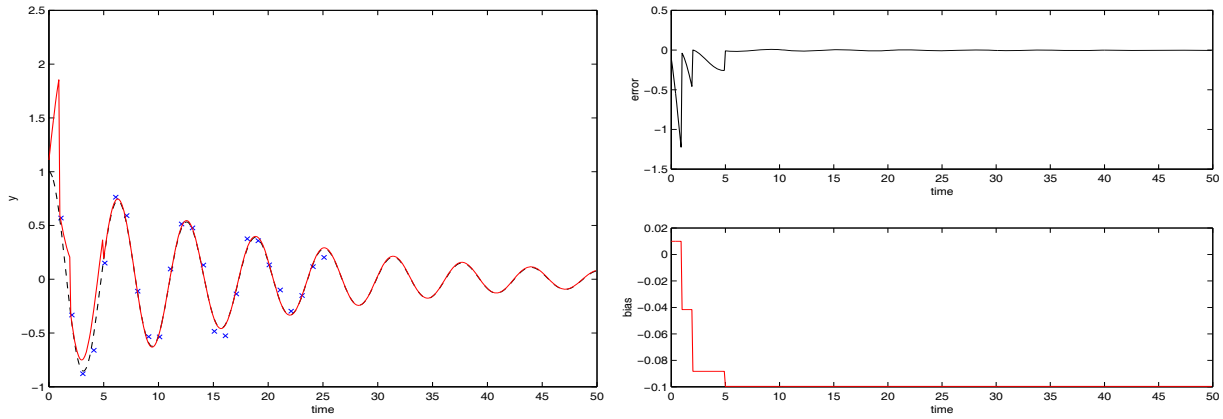

Figure 5.5.7: KF on the oscillating system with observations every time unit. Constant forcing is added to the model and bias correction is performed. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

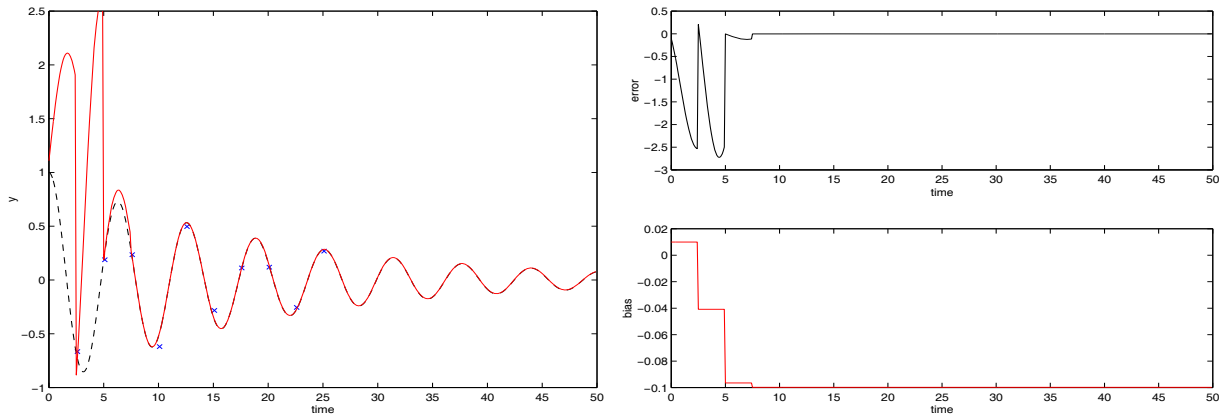

Figure 5.5.8: KF on the oscillating system with observations every 2.5 time units. Constant forcing is added to the model and bias correction is performed. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

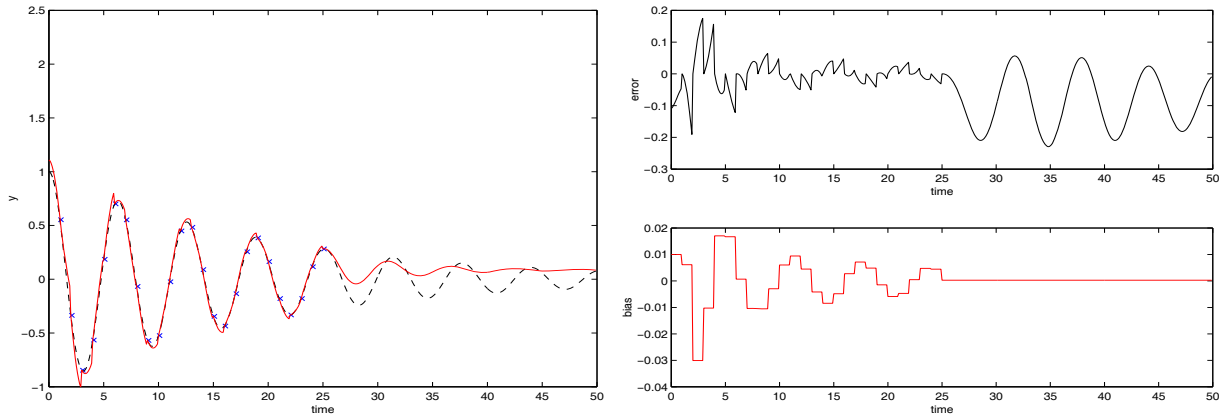

Figure 5.5.9: KF on the oscillating system with observations every time unit. The parameters in the model are altered and a constant bias correction is performed. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

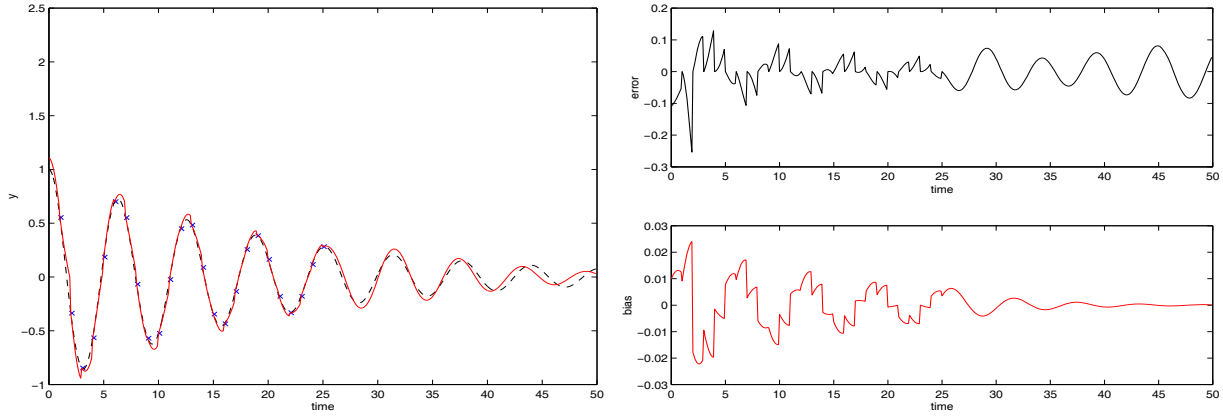

Figure 5.5.10: KF on the oscillating system with observations every time unit. The parameters in the model are altered and the system model is used to forecast the systematic model error. (i) y variable. (ii) Error in y variable. (iii) Analysed estimate of bias in equation (5.5.1).

5.5.2 The Lorenz equations

The method of bias correction has also been applied to the Lorenz system. Here, we take the true system model, $\underline{f}_k$ to be given by the model used in Section 4.4.2, that is equations (4.4.9), (4.4.10) and (4.4.11). The observations are taken of all the variables at every 0.25 time units in one experiment and at every 0.5 time units in the other. Noise was added to the observations which had a Gaussian distribution with variance of two and zero mean. The assimilation is performed over the first 20 time units and the analysis at the end of this period is used as initial conditions for a forecast which lasts for another 10 time units.

A constant bias has been added to each of the discrete equations (4.4.9-4.4.11), to obtain the forecast model system, $\underline{g}_k$. This is given by

$$\begin{aligned} x_{k+1} &= x_k + \sigma\Delta t/2[2(y_k - x_k) + \Delta t(\rho x_k - y_k - x_k y_k) \\ &- \sigma\Delta t(y_k - x_k)] + \alpha, \qquad (5.5.3) \\ y_{k+1} &= y_k + \Delta t/2[\rho x_k - y_k - x_k z_k + \rho(x_k + \sigma\Delta t(y_k - x_k)) - y_k \\ &- \Delta t(\rho x_k - y_k - x_k z_k) - (x_k + \sigma\Delta t(y_k - x_k)) \\ &(z_k + \Delta t(x_k y_k - \beta z_k))] + \beta, \qquad (5.5.4) \\ z_{k+1} &= z_k + \Delta t/2[x_k y_k - \beta z_k + (x_k + \Delta t\sigma(y_k - x_k)) \\ &(y_k + \Delta t(\rho x_k - y_k - x_k z_k)) \\ &- \beta z_k - \Delta t(x_k y_k - \beta z_k)] + \gamma, \qquad (5.5.5) \end{aligned}$$

where $\alpha = 0.05$, $\beta = 0.1$ and $\gamma = 0.15$. The covariance matrices for the OI method were calculated by averaging the statistics over the assimilation period. The model error covariance matrix for the state variables was taken

to be the same as that of Section 4.4.2. The bias model error covariance, Q_k^b and bias-state cross covariances, Q_k^{xb} were taken to be zero. We only show figures for the x variable but the results for the y and z variables are qualitatively similar.

With just the normal OI and EKF data assimilation methods, the analyses during the assimilation period when systematic model errors are present are comparable to those when there were no systematic errors, as shown in Figures 5.5.11 and 5.5.12. However, as soon as the forecast is begun, the solutions converge to a stable equilibrium point which is incorrect, due to the fact that the model is incorrect.

The constant bias correction method was implemented using the OI scheme and the results are shown in Figures 5.5.13 and 5.5.14 for the cases when observations were taken every 0.25 and 0.5 time units respectively. Both these cases show improvements in the forecast. The estimate for the bias in the x variable is quite accurate although there are still large oscillations in both experiments due to the noise on the observations. To try to eliminate these oscillations, similar experiments were performed but with the bias estimate averaged over a 5 assimilation step period, the results of which are shown in Figures 5.5.15 and 5.5.16. The estimates of the bias at the end of the assimilation period improved in both experiments, most noticeably when observations were taken every 0.5 time units. However, due to the chaotic nature of the system, this does not necessarily improve the accuracy of the forecast over a significant time period.

The results of using the constant bias correction with the EKF are shown in Figures 5.5.17 and 5.5.18 for the experiments with the two different ob-

serving frequencies. With observations every 0.25 time units, the analysis is good and the forecast agrees with the true solution for about 5 time units. When observations were available every 0.5 time units however, the solution is worse than that given by OI, due to large variations in the estimate of the bias. The averaging of the bias was again performed but the results were not improved, as can be seen in Figures 5.5.19 and 5.5.20. This could be due to the length of the assimilation interval and the frequency of the observations. We show additional experiments in [63], where observations were taken more frequently and the results were improved.

Key for Figures 5.5.11-5.5.20: true solution (dashed line), analysed solution (solid line) and observations (crosses).

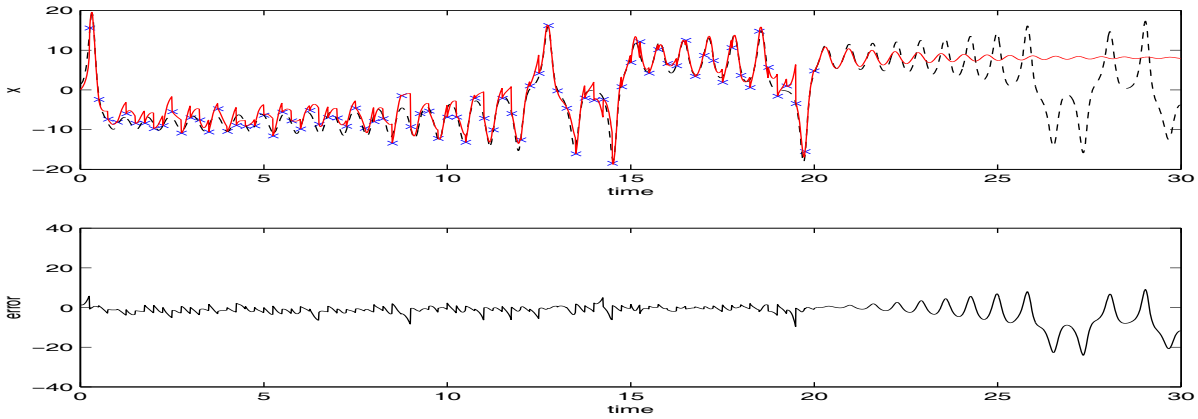

Figure 5.5.11: OI on the Lorenz equations with observations every 0.25 time units. Constant forcing is contained in the model and no bias correction is performed. (i) x variable. (ii) Error in x variable.

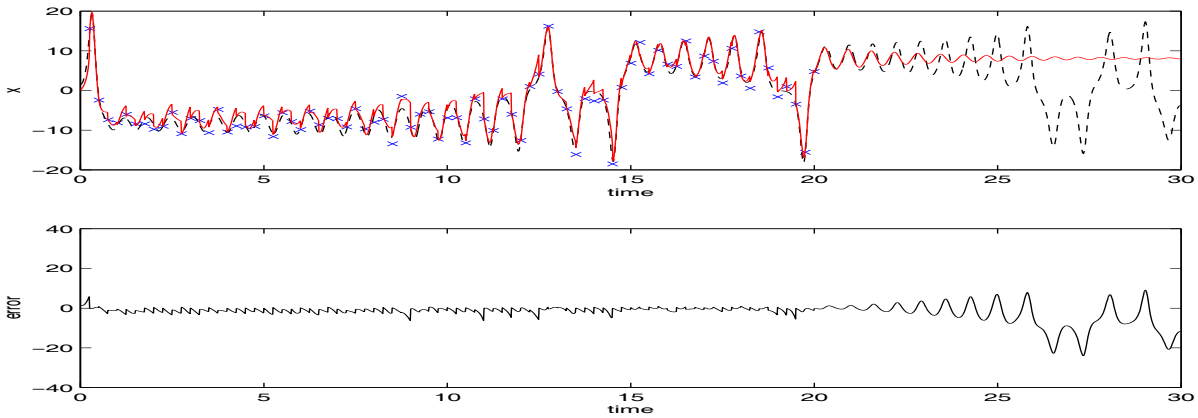

Figure 5.5.12: EKF on the Lorenz equations with observations every 0.25 time units. Constant forcing is contained in the model and no bias correction is performed. (i) x variable. (ii) error in x variable.

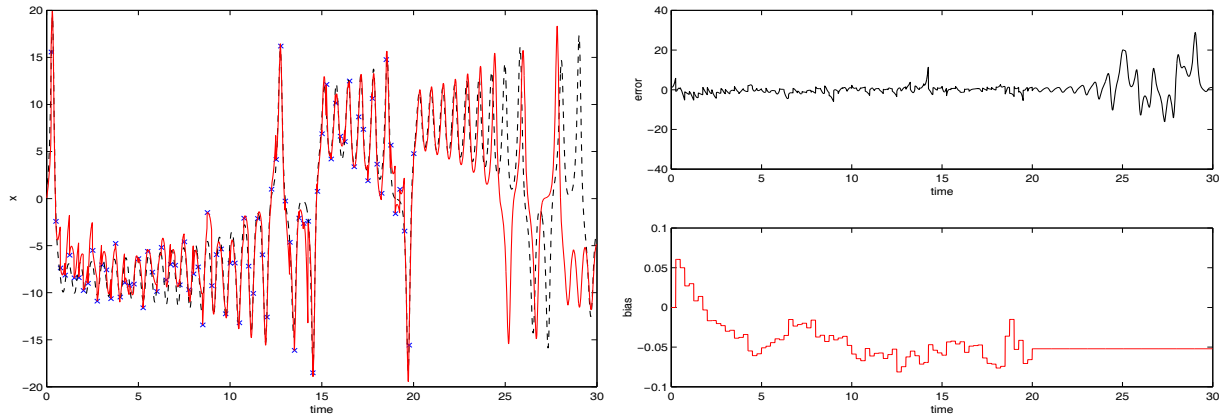

Figure 5.5.13: OI on the Lorenz equations with observations every 0.25 time units. Constant forcing is contained in the model and bias correction is performed. (i) x variable. (ii) Error in x variable. (iii) Analysis of bias in equation (5.5.3).

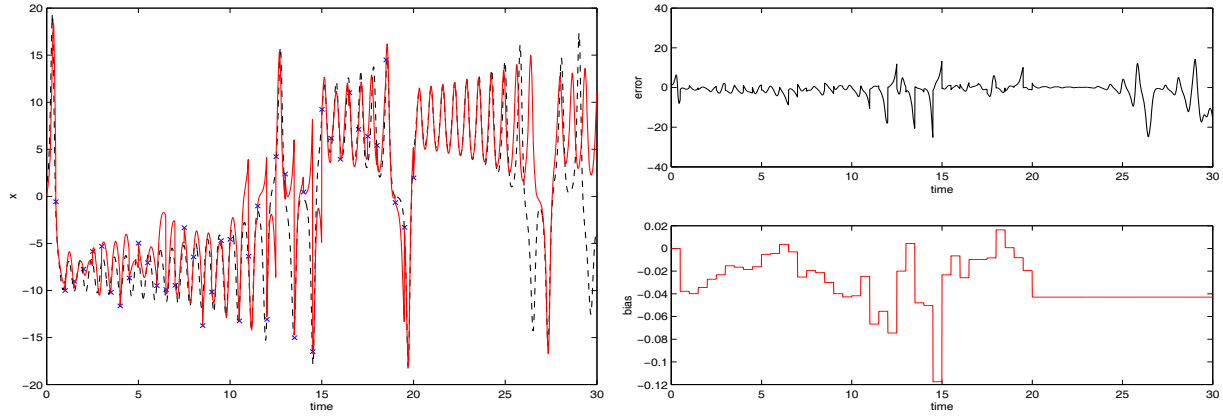

Figure 5.5.14: OI on the Lorenz equations with observations every 0.5 time units. Constant forcing is contained in the model and bias correction is performed. (i) x variable. (ii) Error in x variable. (iii) Analysis bias in equation (5.5.3).

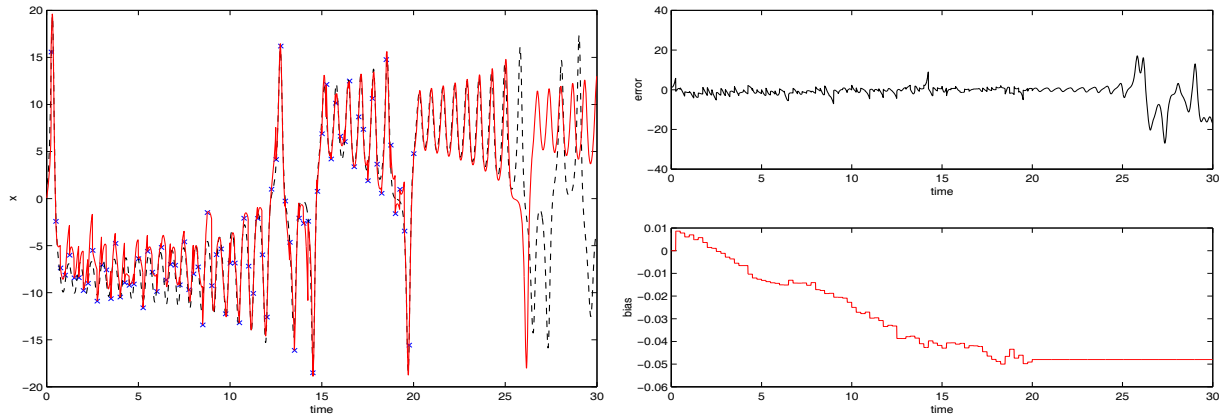

Figure 5.5.15: OI on the Lorenz equations with observations every 0.25 time units. Constant forcing is contained in the model and bias correction is performed. Averaging of the bias is done over 5 assimilation steps. (i) x variable. (ii) Error in x variable. (iii) Analysis of bias in equation (5.5.3).

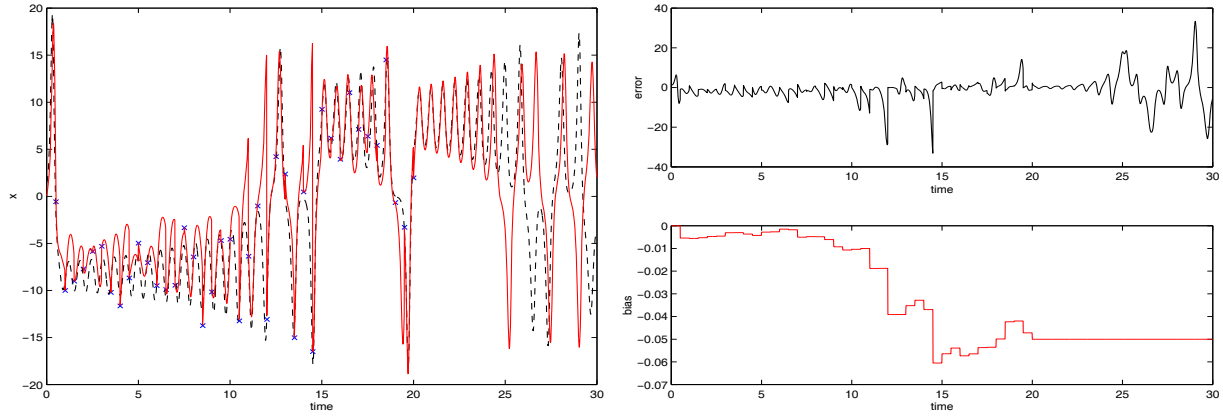

Figure 5.5.16: OI on the Lorenz equations with observations every 0.5 time units. Constant forcing is contained in the model and bias correction is performed. Averaging of the bias is done over 5 assimilation steps. (i) x variable. (ii) Error in x variable. (iii) Analysis of bias in equation (5.5.3).

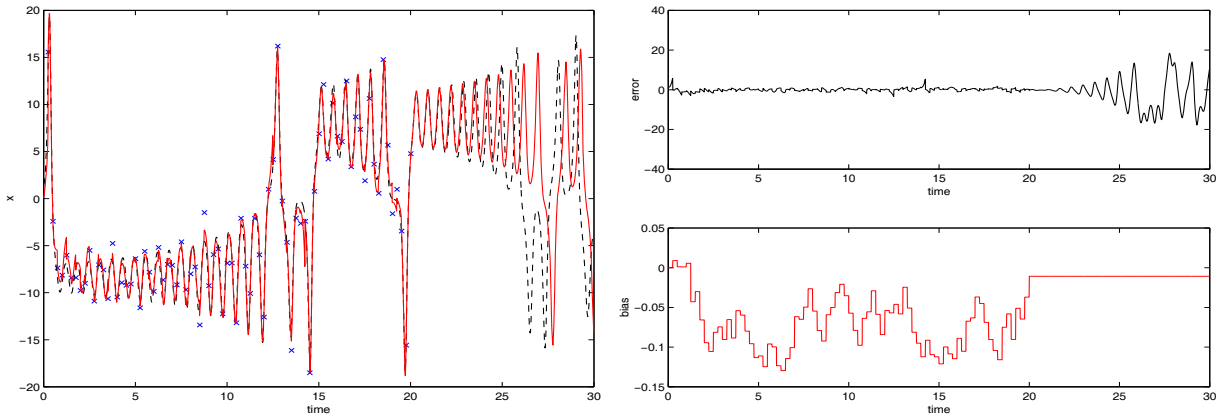

Figure 5.5.17: EKF on the Lorenz equations with observations every 0.25 time units. Constant forcing is contained in the model and bias correction is performed. (i) x variable. (ii) Error in x variable. (iii) Analysis of bias in equation (5.5.3).

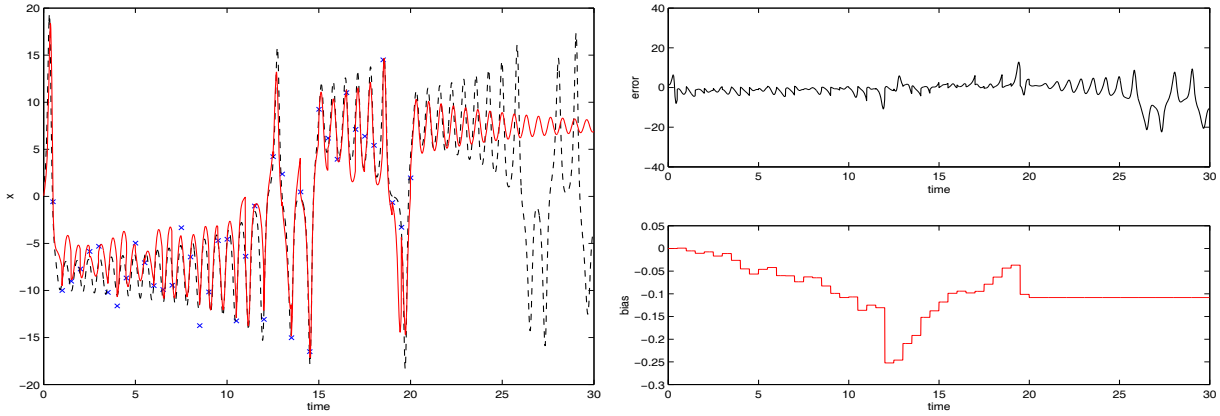

Figure 5.5.18: EKF on the Lorenz equations with observations every 0.5 time units. Constant forcing is contained in the model and bias correction is performed. (i) x variable. (ii) Error in x variable. (iii) Analysis of bias in equation (5.5.3).

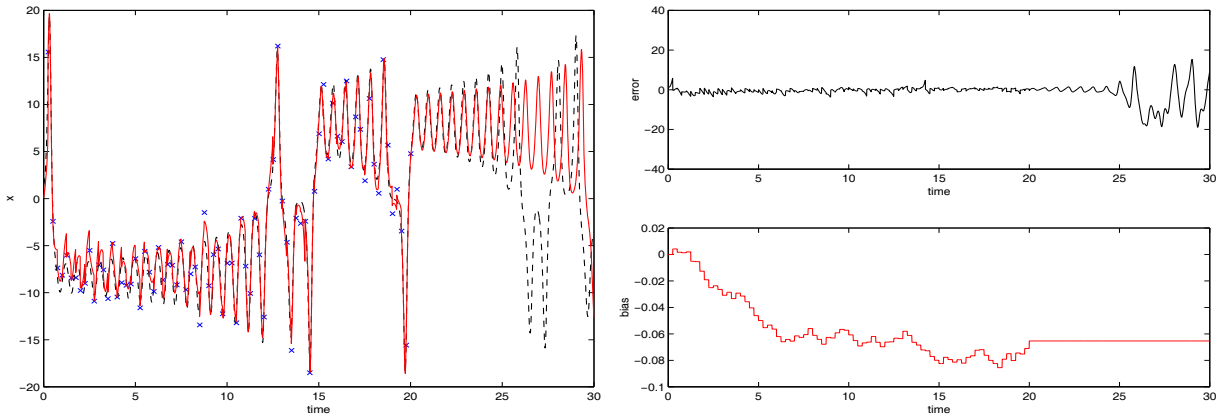

Figure 5.5.19: EKF on the Lorenz equations with observations every 0.25 time units. Constant forcing is contained in the model and bias correction is performed. Averaging of the bias is done over 5 assimilation steps. (i) x variable. (ii) Error in x variable. (iii) Analysis of bias in equation (5.5.3).

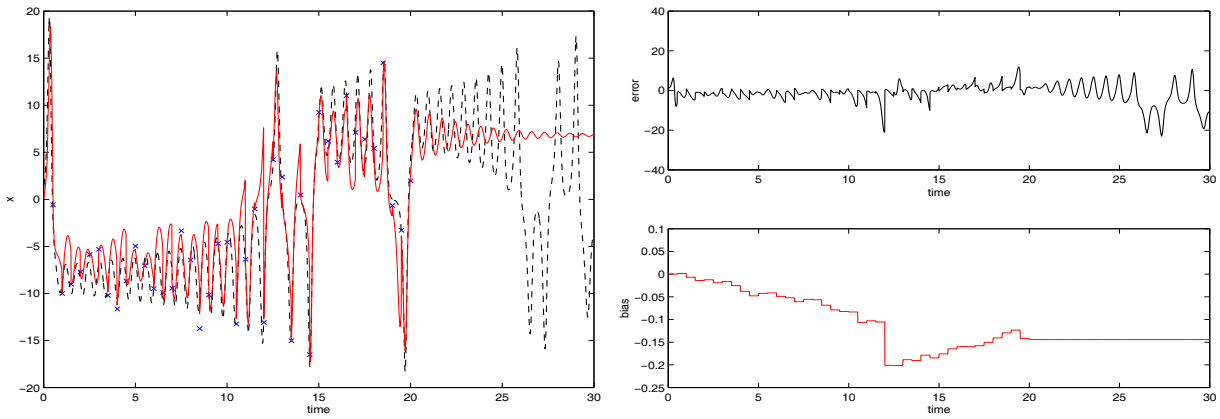

Figure 5.5.20: EKF on the Lorenz equations with observations every 0.5 time units. Constant forcing is contained in the model and bias correction is performed. Averaging of the bias is done over 5 assimilation steps. (i) x variable. (ii) Error in x variable. (iii) Analysis of bias in equation (5.5.3).

5.5.3 Heat equation

In this section, the Heat equation is used to demonstrate how the bias correction method can be used with partial differential equations. The Heat equation is used because it is of a very simple form which is useful when determining which mechanisms are responsible for certain features.

The Heat equation with boundary conditions and initial condition can be written as

$$\begin{aligned} u_t &= \alpha^2 u_{xx}, \quad 0 < x < L, \quad 0 < t \\ u(0,t) &= a, \quad u(L,t) = b, \\ u(x,0) &= f(x). \end{aligned} \tag{5.5.6}$$

The boundary conditions a and b are both given values of 0.5 for the true solution. For the incorrect model solution, $a = 0.5$ and $b = 1.0$. The initial conditions for both solutions are given by

$$f(x) = \begin{cases} 1 & \text{if } \frac{L}{4} < x < \frac{3L}{4} \\ 0 & \text{otherwise.} \end{cases}$$

The steady state solutions are $u(x,t) = \frac{1}{2}$ for the true solution and $u(x,t) = \frac{1}{2} + \frac{x}{2L}$ for the background solution.

The Heat equation has been discretised using an explicit method which gives the discrete equations as

$$\underline{u}_{n+1} = A\underline{u}_n + \underline{r} \tag{5.5.7}$$

where

$$A = \begin{pmatrix} 1-2\nu & \nu & & 0 \\ \nu & 1-2\nu & \nu & \\ & \ddots & \ddots & \ddots \\ 0 & & \nu & 1-2\nu \end{pmatrix},$$

$$\underline{r} = \begin{pmatrix} \nu a \\ 0 \\ \vdots \\ \nu b \end{pmatrix}$$

and

$$\nu = \alpha^2 \frac{\Delta t}{(\Delta x)^2}. \tag{5.5.8}$$

We set $\Delta x = 0.2$ and $\Delta t = 0.005$ which, with $\alpha = 1$, gives a value for ν of 0.125, resulting in a stable numerical scheme. Figures 5.5.21 and 5.5.22 show the true solution and background solution with $\alpha = 1$ and $L = 4$ which are the values these parameters take throughout this section.

Observations are obtained from the true solution and are given everywhere in space and at all times between $t = 0$ and $t = 1$. These observations contain noise, which has a Gaussian distribution with a mean of zero and a variance of 0.1. The data assimilation scheme used in these experiments is the OI method. The assimilation is performed from time $t = 0$ to $t = 1$ and a forecast is then made for a further time unit. The covariance matrices used in the OI scheme were obtained by averaging the statistics over the time of the assimilation. The gain matrix was then calculated using these covariance matrices.

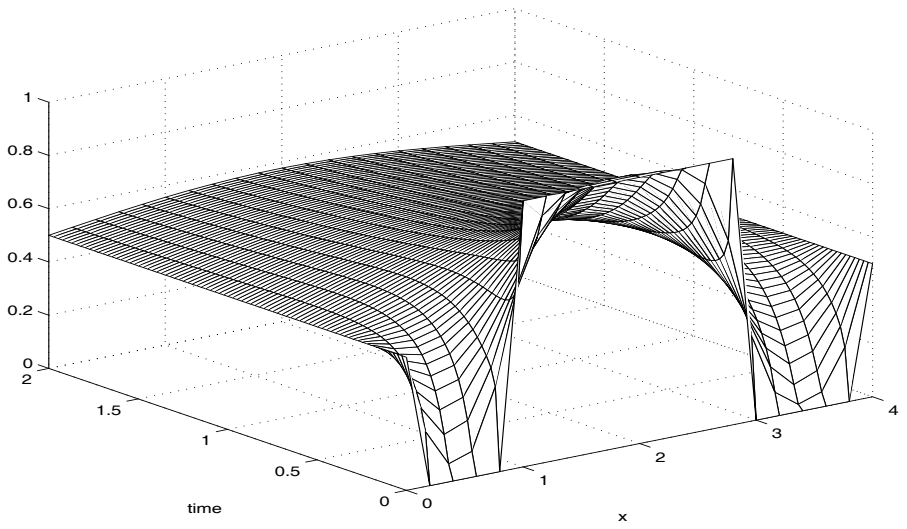

Figure 5.5.21: True solution of the Heat equation.

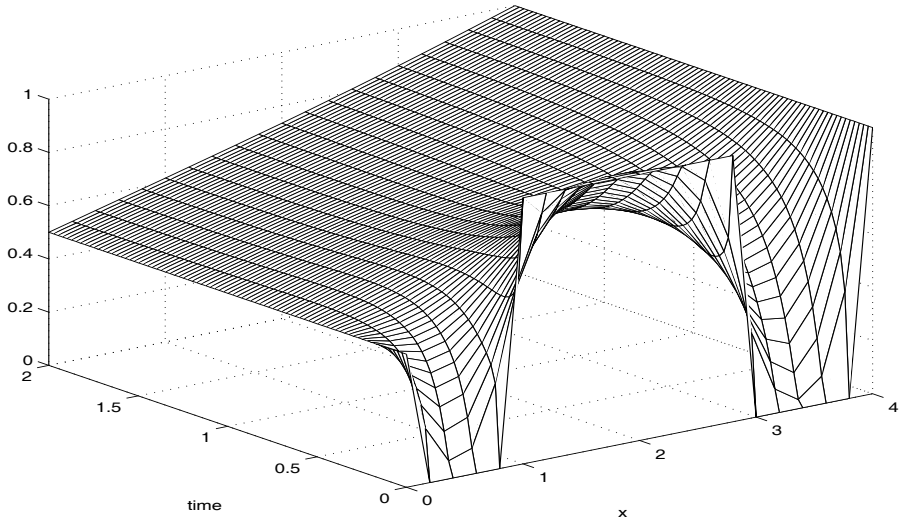

Figure 5.5.22: Background solution of the Heat equation with incorrect boundary condition.

By giving the model solution a different boundary value from the true solution, we are introducing a bias in the model. Therefore, when the analysis at the end of the assimilation period is propagated forward in time, this bias influences the resulting forecast. It is hoped that by using the bias correction term in the assimilation, the forecast will be improved.

The results of four experiments are presented here. The first performs OI on the Heat equation but without any bias correction. Figure 5.5.23 shows that the OI method produces a good analysis during the assimilation period. However, the subsequent forecast is poor because the model contains an incorrect boundary condition. The forecast skews to account for this changed boundary condition and converges to the steady state solution for the model solution. The second experiment implements the bias correction method, where the model for the propagation of the systematic model error is constant, which is the correct model for the bias evolution. When this method is implemented with the correct statistics used in the covariance matrices, the analysis during the assimilation period is good, as shown in Figure 5.5.24. Also, because the correct value was calculated for the bias in the incorrect boundary condition, when this is included in the forecast, we get a very good forecast.

To test the sensitivity of the method with respect to the bias gain matrix, a random perturbation is made to this matrix which has a normal distribution with a mean of zero and variance of 0.03. Here, we obtain the correct boundary condition, but the interior solution is corrupted by noise, as shown in Figure 5.5.25. To account for this, we use the method described in the previous experiments and average the bias over a moving time window of 50

time steps. Figure 5.5.26 shows this solution, which seems to account well for the noise in the bias estimate.

The systematic errors in these experiments only affect one boundary, yet we are performing bias correction over the entire domain. It would be possible to include a matrix, T_k, as in Section 5.2, which would allow us to compute a bias only on the boundary. This would dramatically reduce the computational cost of the method.

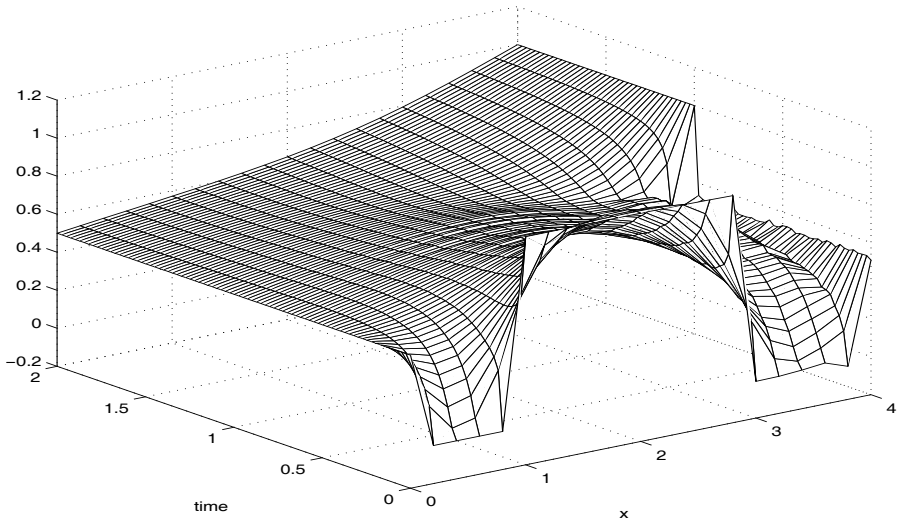

Figure 5.5.23: OI assimilation on the Heat equation without any bias correction.

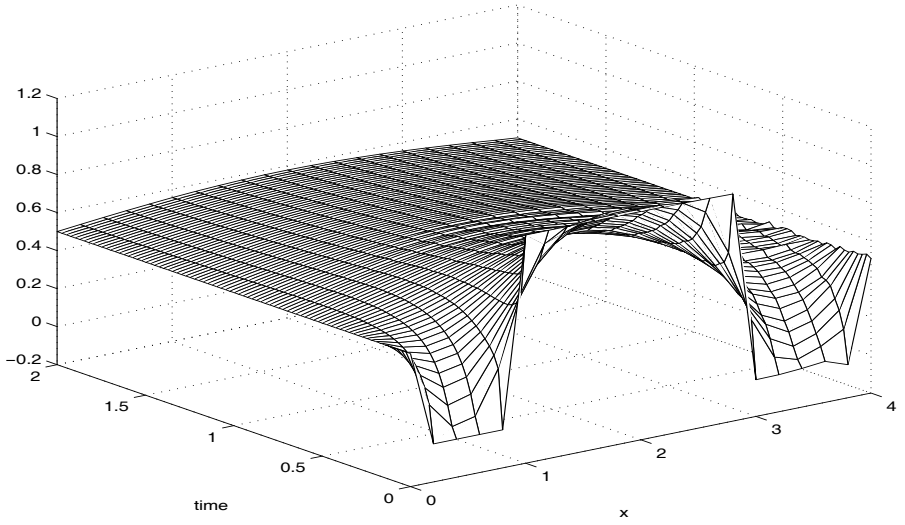

Figure 5.5.24: OI assimilation on the Heat equation with bias correction.

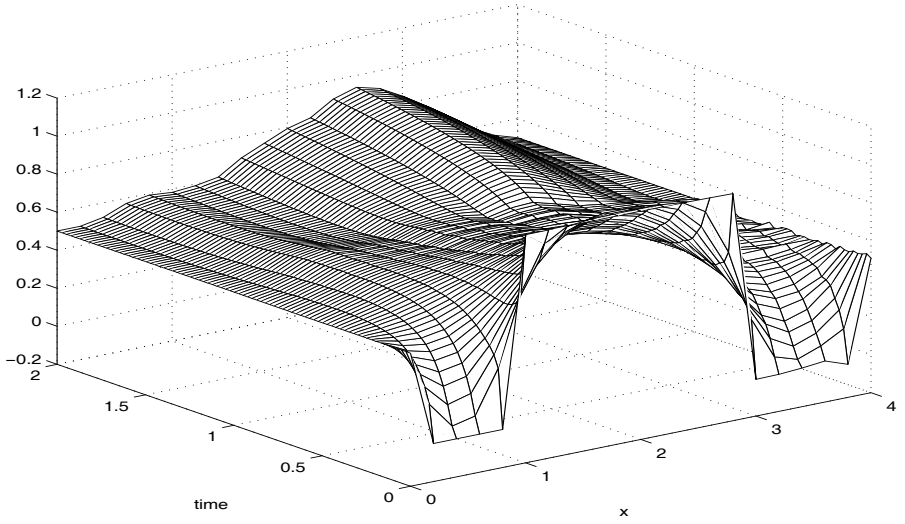

Figure 5.5.25: Same as Figure 5.5.24 but random noise is added to the bias gain matrix.

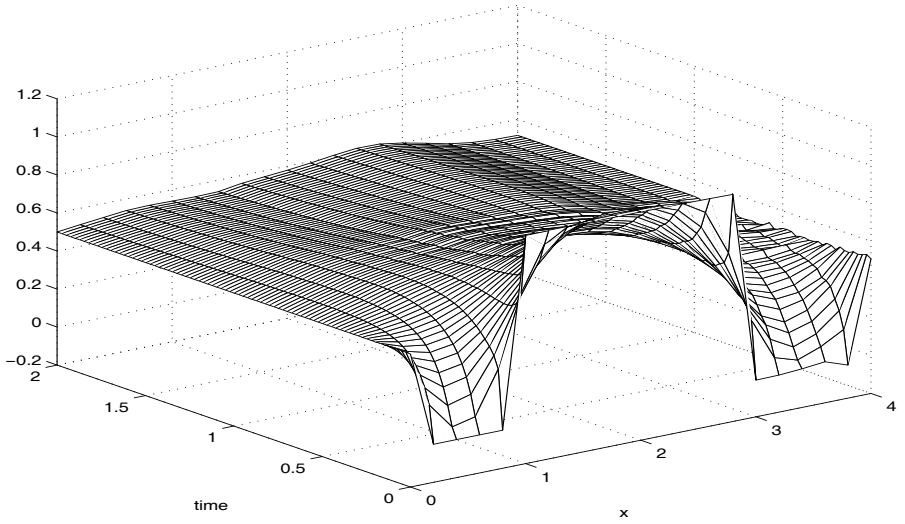

Figure 5.5.26: Same as Figure 5.5.25 except the bias is averaged over a moving window of observations.

5.6 Conclusions

In this chapter, a method for accounting for systematic model error is described. We have shown, in the linear case, that if the correct model for the propagation of the systematic model error is used then both the state and the systematic model error variables should converge to their true values as time increases. The main difficulty with the method is that the model for the propagation of the systematic model error needs to be known *a priori*. Some examples of the types of models which could be used are given. Another difficulty when using the KF is that the covariance matrices for the errors in the systematic model error variables and the cross-correlations between these variables and the state variables need to be known at the initial time. The extra cost of implementing the method in the KF is large, although this could be reduced by using one of the methods described in Section 4.2.4. The extra cost of using the method with the OI and AC schemes is minimal as the matrix inversion involved in computing the gain matrices will have already been performed for the state variables. The only extra information required in these methods is the cross-correlation error covariance matrix, although this might be difficult to ascertain in practice.

Some experiments are presented which perform the systematic model error correction method on two ODEs and one PDE. The experiments with the damped oscillating system show that the bias correction method accounts well for incorrect forcing with a constant bias error. Averaging the bias over a moving time window is shown to produce better initial conditions for a forecast. In these experiments, the KF outperformed OI, as would be expected. When the parameters of the model are altered, the method is shown

to be more accurate if the state model is used to propagate the systematic model error variables. This shows that it would be possible to get an idea of what model should be used for the propagation of the systematic model error variables by implementing the method first of all with a constant bias and observing how the bias evolves in time. It could then be run again with a better guess at the model.

The results for the Lorenz equations are similar to those of the oscillating system for the incorrect forcing, although their chaotic nature leads to a short forecast period for most of the experiments, and the difference between the EKF and OI schemes is diminished. The frequency of the observations appears to be an important factor in the accuracy of the assimilation in these experiments.

The Heat equation experiments show that another type of systematic model error can also be accounted for using a constant bias correction, that is incorrect boundary conditions. In this case, the correct model for the evolution of the bias variables was used. An accurate estimate for the bias gain matrix is shown to be needed here, although averaging of the bias does help when the statistics are incorrect.

The experiments have given us some valuable insight into the advantages and disadvantages of the systematic model error correction method. The main advantage is that systematic model error can be accounted for given certain information. Also, we have an estimate of the systematic model error itself. This could be very important as it would enable a better understanding of the deficiencies in the model, so as to allow some improvements in the modelling of the system under consideration. The main disadvantage, or

difficulty, is that the error covariances need to be known fairly accurately and the information required is not readily available unless a good estimate of the true state of the system exists, which is not necessarily the case in the context of the ocean.

Chapter 6

Applying Bias Correction to Shallow Water Equations

6.1 Introduction

We have seen in the previous chapter that it is possible to account for certain types of systematic model error by using the data. Some simple examples were shown. In reality, the type of model error we are dealing with will vary according to the situation we are in. Here, we concentrate on the equatorial Pacific ocean and try to deal with a particular type of systematic error.

It is known that the forcing of ocean models plays an important role in the types of circulations we observe. As was discussed in Chapter 2, wind forcing near the equator explains much of the structure of the ocean there. Systematic errors in the winds supplied to an ocean model can therefore prove to be problematic when assimilating data.

In this chapter, we compare some different methods for accounting for

incorrect wind forcing and produce a method which will be appropriate for use in a full Primitive equation model, i.e. computationally inexpensive. To analyse the effects of these methods, a linear two-dimensional model is derived in Section 6.2 which retains some of the significant characteristics of the real ocean. Steady state and transient solutions of these equations are discussed. We would also like to examine the effects of our methods using a numerical model. The numerical method and assimilation scheme are therefore described in Sections 6.4.1 and 6.4.2. Experiments and results are given in Section 6.5 with some conclusions in Section 6.6.

6.2 Correcting for incorrect wind forcing

To examine, in a continuous framework, the effects of different methods for accounting for systematic errors in wind forcing, we return to the set of linear equations given in Section 2.4, i.e. equations (2.4.2 - 2.4.6). Throughout this section we assume that the true state of the ocean is given by a reference state u_0, v_0, w_0, ρ_0 with some perturbations, denoted by superscript t, which are given by

$$\rho_0(\frac{\partial u^t}{\partial t} - fv^t) = -\frac{\partial p^t}{\partial x} + \frac{\partial \tau^{x\,t}}{\partial z}, \tag{6.2.1}$$

$$\rho_0(\frac{\partial v^t}{\partial t} + fu^t) = -\frac{\partial p^t}{\partial y} + \frac{\partial \tau^{y\,t}}{\partial z}, \tag{6.2.2}$$

$$\frac{\partial p^t}{\partial z} = -\rho^t g, \tag{6.2.3}$$

$$\frac{\partial u^t}{\partial x} + \frac{\partial v^t}{\partial y} + \frac{\partial w^t}{\partial z} = 0, \tag{6.2.4}$$

$$\frac{\partial \rho^t}{\partial t} + w^t \frac{\partial \rho_0}{\partial z} = 0, \tag{6.2.5}$$

where $\underline{\tau}^t = (\tau^{x\,t}, \tau^{y\,t})$ is the true wind forcing. As we are dealing with the circulation near the equator, we make the β-plane approximation, where $f = \beta y$. These equations allow planetary, gravity and Kelvin waves, as described in Section 2.4.1, which occur in the real ocean.

6.2.1 Data assimilation in the presence of incorrect wind forcing

We now write down the equations for the evolution of our model variables. These model variables are assumed to consist of the same reference solution as the true model, but with different perturbations, denoted by superscript m. The model equations are then written as

$$\rho_0(\frac{\partial u^m}{\partial t} - fv^m) = -\frac{\partial p^m}{\partial x} + \frac{\partial \tau^{x\,m}}{\partial z}, \tag{6.2.6}$$

$$\rho_0(\frac{\partial v^m}{\partial t} + fu^m) = -\frac{\partial p^m}{\partial y} + \frac{\partial \tau^{y\,m}}{\partial z}, \tag{6.2.7}$$

$$\frac{\partial p^m}{\partial z} = -\rho^m g, \tag{6.2.8}$$

$$\frac{\partial u^m}{\partial x} + \frac{\partial v^m}{\partial y} + \frac{\partial w^m}{\partial z} = 0, \tag{6.2.9}$$

where $\underline{\tau}^m = (\tau^{x\,m}, \tau^{y\,m})$ is the incorrect model wind forcing. We modify equation (6.2.5) by introducing a forcing term which represents the data assimilation. For simplicity, we assume that we have complete coverage of observations of the true density field, ρ^t, so that

$$\frac{\partial \rho^m}{\partial t} + w^m \frac{\partial \rho_0}{\partial z} = -\epsilon(\rho^m - \rho^t), \tag{6.2.10}$$

where $\epsilon > 0$ is some coefficient corresponding to the weight given to the observations. This form of data assimilation is known as dynamic relaxation,

[21], [23]. In the discrete version of this equation, this form of the data assimilation corresponds to a gain matrix in the data assimilation which is ϵ multiplied by the identity matrix, i.e. $K^\rho = \epsilon I$. The observation operator, H^ρ, is now equal to the identity matrix, i.e. $H^\rho = I$.

To examine the effects this assimilation has on the error in our model in the presence of incorrect wind forcing, we take the difference between equations (6.2.1-6.2.5) and (6.2.6-6.2.10). We denote the difference between the model and true variables with a prime and obtain

$$\rho_0(\frac{\partial u'}{\partial t} - fv') = -\frac{\partial p'}{\partial x} + \tau'^x, \tag{6.2.11}$$

$$\rho_0(\frac{\partial v'}{\partial t} + fu') = -\frac{\partial p'}{\partial y} + \tau'^y, \tag{6.2.12}$$

$$\frac{\partial p'}{\partial z} = -\rho' g, \tag{6.2.13}$$

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0, \tag{6.2.14}$$

$$\frac{\partial \rho'}{\partial t} + w'\frac{\partial \rho_0}{\partial z} = -\epsilon\rho', \tag{6.2.15}$$

assuming the true and model variables have the same reference state. We then split the solution into a depth dependent part and a part which varies in the horizontal and with time, as in equations (2.4.7-2.4.8), and concentrate on the solution of this latter component. The horizontal structure then satisfies

$$\frac{\partial \hat{u}}{\partial t} - f\hat{v} = -g\frac{\partial \hat{p}}{\partial x} + \tau^x, \tag{6.2.16}$$

$$\frac{\partial \hat{v}}{\partial t} + f\hat{u} = -g\frac{\partial \hat{p}}{\partial y} + \tau^y, \tag{6.2.17}$$

$$\frac{\partial \hat{p}}{\partial t} + \epsilon\hat{p} + H_\epsilon(\frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y}) = 0, \tag{6.2.18}$$

where $\underline{\tau} = (\tau^x, \tau^y)$ is the difference between the true and model wind forcing and the hat over the variables indicates the horizontal structure of the

differences between the true and model solutions.

We now look at the steady state solution of these equations, where we set the time derivatives to zero. It is obvious that with the incorrect wind forcing present, none of the variables needs to be zero, which is one of the requirements for our model to converge to the true solution. This shows that, even in this simple model, the normal data assimilation method will not produce realistic results when systematic errors are present in the wind forcing. We therefore need to modify the data assimilation in some way.

6.2.2 Bias correction

The method we propose to use to account for incorrect forcing is similar to the methods described in the previous chapter. Here, we only have incorrect forcing on two of the equations and so add some bias variables on these equations only, so that equations (6.2.16) and (6.2.17) become

$$\frac{\partial \hat{u}}{\partial t} - f\hat{v} = -g\frac{\partial \hat{p}}{\partial x} + \tau^x + b^x, \tag{6.2.19}$$

$$\frac{\partial \hat{v}}{\partial t} + f\hat{u} = -g\frac{\partial \hat{p}}{\partial y} + \tau^y + b^y. \tag{6.2.20}$$

In the discrete version of these equations, this would be equivalent to including a matrix, T, as in Section 5.2.1, which enables the systematic model error variables to act only on certain model equations. The biases b^x and b^y therefore correspond to entries in the single model error vector, $\underline{b}$.

We now want to calculate b^x and b^y so that the perturbation pressure and velocities are zero in the steady state. We also require that the transient solutions will decay in time so that the model variables will tend towards the true solution as time increases. If we substitute equations (6.2.19) and

(6.2.20) into (6.2.18), and set the time derivatives to zero, we obtain

$$\epsilon\hat{p} + \frac{H}{f}\left\{(-g\frac{\partial^2\hat{p}}{\partial x\partial y} + \frac{\partial\tau^y}{\partial x} + \frac{\partial b^y}{\partial x}) + (g\frac{\partial^2\hat{p}}{\partial y\partial x} - \frac{\partial\tau^x}{\partial y} - \frac{\partial b^x}{\partial y} - \beta\hat{v})\right\} = 0, \tag{6.2.21}$$

where $\epsilon > 0$. So for $\hat{p} = 0$, we require that

$$\frac{\partial\tau^y}{\partial x} + \frac{\partial b^y}{\partial x} = \frac{\partial\tau^x}{\partial y} + \frac{\partial b^x}{\partial y} + \beta\hat{v}. \tag{6.2.22}$$

Differentiating and combining equations (6.2.19) and (6.2.20) and using relationship (6.2.22), we obtain an equation for the steady solution for $\hat{u}$ and $\hat{v}$,

$$\beta y(\frac{\partial\hat{u}}{\partial x} + \frac{\partial\hat{v}}{\partial y}) + \beta\hat{v} = 0. \tag{6.2.23}$$

From equation (6.2.18) with $\hat{p} = 0$, we can see that the horizontal divergence of the velocity field must be zero, i.e. $\frac{\partial\hat{u}}{\partial x} + \frac{\partial\hat{v}}{\partial y} = 0$. Equation (6.2.23) therefore tells us that $\hat{v} = 0$ which in turn implies that $\frac{\partial\hat{u}}{\partial x} = 0$. If we have suitable boundary conditions, we then obtain $\hat{u} = 0$. So, if we choose b^x and b^y so that (6.2.22) holds, the steady state solutions for $\hat{p}$, $\hat{u}$ and $\hat{v}$ will be zero, as required.

We now have the choice of how to choose the form of b^x and b^y. An important consideration for any method that is to be used in a full Primitive equation model is that it is easy to implement and is not computationally expensive. We would therefore like to find an expression for these bias variables so that these conditions are met. Also, we do not want to add unrealistic circulations when performing the bias correction. A scale analysis shows that the acceleration and coriolis terms in the momentum equations are much smaller than the pressure gradients and wind stresses in the zonal direction

in the equatorial Pacific. The main balance close to the equator is therefore between these latter two terms although this is not true away from the equator. We therefore choose to correct for inaccuracies in the wind stresses near the equator by adding a correction of the same form as the pressure gradients, that is

$$b^x = -g\frac{\partial p^c}{\partial x}, \tag{6.2.24}$$

$$b^y = -g\frac{\partial p^c}{\partial y}, \tag{6.2.25}$$

where p^c is some 'pressure correction' field. This form of correction corresponds to assuming that the differences between the true and incorrect model wind forcings have zero curl, as can be seen from equation (6.2.22). The horizontal momentum equations now become

$$\frac{\partial \hat{u}}{\partial t} - f\hat{v} = -g\frac{\partial(\hat{p}+p^c)}{\partial x} + \tau^x, \tag{6.2.26}$$

$$\frac{\partial \hat{v}}{\partial t} + f\hat{u} = -g\frac{\partial(\hat{p}+p^c)}{\partial y} + \tau^y. \tag{6.2.27}$$

Once we have calculated our pressure correction field, we can add it to the pressure field when calculating the momentum of $\hat{u}$ and $\hat{v}$ and so do not have to alter the computer code for these equations. We also only have to calculate and store one field, p^c, when accounting for inaccuracies in both τ^x and τ^y. This choice of b^x and b^y is therefore cheap and easy to implement, but does have the drawback that it does not ensure that $\hat{u}$ and $\hat{v}$ are zero in the steady state unless the curl of the errors in the wind stress is zero, i.e. $\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y} = 0$. This is a drawback since the wind stress near the equator is much stronger in the x direction and so the curl of the wind stess is unlikely to be close to zero here. Away from the equator, the form of the pressure

correction is more likely to account for errors in the wind stress as it has a more rotational field in higher latitudes.

In the steady state, the solution with the pressure correction is given by

$$\hat{p} = 0, \tag{6.2.28}$$

$$\frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y} = 0, \tag{6.2.29}$$

which shows, from equation (6.2.14), that $\frac{\partial \hat{w}}{\partial z} = 0$.

To examine the time dependent solutions, we require an equation for the evolution of the pressure correction term. We model this in a similar way to the augmented data assimilation as

$$\frac{\partial p^c}{\partial t} = \gamma(p^m - p^t), \tag{6.2.30}$$

where $\gamma > 0$ is some weighting coefficient and p^t are observed values of the true pressure field. This model for the evolution of the pressure correction term is equivalent to keeping it constant in time, except when the difference between the model solution and the observed true solution is non-zero, in which case the pressure correction will be forced by a weighting of that difference. When the model's pressure field has a larger magnitude than that of the true pressure field, the pressure correction term will be postitive and so the total pressure will be even larger to compensate for the wind stress errors. This form should therefore restore the balance between the pressure gradient and the incorrect wind stress and therefore reduce the magnitude of the velocity fields.

6.3 Analysis of pressure correction method

In this section we have two main aims. One is to show the types of structure which the pressure correction method, as described above, will introduce into the solution of our linear equations. The second is to show that the method is stable, and to find the timescales at which the solution will converge.

The time dependent solutions for the errors in our model in the horizontal, when the pressure correction is included, satisfy

$$\frac{\partial \hat{u}}{\partial t} - f\hat{v} = -g\frac{\partial(\hat{p}+p^c)}{\partial x} + \tau^x, \tag{6.3.1}$$

$$\frac{\partial \hat{v}}{\partial t} + f\hat{u} = -g\frac{\partial(\hat{p}+p^c)}{\partial y} + \tau^y, \tag{6.3.2}$$

$$\frac{\partial \hat{p}}{\partial t} + \epsilon\hat{p} + H_e(\frac{\partial \hat{u}}{\partial x} + \frac{\partial \hat{v}}{\partial y}) = 0, \tag{6.3.3}$$

$$\frac{\partial p^c}{\partial t} = \gamma\hat{p}. \tag{6.3.4}$$

We want to combine these equations to obtain one equation in $\hat{v}$ so that we can derive the solution, from which the solution for $\hat{u}$, $\hat{p}$ and p^c can be inferred. We obtain

$$\begin{aligned} c^2(\frac{\partial}{\partial t} &+ \gamma)(\frac{\partial}{\partial t}(\nabla^2\hat{v}) + \beta\frac{\partial \hat{v}}{\partial x}) - \frac{\partial}{\partial t}(\frac{\partial}{\partial t}+\epsilon)(\frac{\partial^2\hat{v}}{\partial t^2} + f^2\hat{v}) \\ = c^2(\frac{\partial}{\partial t} &+ \gamma)\frac{\partial}{\partial x}(\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y}) - \frac{\partial}{\partial t}(\frac{\partial}{\partial t}+\epsilon)(\frac{\partial \tau^y}{\partial t} - f\tau^x), \end{aligned} \tag{6.3.5}$$

where $c^2 = gH_e$. The derivation of this equation is described in Appendix A. To investigate the stability of the system described above, we examine the homogeneous equations, where the wind forcing is omitted. This leads to an equation for $\hat{v}$ similar to equation (6.3.5) where the right hand side is zero,

$$c^2(\frac{\partial}{\partial t}+\gamma)(\frac{\partial}{\partial t}(\nabla^2\hat{v}) + \beta\frac{\partial \hat{v}}{\partial x}) - \frac{\partial}{\partial t}(\frac{\partial}{\partial t}+\epsilon)(\frac{\partial^2\hat{v}}{\partial t^2} + f^2\hat{v}) = 0. \tag{6.3.6}$$

We now follow the same strategy as in Section 2.4.1 and look for solutions of the form,

$$\hat{v} = v(y)exp(ikx - i\omega t). \tag{6.3.7}$$

As we are dealing with the errors between the true and analysed solutions, we would like the transient solutions to decay to zero as time increases. This will only happen if the imaginary part of ω is negative; the real part gives us information about the structure of the solutions. Substituting this expression for $\hat{v}$ into equation (6.3.6) yields the following ordinary differential equation for $v(y)$,

$$\frac{d^2v}{dy^2} + \{(\frac{\omega + i\epsilon}{\omega + i\gamma})(\frac{\omega^2 - \beta^2y^2}{c^2}) - k^2 - \frac{\beta k}{\omega}\}v = 0. \tag{6.3.8}$$

We now rescale by introducing a new variable, $y = \Omega\zeta$, where

$$\Omega^2 = \frac{c}{\beta}\left(\frac{\omega + i\gamma}{\omega + i\epsilon}\right)^{1/2}, \tag{6.3.9}$$

which leads to a familiar form of equation,

$$\frac{d^2v}{d\zeta^2} + (\lambda - \zeta^2)v = 0. \tag{6.3.10}$$

The dispersion relation is now more complicated than in Section 2.4.1 and is given by,

$$\lambda = \frac{\omega^2}{\Omega^2\beta^2} - \Omega^2(k^2 + \frac{\beta k}{\omega}). \tag{6.3.11}$$

To show that the system is stable, we need to show that $Im(\omega) < 0$. To do this, we first need to find the values of λ for which the differential equation (6.3.10) has bounded solutions. This is done in the next section. We then need to find the imaginary values of ω which satisfy equation (6.3.11) for these values of λ. We would also like to find the types of structure of the solution of v, which depend on the real part of Ω^2.

Solving the differential equation

We first of all solve equation (6.3.10) for $v(\zeta)$. To do this, we look for solutions of the form,

$$v = v_0 exp(\pm\frac{\zeta^2}{2})J(\zeta), \tag{6.3.12}$$

where $J(\zeta)$ is some polynomial. When we use the plus sign in this expression, the solution will be bounded as $y \to \pm\infty$ as long as the real part of Ω^2 is negative. Similarly, the expression with the minus signs gives bounded solutions as $y \to \pm\infty$ as long as the real part of Ω^2 is positive. Substituting the above expression in equation (6.3.10) gives,

$$\frac{d^2J}{d\zeta^2} - 2\zeta\frac{dJ}{d\zeta} + (\lambda - 1)J = 0, \tag{6.3.13}$$

$$\frac{d^2J}{d\zeta^2} + 2\zeta\frac{dJ}{d\zeta} + (\lambda + 1)J = 0. \tag{6.3.14}$$

We try a solution for both of these equations of the form,

$$J(\zeta) = \zeta^s(c_0 + c_1\zeta + c_2\zeta^2 + \ldots),\ c_0 \neq 0,\ s \geq 0. \tag{6.3.15}$$

For equation (6.3.13) we substitute this expression for J and relate the coefficients in powers of ζ. From the first of these equations, $s = 0$ or $s = 1$. The second gives us $s = 0$, $c_1^- = 0$ or both. Then we can obtain the higher terms by using the recursion relation

$$c^-_{\nu+2} = \frac{(2s + 2\nu + 1 - \lambda)}{(s + \nu + 2)(s + \nu + 1)}c^-_\nu. \tag{6.3.16}$$

The polynomial $J(\zeta)$ can be chosen to be either even or odd. In both cases, this corresponds to choosing $c_1^- = c_3^- = \ldots = 0$. The polynomial is then even or odd depending on whether $s = 0$ or $s = 1$. It can be seen by looking at the

asymptotic behaviour of this series that it must terminate for the solution to be bounded as $\zeta \to \pm\infty$. This means that $\lambda = 2s + 2\nu + 1$ and ν must be an even integer. We can express the cases $s = 0$, $s = 1$ as

$$\lambda = 2n + 1,\ n = 0, 1, 2, \ldots \tag{6.3.17}$$

The polynomials which satisfy the recursion relation, (6.3.16), are known as the Hermite polynomials, H_n, [73], and can be defined by

$$H_n(\zeta) = (-1)^n e^{\zeta^2} \frac{d^n}{d\zeta^n}(e^{-\zeta^2}). \tag{6.3.18}$$

Some properties of the Hermite polynomials are given in Appendix B.

We now look for the solutions of the second expression, (6.3.14) by substituting in equation (6.3.15) and relate coefficients in powers of ζ. Again $s = 0$ or $s = 1$ from the first equation. The second gives us $s = 0$, $c_1^+ = 0$ or both and the recursion relation is

$$c^+_{\nu+2} = -\frac{(2s + 2\nu + 1 + \lambda)}{(s + \nu + 2)(s + \nu + 1)} c^+_\nu. \tag{6.3.19}$$

The same argument for the series terminating applies in this case, in which case λ must satisfy

$$\lambda = -(2n + 1),\ n = 0, 1, 2, \ldots \tag{6.3.20}$$

It is shown in Appendix B that the polynomials which satisfy the differential equation with $\lambda = -(2n + 1)$ are $I_n(\zeta)$, which can be defined by

$$I_n(\zeta) = e^{-\zeta^2} \frac{d^n}{d\zeta^n}(e^{\zeta^2}). \tag{6.3.21}$$

Therefore, if $Re(\Omega^2) > 0$ then the solution of (6.3.10) is given by

$$v = v_0 e^{-\zeta^2/2} H_n(\zeta), \quad \lambda = 2n + 1, \quad n = 0, 1, 2, \ldots \tag{6.3.22}$$

and if $Re(\Omega^2) < 0$ then the solution is given by

$$v = v_0 e^{\zeta^2/2} I_n(\zeta), \quad \lambda = -(2n+1), \quad n = 0, 1, 2, \ldots \tag{6.3.23}$$

These last two expressions show that, when $Re(\Omega^2) > 0$, the values of λ which give bounded solutions to the differential equation (6.3.10) are $\lambda = 2n+1$, $n = 0, 1, 2, \ldots$, and when $Re(\Omega^2) < 0$, the values of λ which give bounded solutions are $\lambda = -(2n+1)$, $n = 0, 1, 2, \ldots$.

Solving the dispersion relation

We now find the solutions of the dispersion relation (6.3.11) for ω so that we can demonstrate the stability of the pressure correction method and determine the structure of the types of waves we expect to observe. Substituting Ω^2 in this equation and rearranging, we get

$$\frac{\omega^2}{\beta c}(\omega + i\epsilon) - \frac{c}{\beta}(\omega + i\gamma)(k^2 + \frac{\beta k}{\omega}) = \lambda(\omega + i\gamma)^{1/2}(\omega + i\epsilon)^{1/2}. \tag{6.3.24}$$

We cannot find the roots of this expression explicitly, so we try to determine the significant characteristics of ω by expanding it in terms of order ϵ, γ. We therefore let $\omega = \omega_0 + \omega_1 + \ldots$, where ω_0 contains the first order terms, ω_1 contains terms of order ϵ, γ and so on. To find the first order terms we set $\epsilon = \gamma = 0$ and obtain,

$$\omega_0^4 - \omega_0^2(\lambda\beta c + (ck)^2) - c^2 k\beta\omega_0 = 0, \tag{6.3.25}$$

which gives solutions

$$\omega_0^3 - \omega_0(\lambda\beta c + (ck)^2) - c^2 k\beta = 0, \tag{6.3.26}$$

$$\omega_0 = 0. \tag{6.3.27}$$

We therefore have four roots, three of which are given by equation (6.3.26) and the fourth by equation (6.3.27). Equation (6.3.26) is the same as equation (2.4.16) in Section 2.4.1 and results in the same gravity and planetary wave solutions as we would obtain without any data assimilation or pressure correction. Equation (6.3.27) is an extra root due to the inclusion of the pressure correction term. We also have Kelvin wave solutions as before, which are described by setting $\hat{v} = 0$ in equations (6.3.1-6.3.4). This satisfies the dispersion relation,

$$(\epsilon - i\omega)\omega^2 = (\gamma - i\omega)(ck)^2, \tag{6.3.28}$$

which, when $\epsilon = \gamma = 0$, gives

$$\omega_0(\omega_0^2 - (ck)^2) = 0. \tag{6.3.29}$$

This has three solutions, one of which is $\omega_0 = 0$ which we already have. The other two are given by $\omega_0^2 = (ck)^2$. The root $\omega_0 = -ck$ is not allowed because of the requirement that the solution be bounded as $y \to \pm\infty$.

So, to first order, we have the following roots:

1. Gravity wave solutions obtained by neglecting the $\frac{\beta k}{\omega}$ term in equation (6.3.24):

$$\omega_0^2 = (ck)^2 + c\beta\lambda. \tag{6.3.30}$$

2. Planetary wave solution obtained by neglecting the $\frac{\omega^2}{\beta c}$ term in equation (6.3.24):

$$\omega_0 = \frac{-\beta kc}{k^2 c + \beta\lambda}. \tag{6.3.31}$$

3. Kelvin wave solution:

$$\omega_0 = ck. \tag{6.3.32}$$

4. Additional wave solution:

$$\omega_0 = 0. \tag{6.3.33}$$

We now need to find the higher order terms in the expansion of ω to ensure that the imaginary parts of ω are negative, i.e. the waves decay to zero.

Before we begin, we expand the quantity Ω^2 as a series in terms of order ϵ and γ. We write

$$\Omega^2 = \frac{c}{\beta}\left(1 + \frac{i(\gamma - \epsilon)}{2(\omega + i\epsilon)} + O(\epsilon^2, \gamma^2, \epsilon\gamma)\right). \tag{6.3.34}$$

So to order ϵ, γ,

$$\Omega^2 = \frac{c}{\beta}(1 + \delta), \tag{6.3.35}$$

where

$$\delta = \frac{i(\gamma - \epsilon)}{2(\omega + i\epsilon)}. \tag{6.3.36}$$

Gravity waves

Here, the standard approximation is that the term $\beta k/\omega$ in equation (6.3.11) is small and so can be neglected. We then have to deal with the equation

$$\omega^2 = \Omega^4\beta^2k^2 + \Omega^2\beta^2\lambda. \tag{6.3.37}$$

We write the first order expression (6.3.30) as,

$$\omega_0^2 = (ck)^2 + c\beta\lambda = (ck)^2(1 + r), \tag{6.3.38}$$

where $r = \lambda\beta c/(kc)^2$. We now substitute $\omega = \omega_0 + \omega_1$ in equation (6.3.37), where $\omega_1 = O(\epsilon, \gamma)$ and use relationship (6.3.35) to obtain an equation for ω_1,

$$\omega_1^2 + 2\omega_0\omega_1 + \omega_0^2 = (ck)^2(1 + 2\delta + \delta^2 + r(1 + \delta)). \tag{6.3.39}$$

We now ignore the ω_1^2, δ^2 terms as they are of order ϵ^2, γ^2 and obtain

$$2\omega_0\omega_1 = \delta(ck)^2(2+r). \tag{6.3.40}$$

We now want to show that the imaginary part of ω_1 is negative, in which case the gravity wave solutions will decay in time. To do this we substitute equation (6.3.36) into equation (6.3.40) and write $\delta = \delta_R + i\delta_I$, $\omega = \omega_0 + \omega_{1R} + i\omega_{1I}$ where subscript R indicates real part, subscript I indicates imaginary part, and w_0 is real. This gives us the equation

$$2\omega_0(\omega_{1R} + i\omega_{1I}) = (ck)^2(2+r)\frac{i(\epsilon-\gamma)(\omega_0+\omega_{1R}-i(\omega_{1I}+\epsilon))}{2((\omega_0+\omega_{1R})^2+(\omega_{1I}+\epsilon)^2)}. \tag{6.3.41}$$

Taking the imaginary part of this equation and only retaining terms of order ϵ, γ leads to the relationship

$$\omega_{1I} = -\frac{1}{4}\frac{(2+r)}{(1+r)}(\epsilon-\gamma). \tag{6.3.42}$$

Taking typical magnitudes for the constants in r, we obtain $r \approx 6\lambda$. If $Re(\Omega^2) > 0$, we have $\lambda = 2n+1$, $n = 0, 1, 2, \ldots$, in which case the bounds on ω_{1I} are given by

$$-\frac{2}{7}(\epsilon-\gamma) \leq \omega_{1I} \leq -\frac{1}{4}(\epsilon-\gamma). \tag{6.3.43}$$

If $Re(\Omega^2) < 0$, we have $\lambda = -(2n+1)$, $n = 0, 1, 2, \ldots$, in which case the bounds on ω_{1I} are given by

$$-\frac{1}{4}(\epsilon-\gamma) \leq \omega_{1I} \leq -\frac{1}{5}(\epsilon-\gamma). \tag{6.3.44}$$

We can therefore say that, to order ϵ, γ, provided $\epsilon - \gamma > 0$, the gravity waves will decay towards zero as time increases.

Planetary waves

For planetary waves, the standard approximation is to neglect the term $\omega^2/(\Omega^2\beta^2)$ in equation (6.3.11). This gives us the equation

$$\omega = -\frac{\Omega^2\beta^2}{k(rc + \Omega^2\beta)}. \tag{6.3.45}$$

We write the first order terms as

$$\omega_0 = -\frac{\beta}{k(1+r)}. \tag{6.3.46}$$

We now substitute $\omega = \omega_0 + \omega_1$ in equation (6.3.45), where again $\omega_1 = O(\epsilon, \gamma)$, and use relationship (6.3.35) to obtain

$$\omega_1 = \frac{r}{1+r}\delta\omega_0. \tag{6.3.47}$$

Using equation (6.3.36) and splitting ω and δ into real and imaginary parts as before, we get

$$\omega_{1R} + \omega_{1I} = \omega_0 \frac{r}{1+r} \frac{i(\epsilon-\gamma)(\omega_0 + \omega_{1R} - i(\omega_{1I}+\epsilon))}{2((\omega_0+\omega_{1R})^2 + (\omega_{1I}+\epsilon)^2)}. \tag{6.3.48}$$

Taking the imaginary parts of this equation and retaining terms of order ϵ, γ, gives

$$\omega_{1I} = -\frac{r}{1+r}\frac{(\epsilon-\gamma)}{2}. \tag{6.3.49}$$

Taking typical values for the constants in r, we can now find bounds on ω_{1I}. If $Re(\Omega^2) > 0$, then the bounds are given by

$$-\frac{1}{2}(\epsilon-\gamma) \leq \omega_{1I} \leq -\frac{3}{7}(\epsilon-\gamma). \tag{6.3.50}$$

If $Re(\Omega^2) < 0$, then the bounds are given by

$$-\frac{3}{5}(\epsilon-\gamma) \leq \omega_{1I} \leq -\frac{1}{2}(\epsilon-\gamma). \tag{6.3.51}$$

This shows that, provided $\epsilon - \gamma > 0$, the planetary waves will also decay towards zero as time increases.

Kelvin waves

We now look for the imaginary parts of the Kelvin wave solution by setting $\omega = \omega_0 + \omega_1$ where $\omega_1 = O(\epsilon, \gamma)$. For the root $\omega_0 = ck$, we find $\omega_1 = -\frac{i}{2}(\epsilon - \gamma)$ so the solution for this root is

$$\omega = ck - \frac{i}{2}(\epsilon - \gamma) + O(\epsilon^2, \gamma^2, \epsilon\gamma). \tag{6.3.52}$$

The fact that this root has negative imaginary parts shows that the Kelvin waves will also decay towards zero as time increases.

Additional wave

We return to the original dispersion relation, (6.3.11), square and rearrange to obtain

$$\begin{aligned} \omega^6(\omega + i\epsilon)^2 &- 2c^2\omega^2(k^2\omega + \beta k)(\omega + i\gamma)(\omega + i\epsilon) + c^4(\omega + i\gamma)^2(k^2\omega + \beta k)^2 \\ &= \lambda^2\beta^2c^2\omega^2(\omega + i\epsilon)(\omega + i\gamma). \end{aligned} \tag{6.3.53}$$

We now substitute $\omega = \omega_0 + \omega_1$ where $\omega_1 = O(\gamma)$ and $\omega_0 = 0$. We assume γ to be the only small quantity here because it turns out that the first term in the expansion is of order γ. We obtain

$$\omega_1 = -i\gamma. \tag{6.3.54}$$

So the leading term in the imaginary part of ω is negative which shows that this additional wave solution will decay in time.

The pressure correction is inducing this additional wave and so we would like to find its structure. The meridional structure of the solution depends on $Re(\Omega^2)$ as can be seen from equations (6.3.22) or (6.3.23) and the relationship

$y = \Omega\zeta$. We therefore look for the next terms in the expansion of ω for which $Re(\Omega^2)$ is non-zero.

Setting $\omega \approx \omega_0 + \omega_1 + \omega_2 = \omega_2 - i\gamma$, where $\omega_2 = O(\epsilon\gamma^2)$, we get

$$\omega_2 = -i\frac{\lambda^2\epsilon\gamma^2}{(ck)^2}. \tag{6.3.55}$$

If we now put this into the expression for Ω^4 we obtain

$$\Omega^4 \approx \frac{c^2}{\beta^2}(\frac{\omega + i\gamma}{i\epsilon}) = -\frac{\lambda^2\gamma^2}{(ck)^2}, \tag{6.3.56}$$

where the term $i\epsilon$ only appears on the bottom because γ is assumed to be smaller than ϵ here. This result shows that, to this order, Ω^2 is a purely imaginary number and so the solution for v will not decay away from the equator. We therefore look for the solution to the next order by setting $\omega = \omega_0 + \omega_1 + \omega_2 + \omega_3$ where $\omega_3 = O(\gamma^3)$. To order γ^5 we obtain

$$\begin{aligned} 2c^2\beta k\epsilon\gamma^3\omega_2 &+ c^4(\beta^2k^2\omega_2^2 + 2\omega_2\omega_3\beta^2k^2 - 2i\beta k^3\gamma\omega_2^2) \\ &= \lambda^2\beta^2c^2(2\epsilon\gamma\omega_2^2 + i\gamma^2(\gamma - \epsilon)\omega_2 - i\epsilon\gamma^2\omega_3), \end{aligned} \tag{6.3.57}$$

which when rearranged gives

$$\omega_3 = \frac{2k\epsilon\gamma^3}{\beta(ck)^2}(\lambda^2 - 1) - i\frac{\lambda^2\gamma^3}{(ck)^2}(\frac{\lambda^2\epsilon^2}{(ck)^2} + 1). \tag{6.3.58}$$

Apart from the case when $n = 0$, the additional wave solution will therefore have a non-zero real part to order $\epsilon\gamma^3$. The meridional structure of the solution depends on Ω^2. Substituting ω to order $\epsilon\gamma^3$ into the expression for Ω^4, we obtain

$$\Omega^4 \approx \frac{c^2}{\beta^2}\left(\frac{\omega + i\gamma}{i\epsilon}\right) = \frac{c^2}{\beta^2\epsilon}(\gamma - i\omega) \tag{6.3.59}$$

$$= -\gamma^2\left(\frac{\lambda^2}{\beta^2k^2}(1 - \frac{\lambda^2\epsilon}{(ck)^2}) - \frac{\gamma}{\beta^2k}(\frac{\lambda^2}{k\epsilon} - 2i\frac{(\lambda^2 - 1)}{\beta})\right) \tag{6.3.60}$$

$$\equiv -p_1^2\gamma^2(1 - ip_2\gamma), \tag{6.3.61}$$

where $p_1^2 = \frac{\lambda^2}{\beta^2 k^2}(1 - \frac{\lambda^2 \epsilon}{(ck)^2})$ and $p_2 = -\frac{(ck)^2}{\lambda^2 \epsilon}(\frac{k\epsilon(\lambda^2-1)+i\lambda^2\beta}{(ck)^2-\lambda^2\epsilon})$. So, to this order, $\Omega^2 = -ip_1\gamma(1 + ip_2\frac{\gamma}{2})$, and splitting p_2 into real and imaginary parts as $p_2 = p_{2R} + ip_{2I}$ gives us

$$Re(\Omega^2) \approx p_1 p_{2R} \frac{\gamma^2}{2}. \tag{6.3.62}$$

Summary

In this section, we first of all showed that the horizontal structure of the differences between the true and model solutions of the forced shallow water equations, when the pressure correction method is included, can be written in terms of one equation for $\hat{v}$. We then assumed that $\hat{v}$ could be written as $\hat{v} = v(y)exp(ikx - i\omega t)$ and found an equation for $v(y)$ in the homogeneous case. We then rescaled the y variable to put the equation for v in the same form as it would be without any assimilation or pressure correction. We then found the values of λ for which the solution is bounded as $y \to \pm\infty$ and found the structures of these solutions, which depend on $Re(\Omega^2)$.

To show that the solution of the equation for $\hat{v}$ will tend to zero as time increases, that is the model solution will tend toward the true solution, we next showed that the imaginary part of ω is negative by solving the dispersion relation (6.3.11). This was done by expanding the various wave types in terms of order ϵ, γ to find the largest imaginary part of ω. In the four types of waves, gravity, planetary, Kelvin and the additional wave, we found that the leading imaginary term in the expansion of ω was in fact negative, provided $\epsilon - \gamma > 0$. This therefore showed that, if we have observations of the true density field and perform data assimilation of the type given in equation (6.2.10) with the pressure correction method included, the model solution

will tend toward the true solution as time increases. It also shows that the weighting for the pressure correction, γ, must be smaller than the weighting for the data assimilation, ϵ, for the method to be stable. It would be useful, for a more realistic analysis, to determine the effects of time varying forcing on the system. This is not discussed here however.

The leading real terms obtained in the expansions of ω for the gravity, planetary and Kelvin waves show the same structure for these waves as we would obtain without any data assimilation or pressure correction, as in Section 2.4.1. The structure of the additional wave solution was found for the first term in the expansion of ω for which $Re(\Omega^2)$ is non-zero. These waves will have much smaller magnitudes than the other types of waves away from the equator, due to the fact that $Re(\Omega^2)$ is of order γ^2 in this case, and so should not significantly affect the overall structure.

6.4 Experimental set-up

6.4.1 Numerical model

In this section, we would like to verify some of the theoretical results of the previous two sections in an experimental setting. One of the results to be examined is that normal data assimilation will not account for incorrect wind forcing in the forecast model. We would also like to examine the improvements, or otherwise, made by the normal bias correction method and the pressure correction method in the presence of incorrect wind forcing. To do this we use a simple numerical model which still produces some of the important features we wish to study. The model equations we use are the

linear two-dimensional shallow water equations, given by

$$\frac{\partial u}{\partial t} - fv = -g\frac{\partial p}{\partial x} + \tau^x, \tag{6.4.1}$$

$$\frac{\partial v}{\partial t} + fu = -g\frac{\partial p}{\partial y} + \tau^y, \tag{6.4.2}$$

$$\frac{\partial p}{\partial t} + H_e(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}) = 0, \tag{6.4.3}$$

where H_e is a separation constant and g is the gravitational constant. We study the flow near the equator and so make the β-plane approximation, $f = \beta y$. The true solution will be given by the above equations forced by the true wind forcing. The model equations will be given by the same equations but forced by some incorrect wind forcing.

Spatial discretisation

A staggered grid has been chosen for the spatial discretisation. This is the B-grid which holds values of u and v at points i, j and values of p at points $i + \frac{1}{2}$, $j + \frac{1}{2}$, [9], as shown in Figure 6.4.1.

We would like the domain of the model to be similar to the size of the equatorial Pacific. We therefore use the following size grid:

- $\Delta x = 2.22 \times 10^5 m = 2^0$, $i = 0, 1, \ldots, 50$ so that $x^i = i\Delta x = 0, 2^0, \ldots, 100^0$.
- $\Delta y = 2.22 \times 10^5 m = 2^0$, $j = -10, -9, \ldots, 10$ so that $y^j = j\Delta y = -20^0, -18^0, \ldots, 20^0$.

We approximate gradients of p at u and v points by

$$\frac{\partial p}{\partial x}^{i,j} = \frac{1}{2\Delta x}\{p^{i+1/2,j+1/2} + p^{i+1/2,j-1/2} - p^{i-1/2,j+1/2} - p^{i-1/2,j-1/2}\},$$

$$\frac{\partial p}{\partial y}^{i,j} = \frac{1}{2\Delta y}\{p^{i+1/2,j+1/2} + p^{i-1/2,j+1/2} - p^{i+1/2,j-1/2} - p^{i-1/2,j-1/2}\}.$$

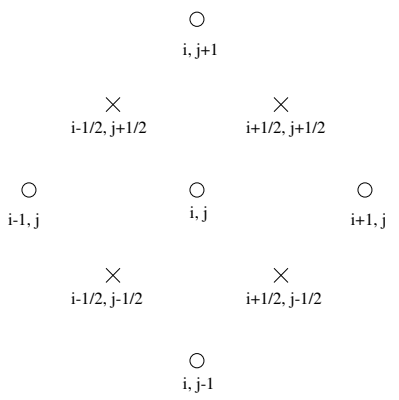

Figure 6.4.1: The B grid: u and v are held at the points indicated by the circles and p is held at points indicated by the crosses.

Gradients of u and v are approximated at p points by

$$\begin{aligned} \frac{\partial u}{\partial x}^{i+1/2,j+1/2} &= \frac{1}{2\Delta x}\{u^{i+1,j+1} + u^{i+1,j} - u^{i,j+1} - u^{i,j}\}, \\ \frac{\partial v}{\partial y}^{i+1/2,j+1/2} &= \frac{1}{2\Delta y}\{v^{i+1,j+1} + v^{i,j+1} - v^{i,j+1} - v^{i,j}\}. \end{aligned}$$

Artificial dissipation of momentum is included in the momentum equations by adding a term of the form

$$\kappa \nabla^2 \underline{u}_k, \tag{6.4.4}$$

where κ is kept as small as possible whilst keeping the model numerically stable. The scheme is second order accurate in space.

Time discretisation

The leap-frog scheme is used as the time discretisation, [67]. Here, the time derivatives are approximated as

$$\frac{\partial \mu_k}{\partial t_k} = \frac{\mu_{k+1} - \mu_{k-1}}{2\Delta t}. \tag{6.4.5}$$

The scheme is second order accurate in time. To avoid producing different solutions at even and odd time steps, a time filter is used, whereby the new value of each variable is given as

$$\bar{\mu}_k = \mu_k + \alpha(\mu_{k+1} - 2\mu_k + \mu_{k-1}). \tag{6.4.6}$$

The first time step is calculated using forward time differencing. We choose a time step of $\Delta t = 8.64 \times 10^3 s$ so that the model is numerically stable. The time steps are given by $k = 0, 1, \ldots, 500$, giving us a time period of 50 days. The speed of disturbances will be of the order of $c^2 = gH = (2.8ms^{-1})^2$.

Initial and boundary conditions

At the initial time we will choose all the variables to be zero, i.e.

$$u(x, y, 0) = 0, \quad v(x, y, 0) = 0, \quad p(x, y, 0) = 0. \tag{6.4.7}$$

The wind stress should then start up a circulation. The form of the wind stress will be

$$\tau^x = \begin{cases} Aexp(-\frac{1}{4}y^2)cos(\pi x/2L) & \text{for } |x| < L \\ 0 & \text{for } |x| > L \end{cases}, \tag{6.4.8}$$

where A is a chosen amplitude and L is a length scale for the extent of the forcing in the x direction. This is shown in Figure 6.4.2 for $L = 1 \times 10^6 m$, $A = 2 \times 10^{-5} m^2 sec^{-1}$, which is the amplitude used by [81]. We set $\tau^y = 0$ everywhere for simplicity.

The boundary conditions are that u and v have no slip boundaries at the north and south of the domain and p is periodic in the x direction, i.e.

$$u(x, -20, t) \quad = \quad u(x, 20, t) = 0 \tag{6.4.9}$$

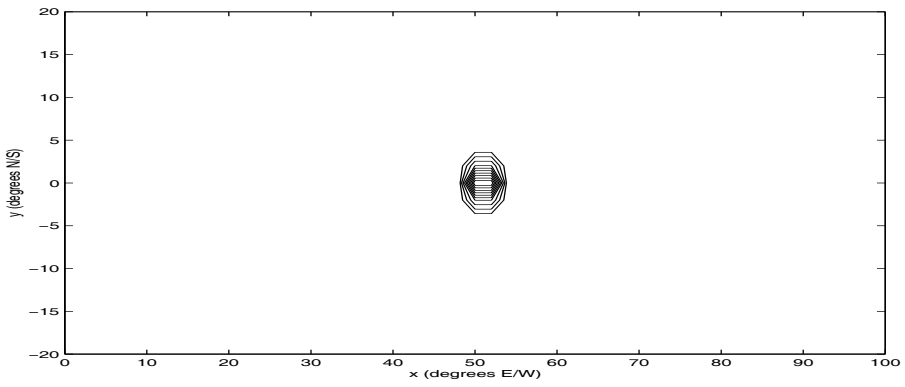

Figure 6.4.2: Wind stress τ^x, $A = 2 \times 10^{-5} m^2 s^{-1}$, $L = 1 \times 10^6 m$.

$$v(x, -20, t) = v(x, 20, t) = 0 \tag{6.4.10}$$

$$p(0, y, t) = p(100, y, t) \tag{6.4.11}$$

6.4.2 Data assimilation scheme

To retain some similarities with the methods used in oceanography, only observations of p will be available. These will be at similar locations to that of the TAO array in the equatorial Pacific and will be available for assimilation every time step. The exact location of the observations is given in Figure 6.4.3.

The method which we will use to assimilate these observations is a univariate Optimal Interpolation. Here we will only make increments to the p variable based on the differences between the observed and forecast values of that variable. The velocities u and v will not be altered in the assimilation step. The OI formula is

$$\underline{p}_k^a = \underline{p}_k^f + K^p[\underline{p}_k^{ob} - H^p \underline{p}_k^f], \tag{6.4.12}$$

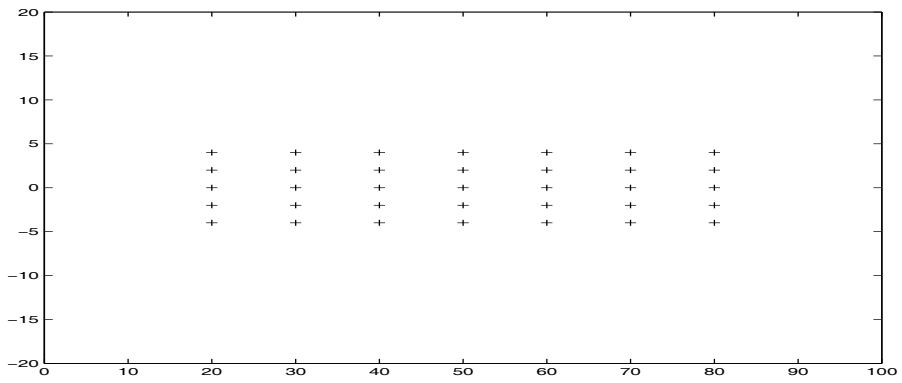

Figure 6.4.3: Position of observations of p at each time step

where $\underline{p}_k$ contains the values of p at all the grid points at time step k, K^p is the usual OI weighting matrix given by $K^p = B^p H^{p\,T}[H^p B^p H^{p\,T} + R]^{-1}$, H^p is the linear observation operator that interpolates from model grid to the observation grid, B^p is the forecast error covariance matrix and R is the observation error covariance matrix. The method used to solve equation (6.4.12) at each time step is:

1. Let $\underline{p}_k^{ob\_inc} = \underline{p}_k^{ob} - H^p \underline{p}_k^f$ be the observation increments.

2. Calculate the LU decomposition of $H^p B^p H^{p\,T} + R$.

3. Solve $\underline{w}_k = [H^p B^p H^{p\,T} + R]^{-1} \underline{p}_k^{ob\_inc}$ using back and forward substitution.

4. Calculate the analysis increments, $\underline{p}_k^{an\_inc} = B^p H^{p\,T} \underline{w}_k$.

5. Then the analysis is $\underline{p}_k^a = \underline{p}_k^f + \underline{p}_k^{an\_inc}$.

Background error covariance matrix

The specification of the forecast error covariance matrix, B^p, is a very important part of the method. B^p is rarely known accurately in practice and is approximated using various techniques. The one used in these experiments is to specify the variances of the forecast error, σ_i, which appear on the diagonal of the matrix. The off diagonal terms are then calculated using a correlation function which shows how much correlation one point in the domain has with another. The correlation function χ used here, as suggested by [21], relates point (x^i, y^i) to a point (x^j, y^j) using the formula

$$\chi((x^i, y^i), (x^j, y^j)) = exp(\frac{-(x^i - x^j)^2}{2\lambda_x^2})exp(\frac{-(y^i - y^j)^2}{2\lambda_y^2}) \qquad (6.4.13)$$

where λ_x and λ_y are given length scales. To determine what values these length scales should take, a simple experiment was performed with a given set of observations of value 1. The effects the weighting matrix K^p has on these observations for different values of length scales was studied. The choice for the rest of the experiments is $\lambda_x = 6 \times 10^5 m$, $\lambda_y = 1.5 \times 10^5 m$ as this spreads the observations out so that most of the region is covered.

Observation error covariance matrix

For simplicity, we set the observation error covariance matrix, R, to be a diagonal matrix. The diagonal entries are equal to the variance of the noise on the observations.

Bias correction methods

We would like to test the effectiveness of the systematic model error correction technique described in the previous chapter and the special case of this, the pressure correction method described in this chapter, in accounting for systematic errors in the wind forcing of the linear shallow water equations. Here we describe how we implement these methods.

For the normal bias correction method, we introduce two new variables, $\underline{b}^u$ and $\underline{b}^v$, which are added to the u and v momentum equations respectively and have the same dimensions as those variables. These are kept constant in the forecast step, but are altered in the data assimilation by the formulae

$$\underline{b}_k^{u\,a} = \underline{b}_k^{u\,f} + K^b[\underline{p}_k^{ob} - H^p \underline{p}_k^f], \tag{6.4.14}$$

$$\underline{b}_k^{v\,a} = \underline{b}_k^{v\,f} + K^b[\underline{p}_k^{ob} - H^p \underline{p}_k^f], \tag{6.4.15}$$

where the bias weighting matrix is given by $K^b = B^{pb\,T} H^{p\,T}[H^p B^p H^{p\,T} + R]^{-1}$ and B^{pb} is the cross-correlation between the errors in the variables $\underline{p}_k^f$ with the errors in the bias variables. It is difficult to obtain an accurate estimate of this covariance matrix. Here, we use a very simple estimate, whereby $B^{pb} = \alpha B^p$, and we leave the choice of α for the next section. This form of the covariance matrix is not necessarily a good choice because the bias variables are biases on the velocities rather than the pressure variable. We choose this form so that the intercomparison with the pressure correction method is made using the same assumptions.

For the pressure correction method, the new variable $\underline{p}_k^c$ is added onto the variable $\underline{p}_k$ when calculating the pressure gradients in the horizontal momentum equations. The forecast of $\underline{p}_k$ is left unaltered. Again, the update

of the pressure correction term is left unchanged in the forecast model, so that

$$\underline{p}^c_{k+1} = \underline{p}^c_k. \tag{6.4.16}$$

These variables are calculated in the assimilation by

$$\underline{p}^{c\,a}_k = \underline{p}^{c\,f}_k + K^c[\underline{p}^{ob}_k - H\underline{p}^f_k], \tag{6.4.17}$$

where $K^c = B^{pc\,T}H^{p\,T}[H^pB^pH^{p\,T} + R]^{-1}$ and B^{pc} is the cross-correlation between the errors in the variable $\underline{p}^f_k$ with the errors in the pressure correction variable. To keep this method comparable to the bias correction method, we approximate the B^{pc} matrix by $B^{pc} = \gamma B^p$. This choice is more reasonable than that for the bias correction covariance matrix as we are approximating B^{pc} using a covariance matrix which is calculated for the pressure variable.

6.5 Results of experiments

To test the numerical code used in this section, an experiment was performed with forcing on equation (6.4.3) rather than on the momentum equations. This experiment corresponds to some theoretical results given by [34]. The results of this experiment were qualitatively similar to the theoretical results and verified that the programming of the numerical code was correct. The results of this are not shown here.

For all the following experiments, we set the gravitational constant, g, to be $9.8ms^{-2}$ and the separation constant, H_e, to be $0.8m$ to give a value for $c = \sqrt{gH_e}$ of $2.8ms^{-1}$. As we are dealing with a linear model, it is convenient to set the forcing for the true solution to be zero. The initial conditions are

zero everywhere and so the true solution is also zero for all time. We obtain observations from this solution at the points indicated previously at every time step and for simplicity do not introduce any noise. We therefore set the observation error covariance matrix, R, to be zero.

We introduce systematic errors in the wind forcing by altering its amplitude and length scale to those given in the previous section, that is

$$\tau^x = \begin{cases} 2 \times 10^{-5} exp(-\frac{1}{4}y^2) cos(\frac{\pi x}{2 \times 10^6}) & \text{for } |x| < L \\ 0 & \text{for } |x| > L \end{cases} . \tag{6.5.1}$$

The errors in the wind stresses will not have zero curl, which is an assumption of the pressure correction method, so this form of the wind stress should provide a tough test for the method.

A run has been performed for 50 days using this model which gives us the background solution, i.e. the one we obtain without any data assimilation. The results of this integration are displayed in Figures 6.5.1, 6.5.2, 6.5.3 and 6.5.4. These show a wave travelling eastwards along the equator from the region of the wind forcing with a circulation set up in the area of the wind forcing. All the disturbances are contained within about $\pm 10^0$ of the equator, as we would expect.

6.5.1 Optimal Interpolation in the presence of incorrect wind forcing

We now perform OI on this system at every time step, as described in the previous section. Here we set the value for the variance of the background error covariance matrix for p to be $\sigma = 1$. The results after 50 days are

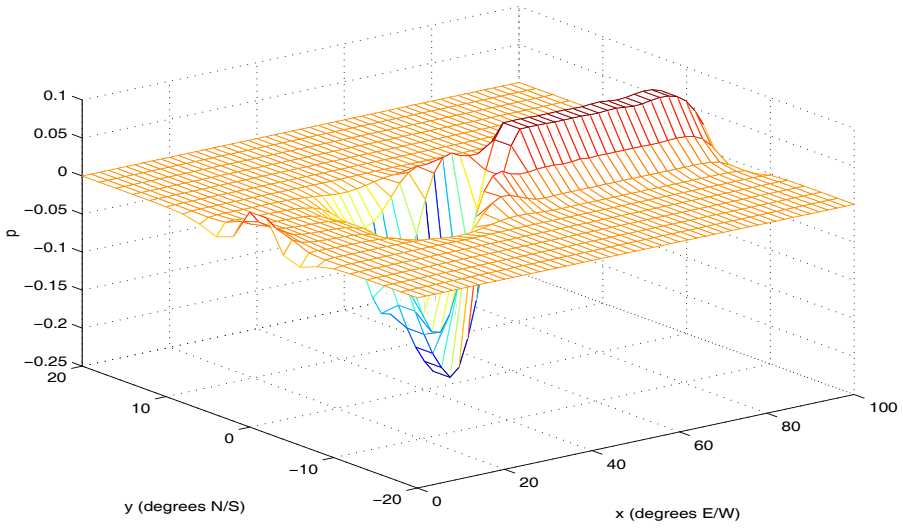

Figure 6.5.1: Errors in p without any data assimilation after 50 days

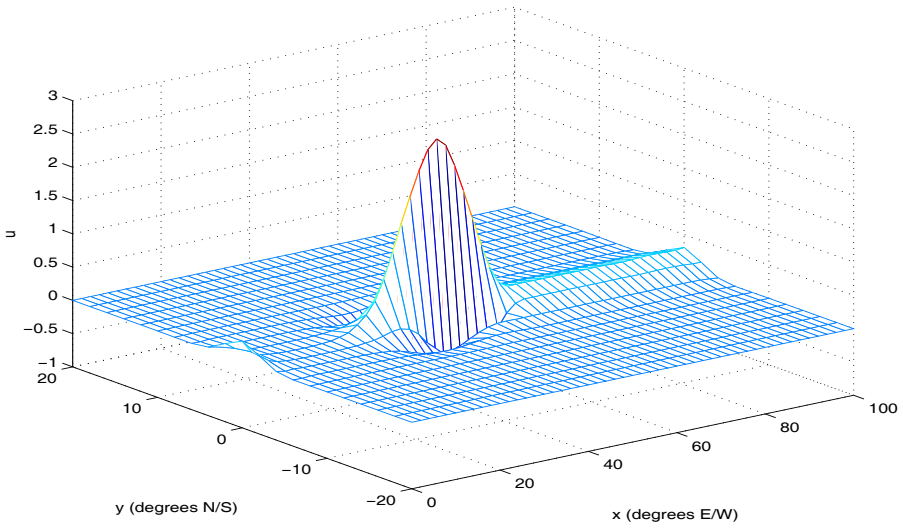

Figure 6.5.2: Errors in u velocity without any data assimilation after 50 days

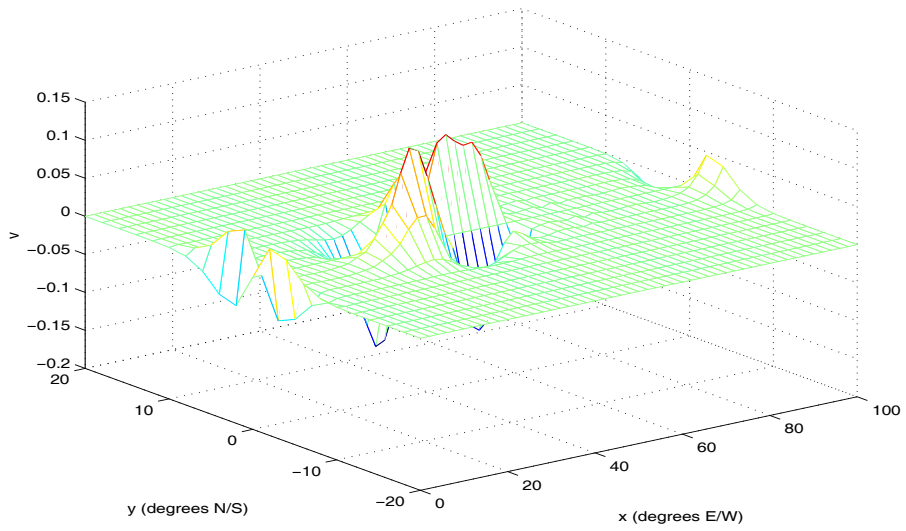

Figure 6.5.3: Errors in v velocity without any data assimilation after 50 days

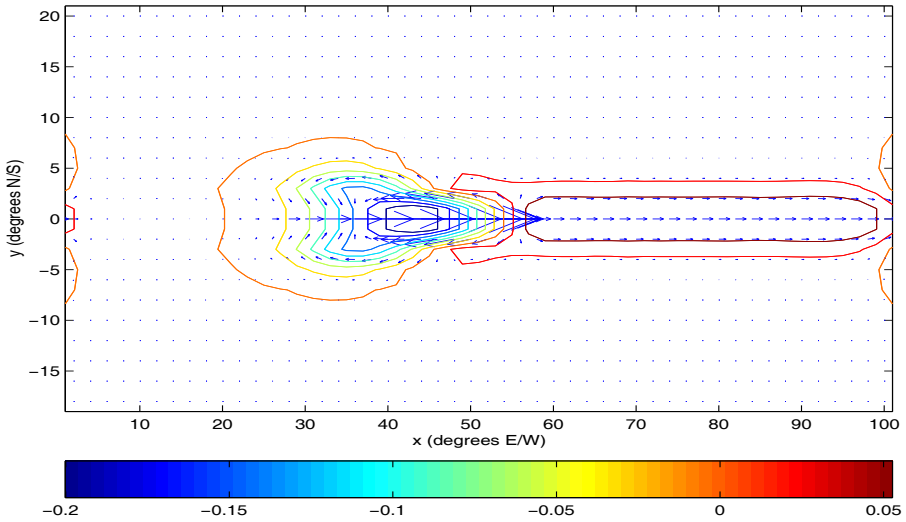

Figure 6.5.4: Errors in p (contours) and currents (arrows) without any data assimilation after 50 days

shown in Figures 6.5.5, 6.5.6, 6.5.7 and 6.5.8. It is clear from these figures that the structure of the solution is completely altered. If there were no systematic errors present, we would expect OI to reduce the errors in the system significantly. The pressure has only slightly smaller errors however, with the largest errors occurring to the west of the domain and in the region of wind forcing. The solution is also rather noisy. Errors in the u velocity are larger after performing OI than without data assimilation, especially in the region of the wind forcing, with wavy features elsewhere. The v velocity has slightly smaller errors, although there is a large peak in the west. The problem we would like to correct for is the larger errors in the u velocity as the u momentum equation is the one containing the errors in the wind stress. We would also like to keep the reduction of the errors in the p variable.

6.5.2 Accounting for the incorrect wind forcing

Normal Bias Correction

We now perform the same experiments but with the bias correction method included. If we select the value of α, the weighting of the bias background covariance matrix, above $\alpha = 1 \times 10^{-3}$, there is divergence. For values of $\alpha = 1 \times 10^{-4}$ and $\alpha = 1 \times 10^{-3}$, the L_2 norm of p, u and v are shown in Figures 6.5.9, 6.5.10 and 6.5.11 respectively, together with the values when there is no bias correction, i.e. $\alpha = 0$. These figures show that there is little improvement in any of the errors, although the results obtained with $\alpha = 1 \times 10^{-4}$ produce slightly smaller errors in the pressure and v velocity at the end of the period. However, the bias correction method does not seem

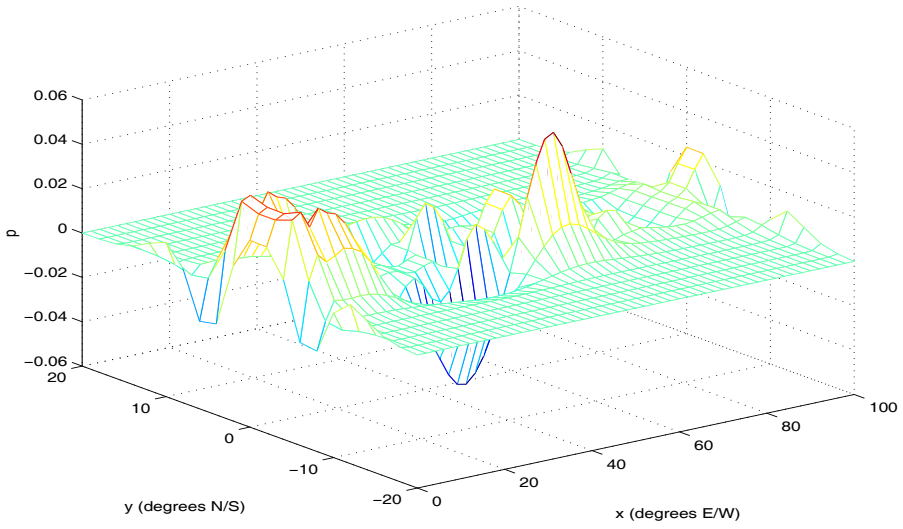

Figure 6.5.5: Errors in p after performing OI for 50 days

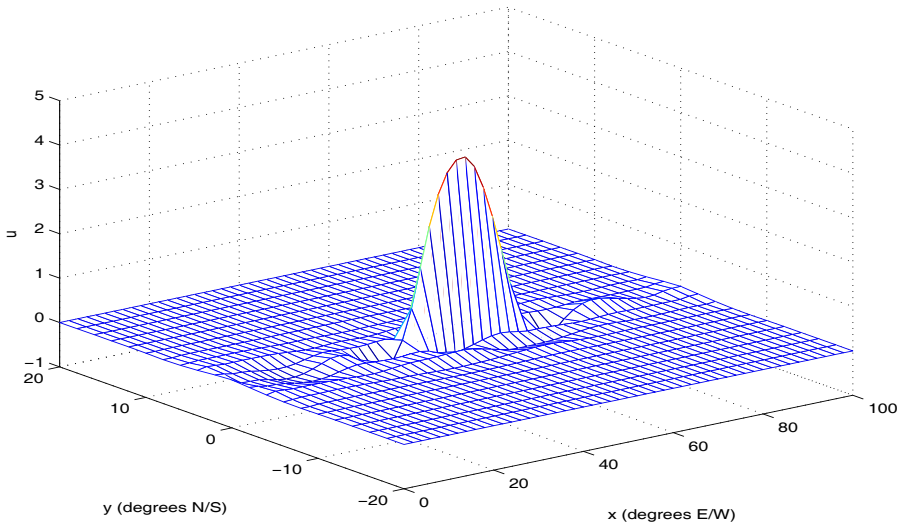

Figure 6.5.6: Errors in u velocity after performing OI for 50 days

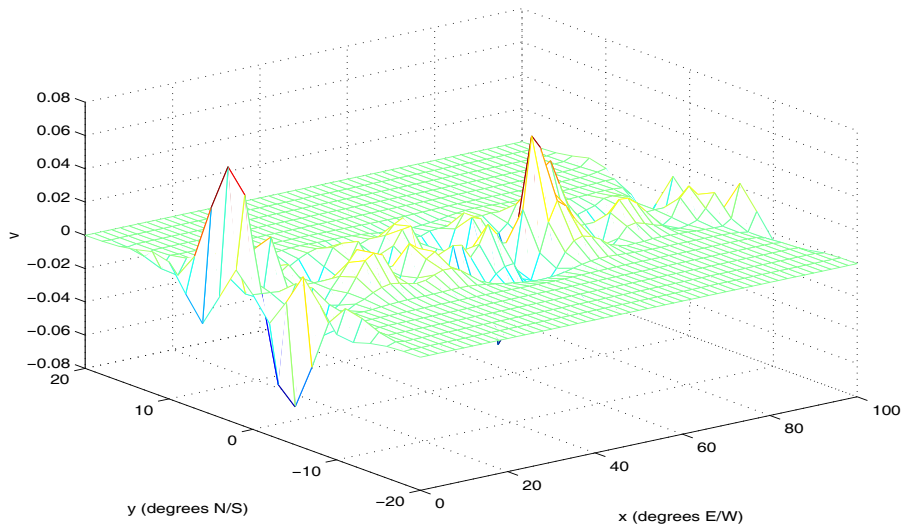

Figure 6.5.7: Errors in v velocity after performing OI for 50 days

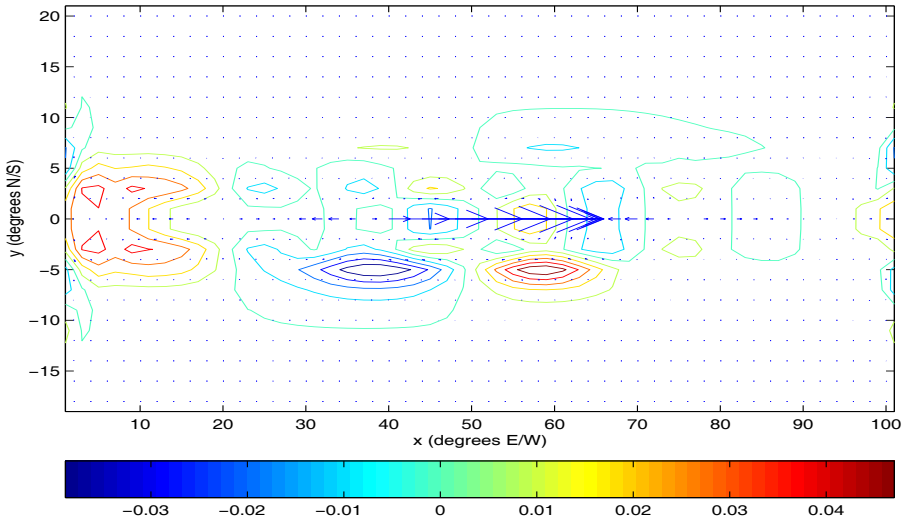

Figure 6.5.8: Errors in p (contours) and currents (arrows) after performing OI for 50 days

able to reduce errors in the u velocity, perhaps due to the simple choice of gain matrix used for the bias variables. If we were to choose the gain matrix based on statistics of the velocity field and its correlation with the pressure variable, this form of bias correction should provide better results. These statistics are difficult to obtain in practice.

From these results we choose a value of $\alpha = 1 \times 10^{-4}$ for comparison with other experiments as this appears to be the value for which the results are the best. We now plot the pressure, the u velocity and the v velocity for a run with this value of α in Figures 6.5.12, 6.5.13, 6.5.14 and 6.5.15. These show the structure of the solution to be similar to the one obtained using the normal data assimilation scheme.

Pressure Correction Method

To determine the best value of γ for the pressure correction method, some runs have been performed with various different magnitudes for this variable. The L_2 norm of the pressure, u velocity and v velocity are shown in Figures 6.5.16, 6.5.17 and 6.5.18 respectively. These results show that the errors in the pressure and u velocity have been reduced when compared to the run with normal data assimilation. Errors in the v velocity appear to be of similar magnitude to the run with normal data assimilation. This is surprising as the dynamics of the shallow water equations should mean that any reduction in u leads to a reduction in v. We choose a value of $\gamma = -0.1$ for the remainder of the experiments.

For the run with $\gamma = -0.1$ the pressure, u velocity and v velocity fields are shown after 50 days of integration in Figures 6.5.19, 6.5.20, 6.5.21 and

Key to Figures 6.5.9 - 6.5.11: normal OI (solid line), bias correction; $\alpha = 1 \times 10^{-4}$ (dashed line), bias correction; $\alpha = 1 \times 10^{-3}$ (dotted line).

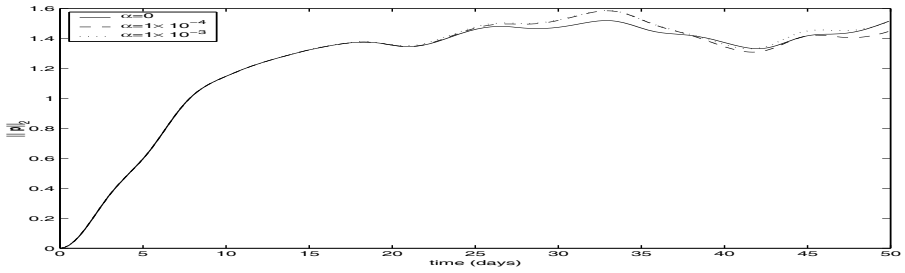

Figure 6.5.9: L_2 norm of the errors in the pressure field for runs with different values of α.

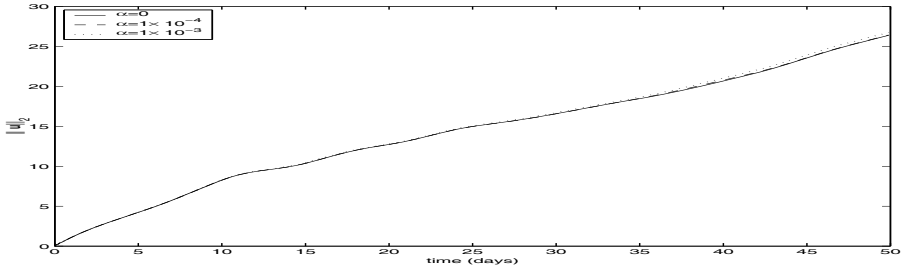

Figure 6.5.10: L_2 norm of the errors in the u velocity field for runs with different values of α.

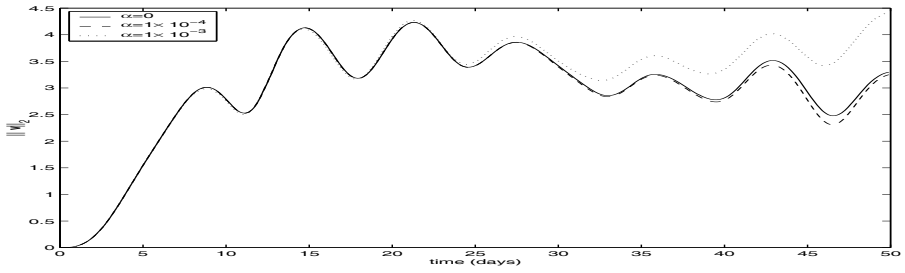

Figure 6.5.11: L_2 norm of the errors in the v velocity field for runs with different values of α.

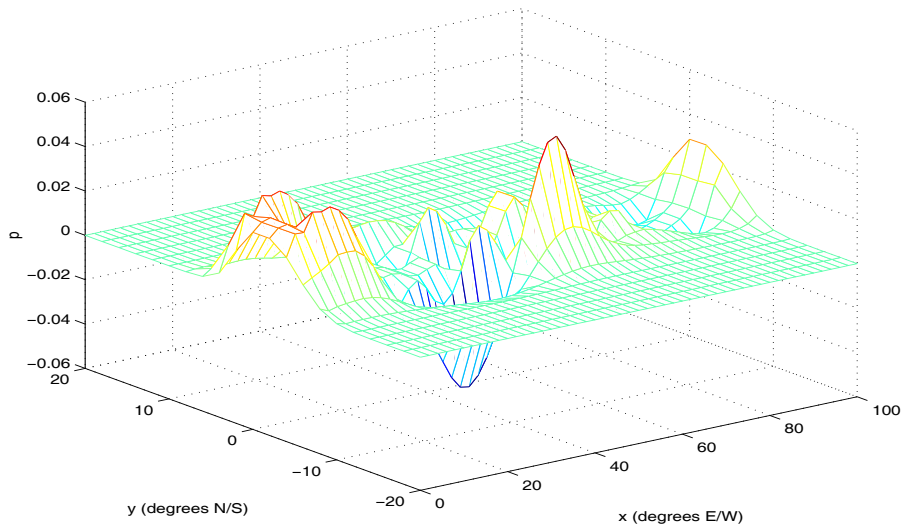

Figure 6.5.12: Error in p after performing OI with bias correction for 50 days

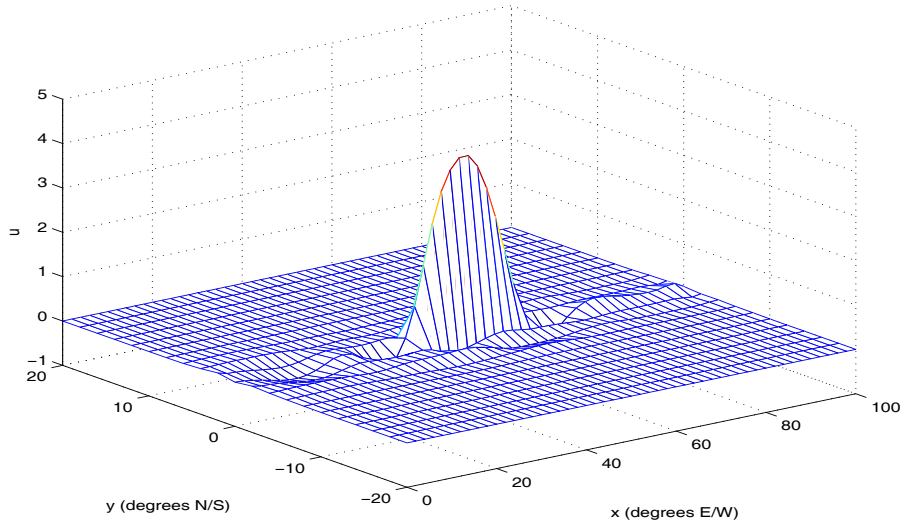

Figure 6.5.13: Error in u velocity after performing OI with bias correction for 50 days

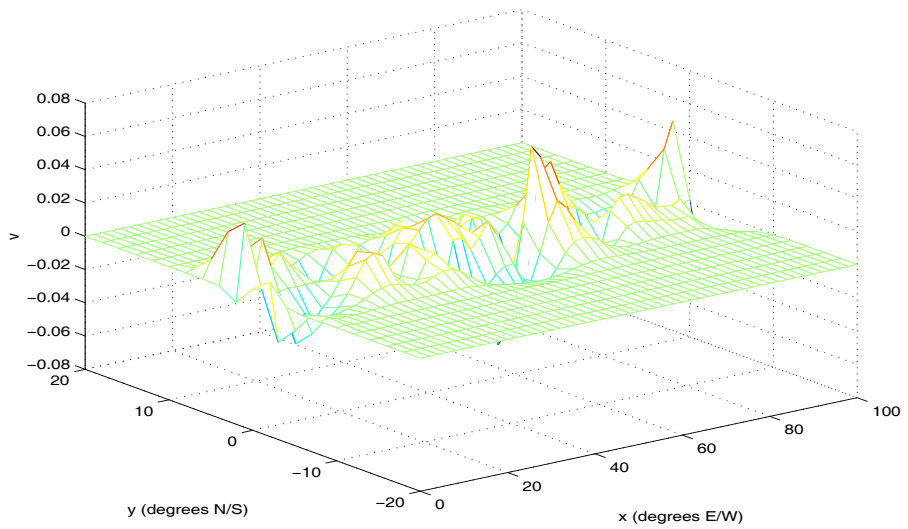

Figure 6.5.14: Error in v velocity after performing OI with bias correction for 50 days

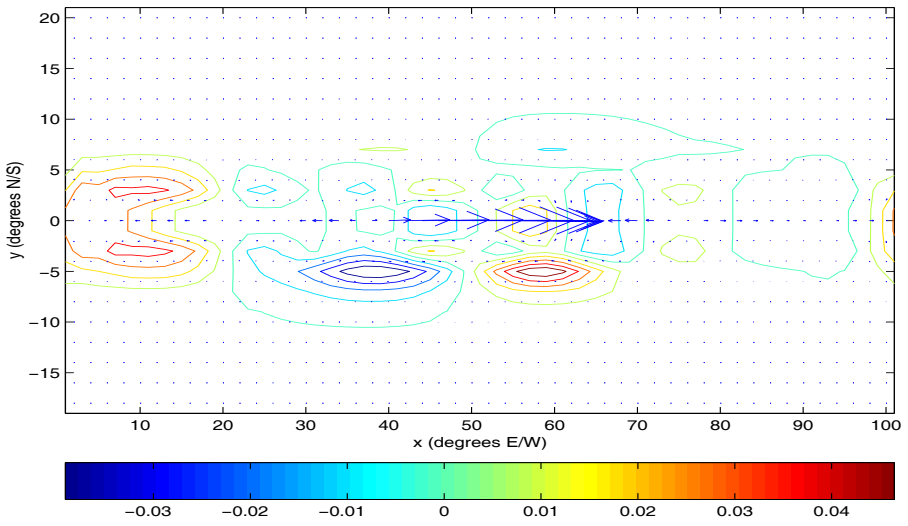

Figure 6.5.15: Error in p (contours) and currents (arrows) after performing OI with bias correction for 50 days

6.5.22. These figures reinforce the results described above. The errors in the pressure are slightly smaller, there is a significant reduction in the errors of the u velocity, and the v velocity has errors of similar magnitude but they are concentrated in the region of the wind forcing now.

Key to Figures 6.5.16 - 6.5.18: normal OI (solid line), pressure correction; $\gamma = -0.1$ (dashed line), pressure correction; $\gamma = -0.3$ (dash-dot line), pressure correction; $\gamma = -0.5$ (dotted line).

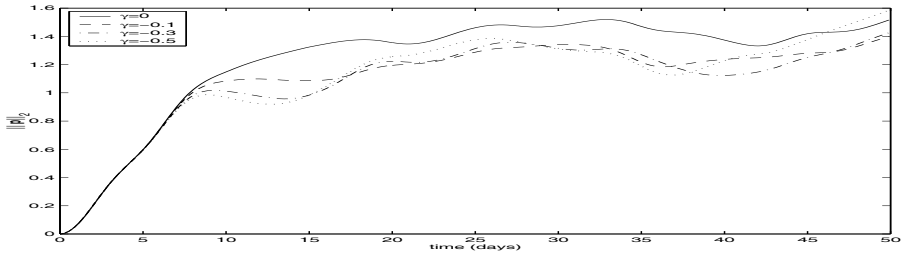

Figure 6.5.16: L_2 norm of the errors in the pressure field for runs with different values of γ.

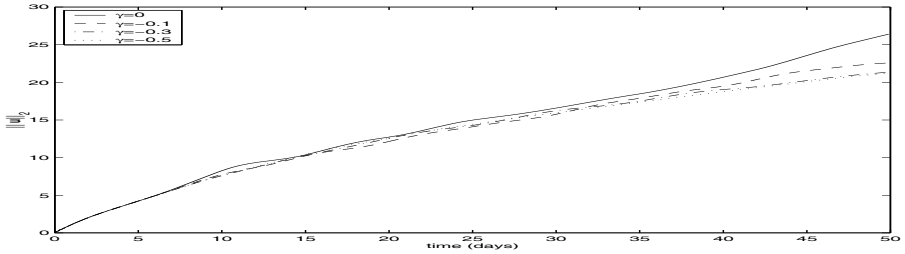

Figure 6.5.17: L_2 norm of the errors in the u velocity field for runs with different values of γ.

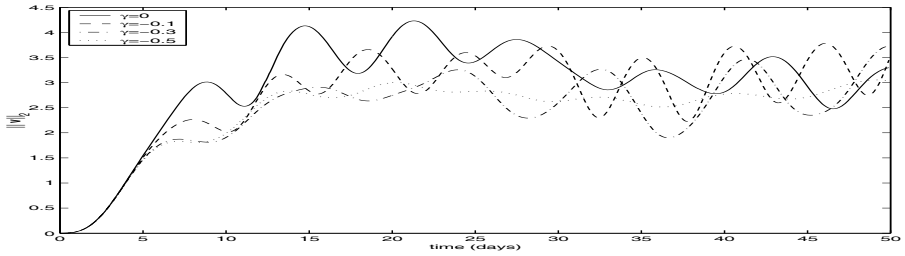

Figure 6.5.18: L_2 norm of the errors in the v velocity field for runs with different values of γ.

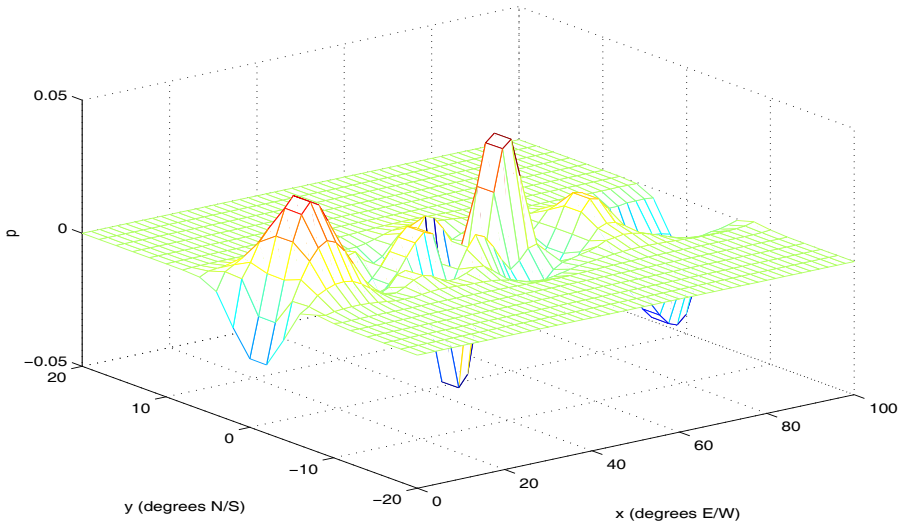

Figure 6.5.19: Errors in p after performing OI with pressure correction for 50 days

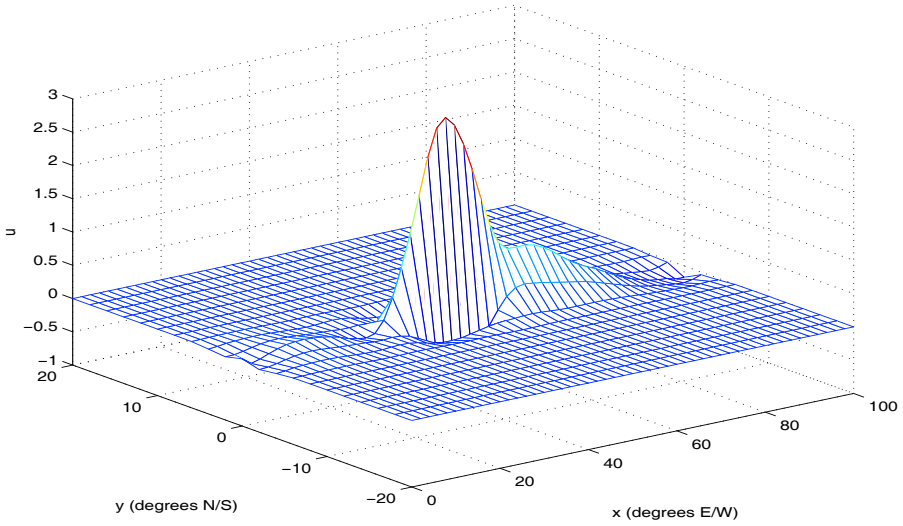

Figure 6.5.20: Errors in u velocity after performing OI with pressure correction for 50 days

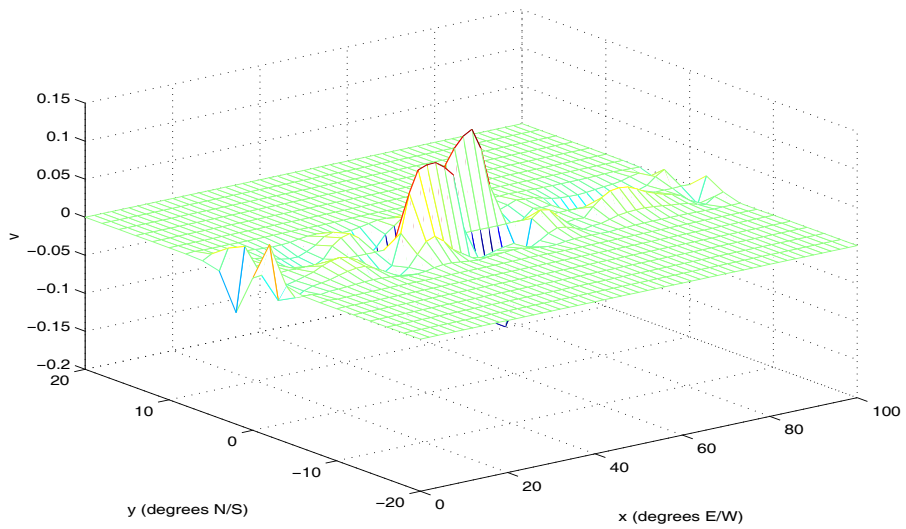

Figure 6.5.21: Errors in v velocity after performing OI with pressure correction for 50 days

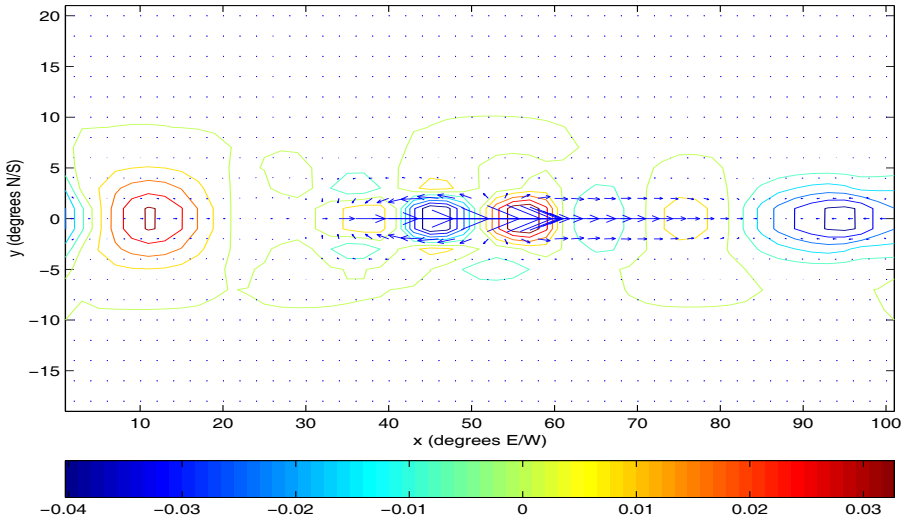

Figure 6.5.22: Errors in p (contours) and currents (arrows) after performing OI with pressure correction for 50 days

6.5.3 Summary

Figures 6.5.23, 6.5.24 and 6.5.25 show the L_2 norm of the p, u and v variables respectively for the four experiments described. All the experiments assimilating data show a marked improvement in the errors in the variable p over the solution without assimilation, as is to be expected when the observations are of the variable p. The pressure correction method produced slightly more accurate estimates of the variable p than all the other experiments.

The u variable shows very different results. All the runs with data assimilation included have higher errors than the run without assimilation. This is due to the fact that the data assimilation excites stronger velocities in the presence of systematic errors in the wind forcing. The method with a simple bias correction does no better than the normal OI data assimilation here. The pressure correction does reduce the errors in u as it was designed to do, although it still does not do as well as was expected. This could be due to the fact that the data is sparse compared to the size of the domain and so there is not enough information in the observations for the pressure correction field to be very accurate. Also, a multivariate data assimilation scheme might produce better results if knowledge of how u and p are related is used.

The results for the v variable show that for the first 40 days of integration, the experiments with assimilation included again do worse than without assimilation. Towards the end of the period however, all of the experiments have approximately the same amount of error, although there is more variability in the assimilation experiments.

It would be possible to make many improvements on the results of the

data assimilation experiments. As has already been said, a scheme which updated u and v based on the pressure observations would produce better results. Also, more observations spread throughout the domain would reduce the errors. In reality, there are more observations than the number used in these experiments, as data is obtained from ships and satellite altimeters, as well as the TAO array. However, the pressure correction method has certainly improved the results over the normal data assimilation and bias correction experiments here.

Key to Figures 6.5.23 - 6.5.25: no assimilation (solid line), normal OI (dashed line), bias correction; $\alpha = 1 \times 10^{-4}$ (dash-dot line), pressure correction; $\gamma = -0.1$ (dotted line).

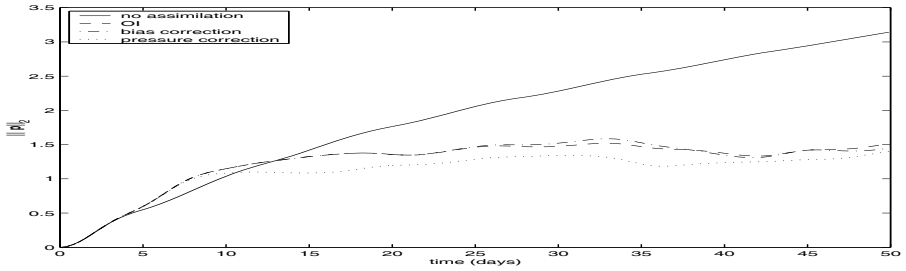

Figure 6.5.23: Time series of the L_2 norm of the errors in the pressure field.

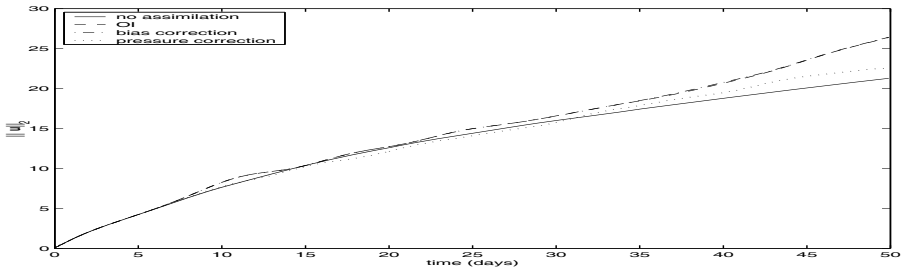

Figure 6.5.24: Time series of the L_2 norm of the errors in the u velocity field.

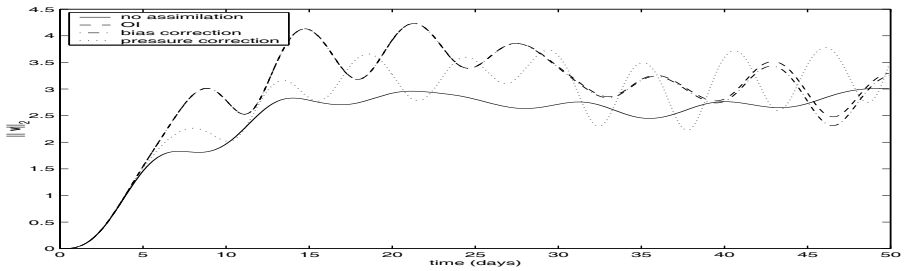

Figure 6.5.25: Time series of the L_2 norm of the errors in the v velocity field.

6.6 Conclusions

In the first part of this chapter, the linear, two-dimensional shallow water equations were derived on a β-plane. We assumed that we had the correct continuous forecast model apart from incorrectly specified wind stresses. A simple form of continuous data assimilation, where the density was assumed to be known everywhere, was shown not to account for the incorrect wind stresses. The bias correction method of the previous chapter was then used to attempt to correct for these incorrect wind stresses. A special form of this bias correction was also put forward as a simple and easy to implement method which could account for errors in the wind stresses. An attempt to analyse this pressure correction method was then made. This analysis on the transient solutions of the homogeneous equations was done to determine the stability requirements of the method. It was shown that the errors in the solution would converge to zero for the types of waves expected in the tropical Pacific Ocean, as long as the weighting coefficient for the pressure correction method is less than that for the data assimilation, i.e. $\epsilon - \gamma > 0$.

To verify these theoretical results, a numerical model of the linear, two-dimensional shallow water equations on a β-plane was used with systematic errors in the wind forcing. Observations of one variable only were available at the approximate locations of the TAO array in the tropical Pacific, which are sparse compared to the size of the domain. The errors in the pressure variable were reduced for all the runs which included data assimilation. However, normal data assimilation produced large errors in u and v, as did the normal bias correction method with a simple form of weighting matrix. The pressure correction method reduced the errors in u and v, although the sparseness of

the observations and the simple form of the weighting matrix meant that the reduction of errors was not as significant as was hoped.

Systematic errors in the wind forcing and its parameterisation was described in Chapter 2 to be a major source of error in the tropical Pacific when assimilating data. We would therefore like to test the pressure correction method in a three-dimensional Primitive equation model of the ocean.

Chapter 7

Experiments with FOAM System

7.1 Introduction

One of the main aims of this thesis is to produce a method which will account for systematic errors in the wind forcing and its parameterisation in a Primitive equation ocean model, using data assimilation. In the previous chapter, the pressure correction method was described in terms of a simple two dimensional shallow water model and was shown, both theoretically and numerically, to produce more accurate analyses than normal data assimilation when systematic errors were present in the wind forcing of the model. We now test this method in a Primitive equation, three dimensional model. To do this we use the Forecasting Ocean-Atmosphere Model (FOAM) developed and used operationally at the Met. Office. In [4], it is shown that the FOAM system and forcing fluxes contain systematic errors. A compar-

ison between another three ocean general circulation models (OGCMs) and their response to wind forcing in the tropical Pacific is given in [28]. One of their conclusions is that some of the physics is incorrectly represented in the OGCMs which leads to unreliable responses to a given wind forcing. Some attempts to account for model biases have been made in [77] and [84] for instance. In the first paper, a type of ensemble approach is used to estimate the climate drift of a coupled ocean-atmosphere model whilst in the second, variational assimilation is used to estimate some of the important parameters which affect the impact of the wind forcing. In [66], it is suggested that the assumption of serially uncorrelated model errors should be relaxed when assimilating data. The pressure correction method attempts to account for the serially correlated errors in the wind forcing and the way it is parameterised by the ocean model.

In Section 7.2, we describe the relevant components of the Bryan-Cox model used in the FOAM model and the discretisation of these equations. The Analysis Correction scheme, briefly described in Chapter 4, is rewritten in Section 7.3 for the approximations used in the FOAM system. A summary of the FOAM system is also given. A physical explanation of the reasons systematic errors in the wind forcing cause problems is given in Section 7.3.2, together with the results of some experiments with and without data assimilation. The way in which we implement the pressure correction method to tackle these problems is described in Section 7.4. In Section 7.4.2, we describe the results of the experiments. Conclusions are given in Section 7.5.

7.2 The FOAM model

The FOAM model is based on the Bryan-Cox code for a Primitive equation, three dimensional ocean. The model is described in [5], but a summary is given here. The equations are based on those given in Section 2.2.1 but are written in terms of spherical coordinates ϕ and λ which are the latitude and longitude respectively. The vertical coordinate is depth, z.

Basic equations

The equations of motion are

$$\frac{\partial u}{\partial t} + \Gamma(u) - fv = -\frac{m}{a}\frac{\partial}{\partial \lambda}(\frac{p}{\rho_0}) + F^{\lambda}, \tag{7.2.1}$$

$$\frac{\partial v}{\partial t} + \Gamma(v) + fu = \frac{1}{a}\frac{\partial}{\partial \phi}(\frac{p}{\rho_0}) + F^{\phi}, \tag{7.2.2}$$

where $m = sec\phi$, $f = 2\Omega sin\phi$, $u = (a/m)\frac{\partial \lambda}{\partial t}$, $v = a\frac{\partial \phi}{\partial t}$, a is the radius of the earth and ρ_0 is the density of sea water at surface pressure and standard temperature and salinity. The advective operator Γ is defined by

$$\Gamma(\mu) = \frac{m}{a}\left[\frac{\partial}{\partial \lambda}(u\mu) + \frac{\partial}{\partial \phi}(v\frac{\mu}{m})\right] + \frac{\partial}{\partial z}(w\mu). \tag{7.2.3}$$

The hydrostatic equation is

$$\frac{\partial p}{\partial z} = -\rho g, \tag{7.2.4}$$

and the continuity equation is

$$\frac{\partial w}{\partial z} + \frac{m}{a}[\frac{\partial u}{\partial \lambda} + \frac{\partial}{\partial \phi}(\frac{v}{m})] = 0. \tag{7.2.5}$$

Tracers such as potential temperature, θ, and salinity S are modelled by the conservation equations

$$\frac{\partial \theta}{\partial t} + \Gamma(\theta) = F^{\theta}, \tag{7.2.6}$$

$$\frac{\partial S}{\partial t} + \Gamma(S) \;=\; F^S. \tag{7.2.7}$$

The equation of state has the form

$$\rho = \rho(\theta, S, z), \tag{7.2.8}$$

where the function is described by a polynomial fit to the Knudsen formula. More details of this expression are given in [35].

Forcing terms

The terms F^λ, F^ϕ, F^θ and F^S represent the effects of surface forcing, turbulent mixing and diffusion, and are given by

$$\begin{aligned} F^\lambda &= A_{MV}\frac{\partial^2 u}{\partial z^2} + \frac{A_{MH}}{a^2}[m^2\frac{\partial^2 u}{\partial \lambda^2} + m\frac{\partial}{\partial \phi}(\frac{1}{m}\frac{\partial u}{\partial \phi}) \\ &+ (1-m^2n^2)u - 2nm^2\frac{\partial v}{\partial \lambda}], \end{aligned} \tag{7.2.9}$$

$$\begin{aligned} F^\phi &= A_{MV}\frac{\partial^2 v}{\partial z^2} + \frac{A_{MH}}{a^2}[m^2\frac{\partial^2 v}{\partial \lambda^2} + m\frac{\partial}{\partial \phi}(\frac{1}{m}\frac{\partial v}{\partial \phi}) \\ &+ (1-m^2n^2)v + 2nm^2\frac{\partial u}{\partial \lambda}], \end{aligned} \tag{7.2.10}$$

$$F^\theta = \frac{A_{\theta V}}{\delta}\frac{\partial^2 \theta}{\partial z^2} + \frac{A_{\theta H}}{a^2}[m^2\frac{\partial^2 \theta}{\partial \lambda^2} + m\frac{\partial}{\partial \phi}(\frac{1}{m}\frac{\partial \theta}{\partial \phi})], \tag{7.2.11}$$

$$F^S = \frac{A_{SV}}{\delta}\frac{\partial^2 S}{\partial z^2} + \frac{A_{SH}}{a^2}[m^2\frac{\partial^2 S}{\partial \lambda^2} + m\frac{\partial^2}{\partial \phi}(\frac{1}{m}\frac{\partial^2 S}{\partial \phi})], \tag{7.2.12}$$

where

$$\delta = \begin{cases} 1 & \text{if } \frac{\partial \rho''}{\partial z} < 0 \\ 0 & \text{if } \frac{\partial \rho''}{\partial z} > 0\ , \end{cases} \tag{7.2.13}$$

and ρ'' is the density which a parcel of water would have if the *in situ* pressure is reduced to surface pressure. Having $\frac{\partial \rho''}{\partial z} > 0$ implies that the fluid is unstable so setting $\delta = 0$ means that there is infinite mixing. The mixing coefficients A_{MV}, A_{MH}, $A_{\theta V}$, $A_{\theta H}$, A_{SV} and A_{SH} are assumed to be given.

Boundary conditions

Boundary conditions at lateral walls are given by

$$u = v = 0, \tag{7.2.14}$$

$$\frac{\partial \theta}{\partial n} = \frac{\partial S}{\partial n} = 0, \tag{7.2.15}$$

where $\frac{\partial}{\partial n}$ denotes a local derivative with respect to the coordinate normal to the wall. At the surface, the rigid-lid approximation is made, where $w = 0$. This has the advantage of filtering out high speed external gravity waves, and so allows the time-step of the numerical integration to be larger, [10]. Also at the surface, we set

$$\rho_0 A_{MV} \frac{\partial u}{\partial z} = \tau^{\lambda}, \tag{7.2.16}$$

$$\rho_0 A_{MV} \frac{\partial v}{\partial z} = \tau^{\phi}, \tag{7.2.17}$$

$$A_{\theta V} \frac{\partial \theta}{\partial z} = \eta^{\theta}, \tag{7.2.18}$$

$$A_{SV} \frac{\partial S}{\partial z} = \eta^{S}, \tag{7.2.19}$$

where τ^{λ} and τ^{ϕ} are the zonal and meridional components of the surface wind stress and η^{θ} and η^{S} are the fluxes through the surface of temperature and salinity respectively. For the lower boundary, the condition is set to take bottom friction into account.

Splitting the velocities

In the model, the horizontal momentum equations are combined with the hydrostatic equation. This enables the horizontal velocities to be split into

a baroclinic component, which varies with depth, and a barotropic streamfunction. Using the continuity equation (7.2.5), and the boundary conditions for w at the surface and bottom, it is possible to define a streamfunction as

$$\frac{\partial \psi}{\partial \lambda} = \frac{a}{m}\int_{-H}^{0} \rho_0 v dz, \tag{7.2.20}$$

$$\frac{\partial \psi}{\partial \phi} = -a\int_{-H}^{0} \rho_0 u dz. \tag{7.2.21}$$

To obtain a predictive equation for this streamfunction, the horizontal momentum equations, (7.2.1) and (7.2.2) are integrated with respect to height, z, and multiplied by $\frac{a\rho_0}{mH}$ and $\frac{a\rho_0}{H}$ respectively. They can then be combined into one equation,

$$\begin{aligned}\frac{\partial^2}{\partial\lambda\partial t}(\frac{m}{H}\frac{\partial\psi}{\partial\lambda}) &+ \frac{\partial^2}{\partial\phi\partial t}(\frac{1}{Hm}\frac{\partial\psi}{\partial\phi}) = \frac{\partial}{\partial\lambda}(FV) - \frac{\partial}{\partial\phi}(\frac{FU}{m}) \\ &- \frac{\partial\psi}{\partial\lambda}\frac{\partial}{\partial\phi}(2\Omega n/H) + \frac{\partial\psi}{\partial\phi}\frac{\partial}{\partial\lambda}(2\Omega n/H),\end{aligned} \tag{7.2.22}$$

where

$$FU = -\frac{a\rho_0}{H}\int_{-H}^{0}[\Gamma(u) - \frac{mn}{a}uv - F^{\lambda} + \frac{gm}{a\rho_0}\int_{z}^{0}\frac{\partial\rho}{\partial\lambda}dz']dz, \tag{7.2.23}$$

$$FV = -\frac{a\rho_0}{H}\int_{-H}^{0}[\Gamma(v) + \frac{mn}{a}u^2 - F^{\phi} + \frac{g}{a\rho_0}\int_{z}^{0}\frac{\partial\rho}{\partial\phi}dz']dz. \tag{7.2.24}$$

We write the horizontal velocities as

$$u = \bar{u} + \hat{u}, \tag{7.2.25}$$

$$v = \bar{v} + \hat{v}, \tag{7.2.26}$$

where the overbar indicates a vertical average over a water column and the hat indicates the deviation from this average. The vertically averaged terms

can be written as

$$\bar{u} = \frac{1}{H}\int_{-H}^{0} u dz = -\frac{H}{\rho_0 a}\frac{\partial \psi}{\partial \phi}, \tag{7.2.27}$$

$$\bar{v} = \frac{1}{H}\int_{-H}^{0} v dz = \frac{Hm}{\rho_0 a}\frac{\partial \psi}{\partial \lambda}. \tag{7.2.28}$$

To calculate the deviation from this average we set

$$\hat{u} = u' - \bar{u'}, \tag{7.2.29}$$

$$\hat{v} = v' - \bar{v'}, \tag{7.2.30}$$

where $\bar{u'}$, $\bar{v'}$ denote the vertical average of the deviations and u', v' are obtained from the horizontal momentum equations and the hydrostatic equation, giving

$$\frac{\partial u'}{\partial t} = -\Gamma(u) + fv - \frac{mg}{\rho_0 a}\frac{\partial}{\partial \lambda}\left(\int_z^0 \rho \, dz'\right) + F^{\lambda}, \tag{7.2.31}$$

$$\frac{\partial v'}{\partial t} = -\Gamma(v) - fu - \frac{g}{\rho_0 a}\frac{\partial}{\partial \phi}\left(\int_z^0 \rho \, dz'\right) + F^{\phi}. \tag{7.2.32}$$

Numerical procedure

The procedure for the numerical integration of these equations is summarised as follows:

1. Predict the temperature, θ and salinity, S at the new time step from a discretisation of equations (7.2.6) and (7.2.7).

2. Predict u' and v' from discretisations of equations (7.2.31) and (7.2.32) and use these to find the new values of $\hat{u}$ and $\hat{v}$ from equations (7.2.29) and (7.2.30).

3. Predict the new value of the streamfunction, ψ, from a discretisation of equation (7.2.22) and use this in equations (7.2.27) and (7.2.28) to calculate $\bar{u}$ and $\bar{v}$. The full velocity field is then given by equations (7.2.25) and (7.2.26).

4. Use the continuity equation (7.2.5) and the equation of state (7.2.8) to find w and ρ at the new time level.

The horizontal grid used for the discretisation of these equations is a B-grid, [10], where the baroclinic velocities u and v are held at points i, j, and variables θ, S and the barotropic streamfunction ψ are held at points $i+1/2$, $j+1/2$. In the vertical, model levels are close together near the surface with the resolution decreasing with depth. Spatial derivatives are approximated by centred differences. The time differencing is done by using the leap-frog scheme, [67], where

$$\frac{\partial \mu_k}{\partial t} = \frac{\mu_{k+1} - \mu_{k-1}}{2\Delta t}, \tag{7.2.33}$$

for a variable μ, where k denotes time step t_k and Δt is the length of the time step. The horizontal resolution of the model is $1^0 \times 1^0$ and there are 20 vertical layers. A time step of one hour is used. More details of the numerical scheme are given in [5].

The resolution of the model means that eddies, which are an important mechanism for the transfer of energy, are not resolved. Mixing in the vertical is also not well resolved by the model. Much of the accuracy of the model is therefore dependent on the parameterisations of these sub-grid scale processes. This is especially important in the mixed layer, the region above the thermocline which is approximately 100m deep, where most of the vertical

mixing takes place. The details of these parameterisations can be found in [5] and references therein.

Supplied forcings

The surface forcing fluxes used to drive the model are available from the Numerical Weather Prediction (NWP) model. In the operational version, FOAM is forced using 6-hourly mean fluxes derived from the global NWP model analyses. Monthly mean climatological fluxes are also available.

The wind forcing enters the model through equations (7.2.16), (7.2.17) and (7.2.9), (7.2.10). This process involves parameterisations which approximate the vertical mixing of momentum. An important part of these parameterisations is the choice of the mixing coefficients A_{MV}. Systematic errors in the supplied wind forcing and errors in the parameterisations will therefore both contribute to the systematic errors in the forcing of the model.

7.3 Assimilation of temperature data

7.3.1 Assimilation scheme

The data assimilation scheme used in FOAM is an approximation to the Analysis Correction scheme described in Section 4.2.3 where corrections are only made to the model state, not the observations. We can therefore write the scheme as

$$\underline{x}_k^{j+1} = \underline{x}_k^j + \lambda W_k V_k [\underline{y}_k^j - \underline{h}_k(\underline{x}_k^j)], \tag{7.3.1}$$

where λ is a relaxation factor which controls the step size of the iterations, $W_k = B_k H_k^T R_k^{-1}$, $V_k = (H_k W_k + I)^{-1}$ and B_k, H_k, R_k are as before.

A number of approximations are made to this scheme, [3]:

- V_k is approximated by a diagonal matrix of normalisation factors. This affects the rate of convergence of the scheme, but not the final limit, [61].

- The relaxation factor, λ and the number of iterations are chosen so that a single observation would give results close to the OI solution.

- The background error covariance matrix B_k is approximated. It now depends on the distance between observation and model grid points and on a correlation scale which varies with location and direction.

There are four main types of data assimilated into FOAM, sea surface temperature (SST), temperature profile data, salinity and surface height. These different data groups are assimilated sequentially. The first group is the SST data, which are used to form analysis increments which are applied within the mixed layer. The temperature profile data is used next, where increments are applied down to 1000m, below which there are very few observations. Velocity increments are calculated which balance the temperature increments through the hydrostatic balance and a modified geostrophic relation. SST observations are assimilated into the model five days either side of their validity time, while temperature profile data are nudged in over a longer period of ten days. The other two data groups are not used in the experiments of this chapter and so details of their assimilation are omitted

here. For more details of the practical implementation of this assimilation scheme, see [5].

Summary

A summary of the forecast/analysis procedure used by FOAM is written here for comparison with the modified method which is described in Section 7.4.

1. Starting from a forecast $\underline{x}_k^f = \underline{x}_k^0$ at time t_k, use observations of temperature to produce an analysis using the following equation:

$$\underline{x}_k^{j+1} = \underline{x}_k^j + \lambda W_k V_k [\underline{y}_k - H_k \underline{x}_k^j], \tag{7.3.2}$$

where $\underline{x}_k^a = \underline{x}_k^i$ is the analysis after i iterations. This can be written as

$$\begin{pmatrix} \underline{\theta}_k^{j+1} \\ \underline{S}_k^{j+1} \\ \underline{u}_k^{j+1} \\ \underline{v}_k^{j+1} \end{pmatrix} = \begin{pmatrix} \underline{\theta}_k^j \\ \underline{S}_k^j \\ \underline{u}_k^j \\ \underline{v}_k^j \end{pmatrix} + \lambda \begin{pmatrix} W_k^\theta V_k^\theta \\ W_k^S V_k^S \\ W_k^u V_k^u \\ W_k^v V_k^v \end{pmatrix} [\underline{\theta}_k^{obs} - (H_k^\theta, 0, 0, 0) \begin{pmatrix} \underline{\theta}_k^j \\ \underline{S}_k^j \\ \underline{u}_k^j \\ \underline{v}_k^j \end{pmatrix}], \tag{7.3.3}$$

where $W_k^\theta V_k^\theta$ is the weighting for the temperature observation increments, $W_k^u V_k^u$, $W_k^v V_k^v$ convert the temperature observation increments into velocity increments, $W_k^S = V_k^S = 0$ and $\underline{y}_k = \underline{\theta}_k^{obs} = H_k \underline{x}_k^t + \underline{\delta}_k = H_k^\theta \underline{\theta}_k^t + \underline{\delta}_k^\theta$.

2. Calculate $\underline{w}_k$, $\underline{\rho}_k$ from the continuity equation and the equation of state respectively, written here as:

$$\underline{w}_k = \underline{g}_k^w(\underline{u}_k^a, \underline{v}_k^a), \tag{7.3.4}$$

$$\underline{\rho}_k = \underline{g}_k^\rho(\underline{\theta}_k^a, \underline{S}_k^a). \tag{7.3.5}$$

3. Forecast the state variables onto time t_{k+1}:

$$\underline{x}^f_{k+1} = \underline{f}_k(\underline{x}^a_k) + \underline{F}_k. \tag{7.3.6}$$

This can be written as

$$\begin{pmatrix} \underline{\theta}^f_{k+1} \\ \underline{S}^f_{k+1} \\ \underline{u}^f_{k+1} \\ \underline{v}^f_{k+1} \end{pmatrix} = \begin{pmatrix} \underline{f}^\theta_k(\underline{\theta}^a_k, \underline{u}^a_k, \underline{v}^a_k) \\ \underline{f}^S_k(\underline{S}^a_k, \underline{u}^a_k, \underline{v}^a_k) \\ \underline{f}^u_k(\underline{\theta}^a_k, \underline{S}^a_k, \underline{u}^a_k, \underline{v}^a_k) \\ \underline{f}^v_k(\underline{\theta}^a_k, \underline{S}^a_k, \underline{u}^a_k, \underline{v}^a_k) \end{pmatrix} + \begin{pmatrix} \underline{F}^\theta_k \\ \underline{F}^S_k \\ \underline{F}^u_k \\ \underline{F}^v_k \end{pmatrix}. \tag{7.3.7}$$

7.3.2 Response of the model to assimilation of temperature

In this section, we give a physical explanation of the effects of systematic errors in the wind stresses and their parameterisation in the vertical. We reinforce these arguments using results obtained from two integrations of the FOAM system, which show the impacts of using normal data assimilation in the equatorial Pacific when compared with not assimilating data. These results are given at the end of this chapter to enable comparison with the experiments described in Section 7.4.2.

We first describe the set-up of the experiments performed in this section and in Section 7.4.2. The integrations start from initial conditions derived from the Levitus (1994) climatology, [55], for 1st May. They are driven by climatological monthly mean fluxes including the surface wind stresses of Hellerman and Rosenstein, [42]. It is generally accepted that these wind stresses are too strong in the equatorial region, [47]. These experiments therefore contain larger systematic errors than would be present operationally, but

the same problems exist in operational runs. The assimilation integrations use observations of SST and temperature profiles taken from 1st May 1995 to 30th April 1996, a period in which there were no El Niño or La Niña events. All of the integrations are run for two years.

The dominant balance in the zonal direction in ocean models of the equatorial Pacific is between the wind forcing and a pressure gradient. The normal direction of winds in this region are towards the west. The stresses are mixed down over the top 100m of the ocean and result in an east-west pressure gradient. A schematic diagram of this balance is shown in Figure 7.3.1(a). When no observations are assimilated into the model, a balance between these forces is found. However, the temperature structure of the model is inaccurate, the vertical gradient of temperature in the thermocline being too weak. This can be seen in Figure 7.4.4(a) for the FOAM system. The vertical velocities for this control run are presented in Figures 7.4.8(a) and 7.4.10(a), which show that there is little upwelling or downwelling. The u velocities are shown in Figure 7.4.12(a) where the eastward flowing Equatorial Undercurrent (EUC) can be clearly seen extending right across the region with a maximum in the region of the thermocline. The EUC is quite diffuse here, due to the weak temperature gradients of the thermocline. The westward surface currents can also be seen between 160^0E and 80^0W.

We would now like to assimilate temperature observations to produce a better estimate of the temperature structure. When observations of temperature are assimilated into the model, the thermocline is tighter, as can be seen in Figure 7.4.4(b). The assimilation of these observations also gives the model a good estimate of the pressure gradients. However, systematic errors

in the wind stress and its parameterisation in the vertical lead to an imbalance between these forces and the pressure gradient, a schematic of which is given in Figure 7.3.1(b). This imbalance manifests itself as large temperature increments put in by the assimilation and large vertical velocities. These fields can be seen in Figures 7.4.6 and 7.4.8(b) for the integration using normal data assimilation. The u velocities are shown in Figure 7.4.12(b) for this run. The EUC is now much stronger than without data assimilation due to the tighter thermocline, but does not extend as far to the east as it should. This is due to the stronger vertical velocities in the region of 100^0W. The surface currents are mainly flowing to the west, as we would expect, but there is an eastward flowing section at about 160^0W.

We now give an explanation for the large vertical velocities and temperature increments. The incorrect wind stresses attempt to drive the model's temperature field to a different equilibrium than that given by the temperature observations. This means that the assimilation has to change the temperature substantially. Altering the temperature at some depth will lead to a density change through the equation of state. Hydrostatic balance then implies that the pressure above and below this level will be changed. This results in a divergent/convergent horizontal flow in the North-South direction above the region and a convergent/divergent flow below. The ocean model has non-divergent 3D flow and so these horizontal velocities produce upwelling/downwelling. These vertical velocities will then act to increase/decrease the density at the level in question, and so produce a dynamical change to the density which is of the opposite sign to the one which produced the circulation in the first place. At the next assimilation step,

the observations will lead to another large increment and so the process will begin again.

7.4 Pressure correction method

The method for systematic model error correction described in Chapter 5 was applied in Chapter 6 to the two dimensional shallow water equations. A special case of the method was also described, whereby a correction to the pressure field was used to account for errors in the forcing of the equations. The application of this pressure correction method was shown in Chapter 6 for systematic errors in the forcing of the linear, two-dimensional equations. Here, we describe an implementation of the method for the nonlinear, three-dimensional Primitive equation model together with the Analysis Correction data assimilation scheme described in Sections 7.2 and 7.3.

The problems of using the normal AC data assimilation scheme described in the previous section arise from an imbalance between incorrect wind stresses and the pressure gradient, which is altered by the assimilation of temperature data. The schematic in Figure 7.3.1(c) shows how the pressure correction method should restore the balance between these two forces. This is done, as before, by calculating a corrected pressure field which alters the effective wind forcing in the horizontal momentum equations.

7.4.1 Implementation of the pressure correction method

In FOAM, the pressure is not a state variable as it was in Chapter 6. Also, we are only dealing with observations of temperature, so the correction to the

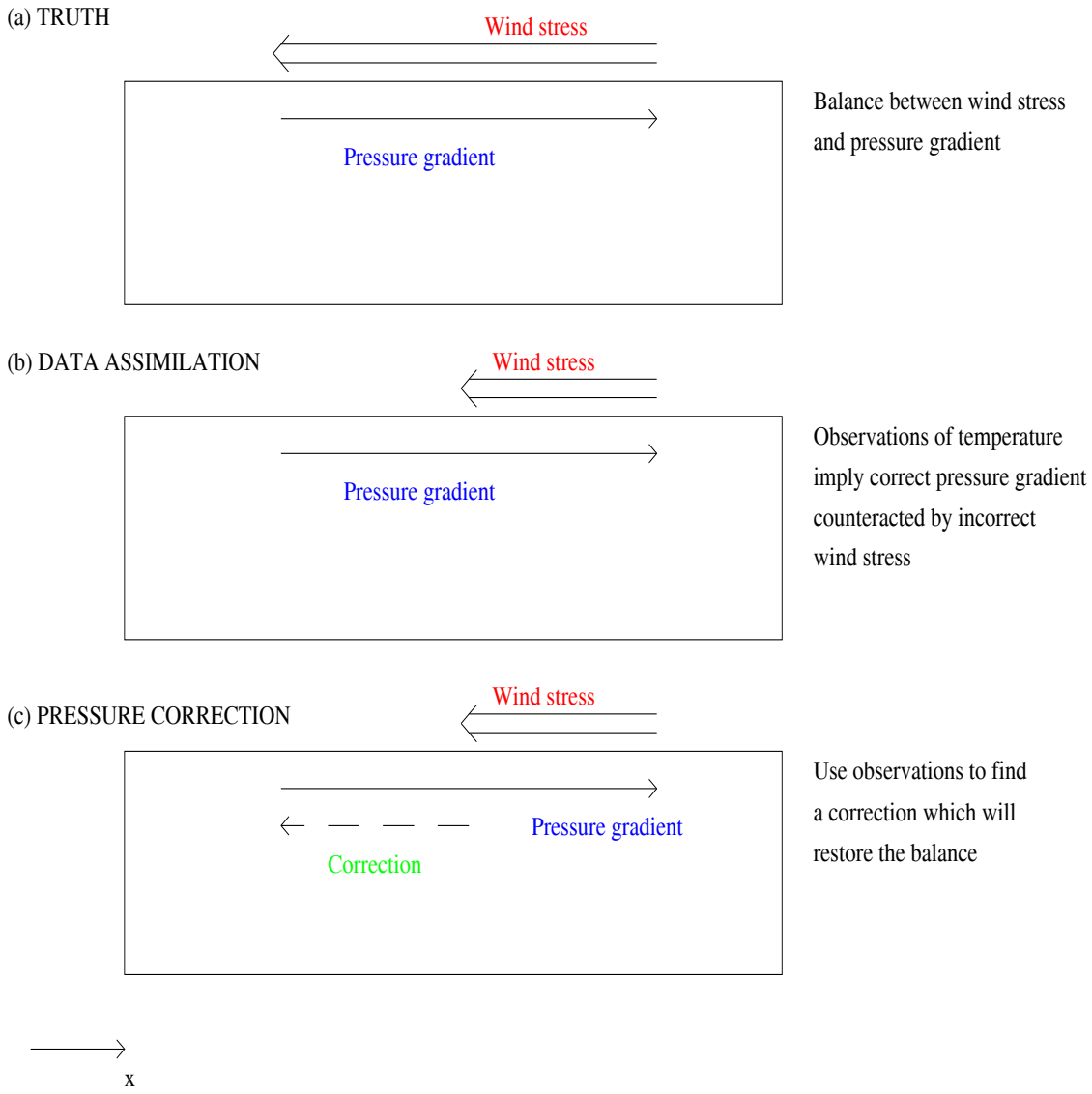

Figure 7.3.1: Schematics of the main balances in the equatorial Pacific and the effects of performing data assimilation and pressure correction on these balances.

pressure field has to be calculated through a compensating temperature field or temperature bias, $\underline{\theta}^c$. In the assimilation, the analysis of this compensating temperature field is calculated, as another state variable, by adding to its forecast a small weighting of the difference between the forecast and observed temperatures, so that

$$\underline{\theta}_k^{c\,a} = \underline{\theta}_k^{c\,f} + \lambda W_k^c V_k^c [\underline{\theta}_k^{obs} - H_k^\theta \underline{\theta}_k^f]. \tag{7.4.1}$$

The forecast of this compensating temperature field is made by keeping it constant from one time step to the next, i.e. $\underline{\theta}_{k+1}^{c\,f} = \underline{\theta}_k^{c\,a}$. It is difficult to obtain an accurate estimate of the weighting matrices W_k^c, V_k^c. We therefore make the following approximation:

$$W_k^c V_k^c = \gamma W_k^\theta V_k^\theta. \tag{7.4.2}$$

The model next calculates a new density field through the equation of state, where the temperature input is altered to be the sum of the temperature and the compensating temperature field, i.e.

$$\underline{\rho}_k = \underline{\rho}_k(\underline{\theta}_k^a + \underline{\theta}_k^{c\,a}, \underline{S}_k, z). \tag{7.4.3}$$

However, the model's temperature field is not affected by this.

The pressure field is now calculated from this new density field using the hydrostatic equation. Finally, the corrected pressure field is used in the calculation of the velocities through the horizontal momentum equations to balance the systematic errors in the forcing of those equations introduced by the wind stresses. A summary of this new forecast/analysis procedure is given below.

1. Starting from a forecast $\underline{z}_k^0 = \underline{z}_k^f$ at time t_k, use observations of temperature to produce an analysis using the following equation:

$$\underline{z}_k^{j+1} = \underline{z}_k^j + \lambda \tilde{W}_k \tilde{V}_k [\underline{y}_k - \underline{\tilde{H}}_k \underline{z}_k^j], \tag{7.4.4}$$

where $\underline{z}_k^j = \begin{pmatrix} \underline{x}_k^j \\ \underline{\theta}_k^{c\,j} \end{pmatrix}$ and $\underline{z}_k^a = \underline{z}_k^i$ is the analysis after i iterations. This can be written as

$$\begin{pmatrix} \underline{\theta}_k^{j+1} \\ \underline{S}_k^{j+1} \\ \underline{u}_k^{j+1} \\ \underline{v}_k^{j+1} \\ \underline{\theta}_k^{c\,j+1} \end{pmatrix} = \begin{pmatrix} \underline{\theta}_k^j \\ \underline{S}_k^j \\ \underline{u}_k^j \\ \underline{v}_k^j \\ \underline{\theta}_k^{c\,j} \end{pmatrix} + \lambda \begin{pmatrix} W_k^\theta V_k^\theta \\ W_k^S V_k^S \\ W_k^u V_k^u \\ W_k^v V_k^v \\ \gamma W_k^\theta V_k^\theta \end{pmatrix} [\underline{\theta}_k^{obs} - (H_k^\theta, 0, 0, 0, 0) \begin{pmatrix} \underline{\theta}_k^j \\ \underline{S}_k^j \\ \underline{u}_k^j \\ \underline{v}_k^j \\ \underline{\theta}_k^{c\,j} \end{pmatrix}], \tag{7.4.5}$$

where $W_k^c V_k^c$ determines how much of the temperature increment signal is present in the compensating temperature field and $\underline{y}_k = \underline{\theta}_k^{obs} = \tilde{H}_k \underline{z}_k^t + \underline{\delta}_k = H_k^\theta \underline{\theta}_k^t + \underline{\delta}_k^\theta$.

2. Calculate $\underline{w}_k$, $\underline{\rho}_k$:

$$\underline{w}_k = \underline{g}_k^w(\underline{u}_k^a, \underline{v}_k^a), \tag{7.4.6}$$

$$\underline{\rho}_k = \underline{g}_k^\rho(\underline{\theta}_k^a + \underline{\theta}_k^{c\,a}, \underline{S}_k^a). \tag{7.4.7}$$

3. Forecast the state variables onto time t_{k+1}:

$$\underline{z}_{k+1}^f = \underline{\tilde{f}}_k(\underline{z}_k^a) + \underline{\tilde{F}}_k. \tag{7.4.8}$$

This can be written as

$$\begin{pmatrix} \underline{\theta}^f_{k+1} \\ \underline{S}^f_{k+1} \\ \underline{u}^f_{k+1} \\ \underline{v}^f_{k+1} \\ \underline{\theta}^{c\,f}_{k+1} \end{pmatrix} = \begin{pmatrix} \underline{f}^{\theta}_k(\underline{\theta}^a_k, \underline{u}^a_k, \underline{v}^a_k) \\ \underline{f}^{S}_k(\underline{S}^a_k, \underline{u}^a_k, \underline{v}^a_k) \\ \underline{f}^{u}_k(\underline{\theta}^a_k + \underline{\theta}^{c\,a}_k, \underline{S}^a_k, \underline{u}^a_k, \underline{v}^a_k) \\ \underline{f}^{v}_k(\underline{\theta}^a_k + \underline{\theta}^{c\,a}_k, \underline{S}^a_k, \underline{u}^a_k, \underline{v}^a_k) \\ \underline{\theta}^{c\,a}_k \end{pmatrix} + \begin{pmatrix} \underline{F}^{\theta}_k \\ \underline{F}^{S}_k \\ \underline{F}^{u}_k \\ \underline{F}^{v}_k \\ 0 \end{pmatrix} . \qquad (7.4.9)$$

7.4.2 Results of experiments

Two experiments were performed using the pressure correction method. These experiments start from the same initial conditions and contain the same forcing fields as those described in Section 7.3. In one of these experiments, the weighting coefficient, γ, is chosen to be -0.1 and in the other it is chosen to be -0.3.

The time mean potential temperature cross sections along the equator for the second year of integration are shown in Figure 7.4.5 for both the pressure corrected experiments. These show that the thermocline still contains the tighter temperature gradients that were obtained in the run with the normal data assimilation, although the structure is slightly different below the thermocline. Table 7.1 gives variances calculated from 1st May - 31st July of the second year of integration. These are for the differences between the temperature observations and the model values at the observation locations obtained before (forecast) and after (analysis) the observations are assimilated. These show that at 50m depth, the run with normal data assimilation ($\gamma = 0$) has the highest variance. The pressure corrected runs both have smaller errors at this depth. At 200m, all the runs have similar variances.

These results are significant as they show that we are fitting the observations better with the pressure correction method than with the normal data assimilation.

Variances Depth (m)	$\gamma = 0$	$\gamma = -0.1$	$\gamma = -0.3$
forecast - observations			
47.85	1.10	0.98	0.85
203.7	0.72	0.72	0.75
analysis - observations			
47.85	1.00	0.88	0.77
203.7	0.67	0.64	0.68

Table 7.1: Variances of model - observed temperatures taken from 1st May to 31st July in the second year of integration.

Figure 7.4.7 shows the time mean cross sections of temperature increments along the equator for the second year of integration. It is clear that the temperature increments are much smaller for the runs with the pressure correction method included than for the normal data assimilation run and are constrained to the top 150m. The maximum values for the pressure corrected runs are about 2^0C per month compared with 4^0C per month for the run with normal data assimilation. Figure 7.4.1 shows the RMS values of the monthly mean temperature increments over the two years of integration for the region $\pm 15^0$N/S, 140^0E-80^0W at 50m depth (note the different axis scales for the two figures). Both pressure corrected runs have much smaller

values over the entire time period, the average value being about half of that for the normal data assimilation run.

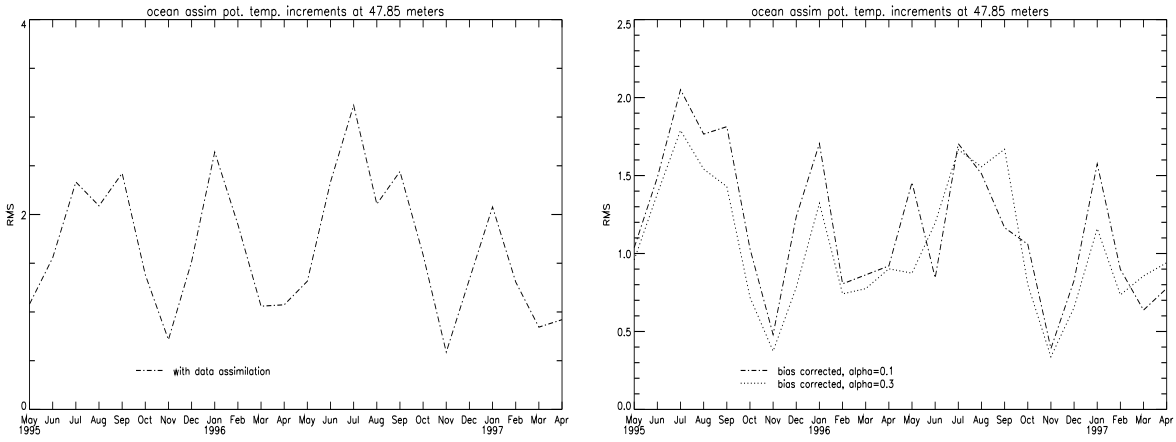

Figure 7.4.1: RMS plots of monthly mean temperature increments (^{0}C per month) at about 50m depth over region $\pm 15^0$N/S, 140^0E-80^0W: (a) normal assimilation (b) bias corrected - $\gamma = -0.1$ (dot-dashed), $\gamma = -0.3$ (dotted)

A vertical cross section of vertical velocities is shown across the equator in Figure 7.4.9 at 110^0W. Here, the large vertical velocities below 150m depth which were present in the run with normal data assimilation have been eliminated. A horizontal view of the vertical velocities at 250m depth is given in Figure 7.4.11. The pressure corrected runs both have significantly smaller vertical velocities at this depth than those for the run with normal data assimilation, as shown in Figure 7.4.10. The magnitudes are now comparable to the control run without data assimilation. To show that the overall magnitude of the vertical velocities has been reduced, the RMS of their monthly mean values are shown in Figure 7.4.2 for the region $\pm 15^0$N/S, 140^0E-80^0W at 250m depth. It can be seen that the pressure correction method has re-

duced the vertical velocities although the values are still not as small as those of the control run.

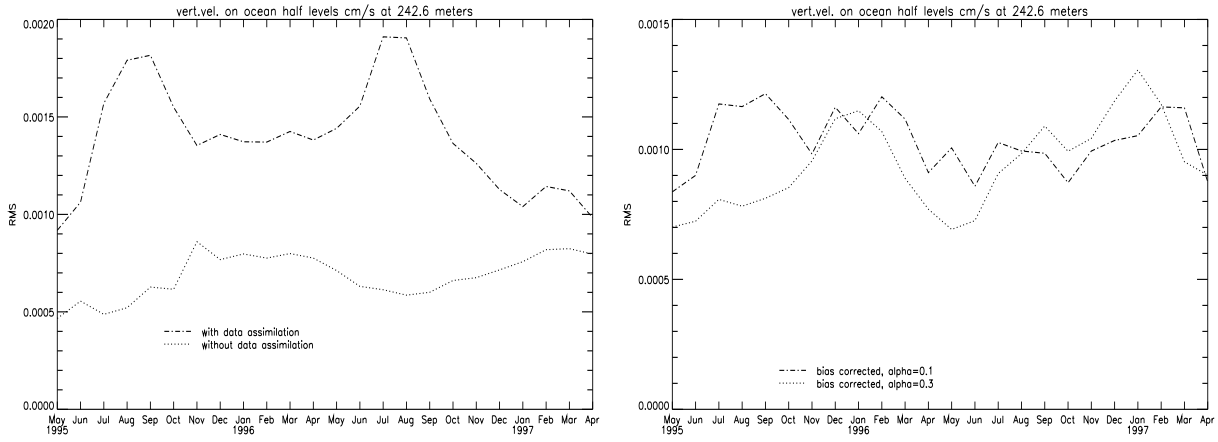

Figure 7.4.2: RMS plots of monthly mean vertical velocities (cm/s) at about 250m depth over region $\pm 15^0$N/S, 140^0E-80^0W: (a) no assimilation (dot-dashed), normal assimilation (dotted) (b) pressure corrected - $\gamma = -0.1$ (dot-dashed), $\gamma = -0.3$ (dotted)

The u velocities for the two pressure corrected runs are presented in Figure 7.4.13. In both experiments the EUC is stronger and less diffuse than for the control run, although the magnitudes are not as large as for the run with normal data assimilation. The EUC now extends to the east because of the reduction in the vertical velocities. The surface currents are now of similar structure to the control run.

The time mean cross sections along the equator for the second year of integration for the compensating temperature field are shown in Figure 7.4.14. The same structure is shown for both values of γ although the run with $\gamma = -0.3$ has much larger magnitude. Most of the field is concentrated in

the top 200m and is sloping down towards the west, the values being positive above the thermocline and negative below it. Figure 7.4.15 shows the compensating temperature field at about 50m depth for the region $\pm 15^0$N/S, 140^0E-80^0W. This shows that most of the field is concentrated within $10-15^0$ of the equator with larger values to the east of the region. The field is positive along the equator and to the south but there is a negative area to the north of the equator at about 120^0W. RMS values of the monthly mean compensating temperature field are given in Figure 7.4.3 for the region $\pm 15^0$N/S, 140^0E-80^0W at 200m depth. Here, the run with $\gamma = -0.1$ appears to converge after 8 or 9 months of integration. However, with $\gamma = -0.3$ the compensating temperature field does not appear to converge which might indicate that this choice of weighting is too large.

The pressure correction field has been calculated and is given in Figure 7.4.16 at the surface for the region $\pm 30^0$N/S, 140^0E-80^0W. This field does not appear to be confined to near the equator, especially with $\gamma = -0.3$, although the largest values are on or close to the equator. To calculate the effective correction to the u momentum equation, for instance, the gradient in the x direction of this field has to be calculated at each model level and then integrated up from the bottom. This has been done to calculate an estimate of the implied wind stress error fields in the x and y directions and is shown in Figures 7.4.17 and 7.4.18 respectively. Surprisingly, the errors in τ^y are larger than those for τ^x. This indicates that the assumption that most of the errors in the momentum equations come from incorrect wind stresses and their parameterisation in the vertical is not necessarily very accurate as there might also be systematic errors in the horizontal transport of momentum, for

instance.

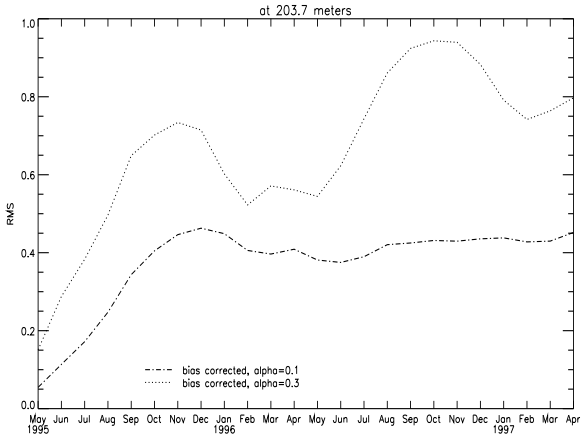

Figure 7.4.3: RMS plots of monthly mean compensating temperature field at about 200m depth over region $\pm 15^0$N/S, 140^0E-80^0W: $\gamma = -0.1$ (dot-dashed), $\gamma = -0.3$ (dotted)

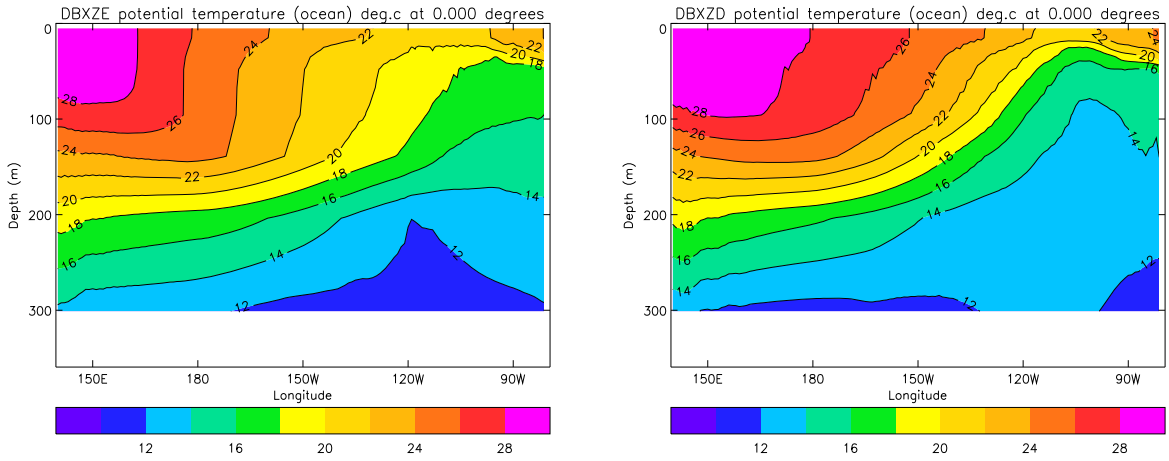

Figure 7.4.4: Annual mean (year 2) potential temperature (^{0}C) cross section along the equator: (a) no assimilation (b) normal assimilation

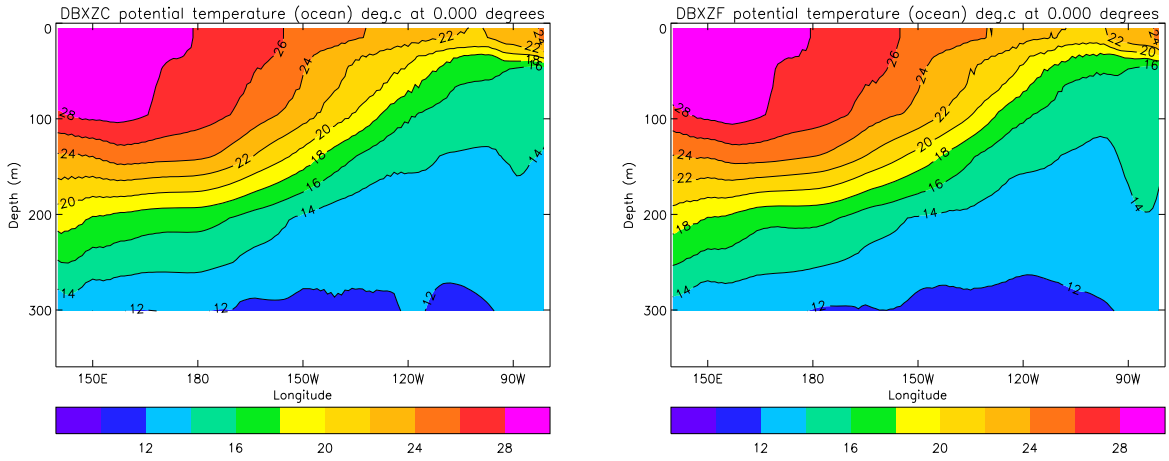

Figure 7.4.5: Annual mean (year 2) potential temperature (^{0}C) cross section along the equator: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

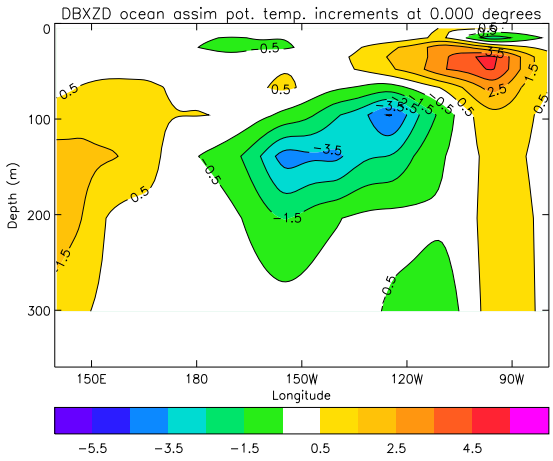

Figure 7.4.6: Annual mean (year 2) potential temperature increments (^{0}C per month) cross section along the equator: normal assimilation

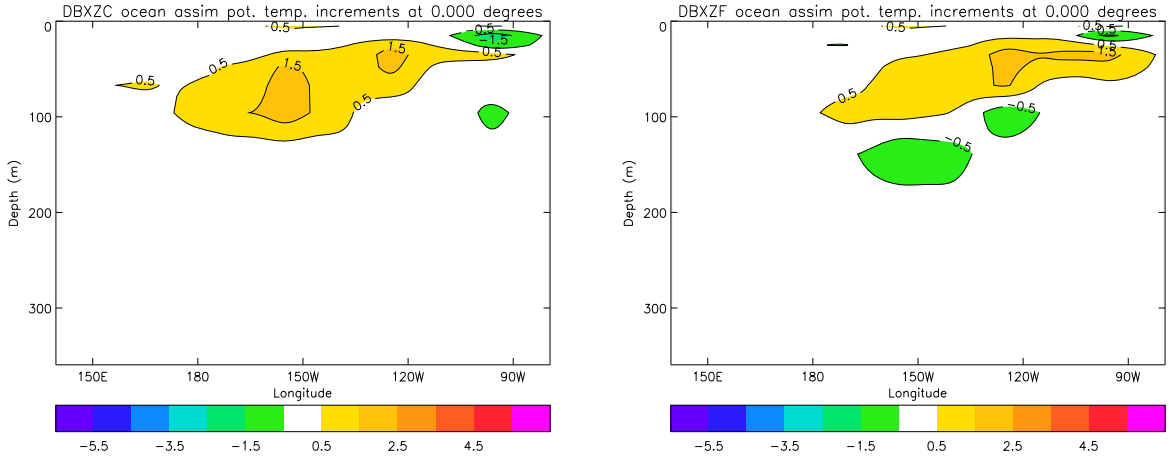

Figure 7.4.7: Annual mean (year 2) potential temperature increments (^{0}C per month) cross section along the equator: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

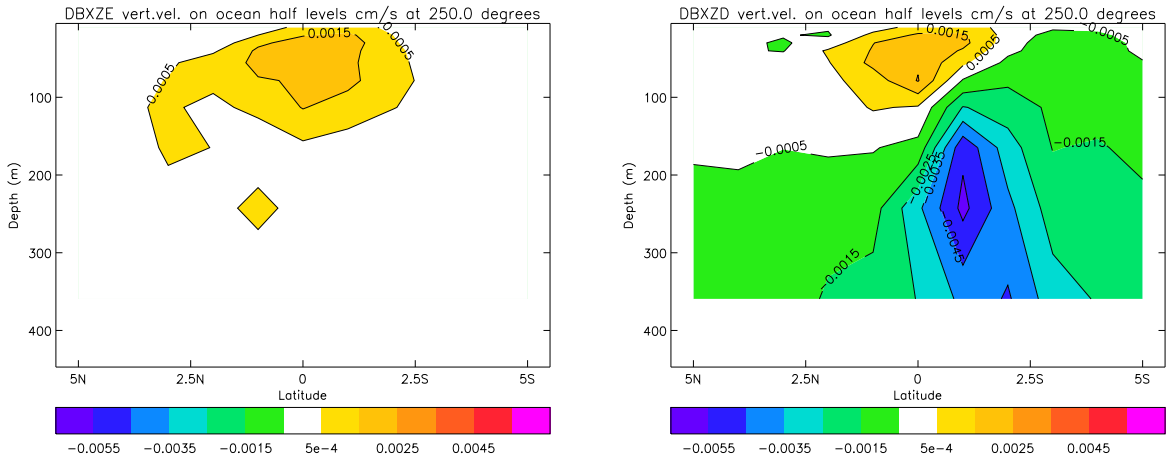

Figure 7.4.8: Annual mean (year 2) vertical velocities (cm/s) cross section across the equator at 110^0W: (a) no assimilation (b) normal assimilation

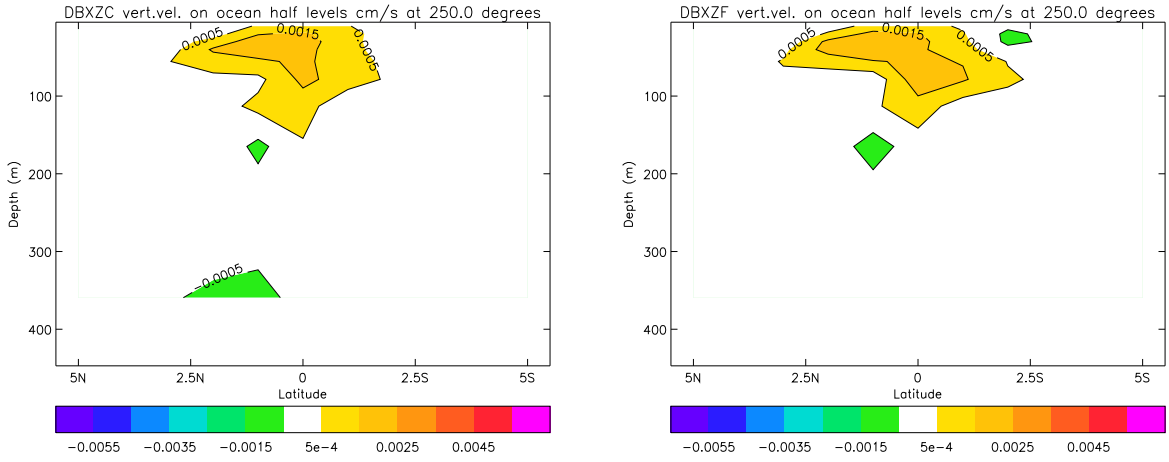

Figure 7.4.9: Annual mean (year 2) vertical velocities (cm/s) cross section across the equator at 110^0W: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

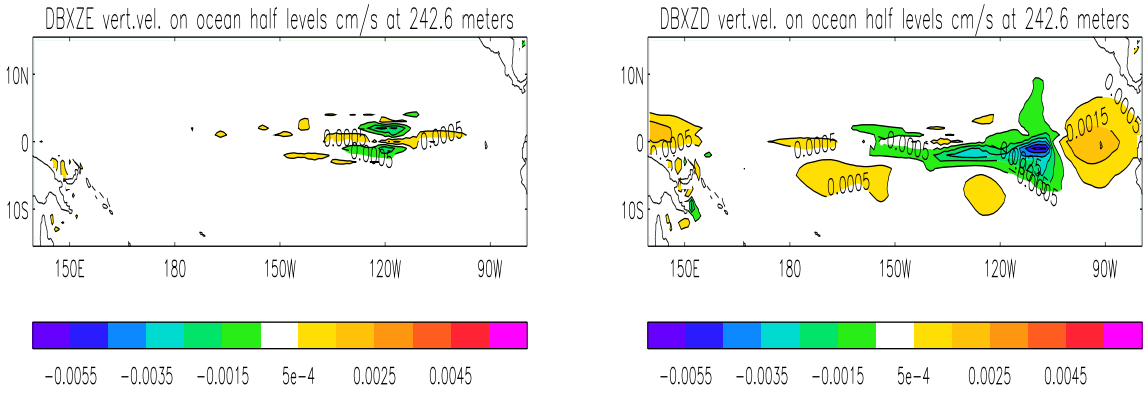

Figure 7.4.10: Annual mean (year 2) vertical velocities (cm/s) at about 250m depth for region $\pm 15^0$N/S, 140^0E-80^0W: (a) no assimilation (b) normal assimilation

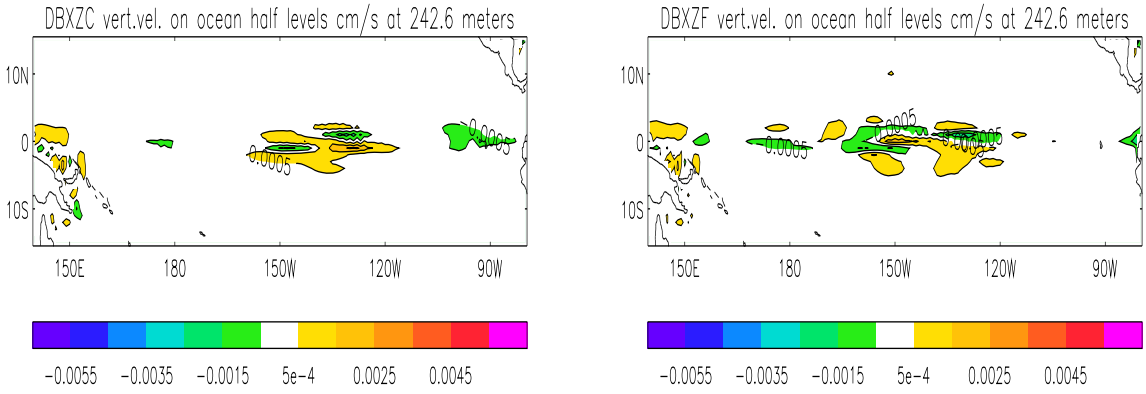

Figure 7.4.11: Annual mean (year 2) vertical velocities (cm/s) at about 250m depth for region $\pm 15^0$N/S, 140^0E-80^0W: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

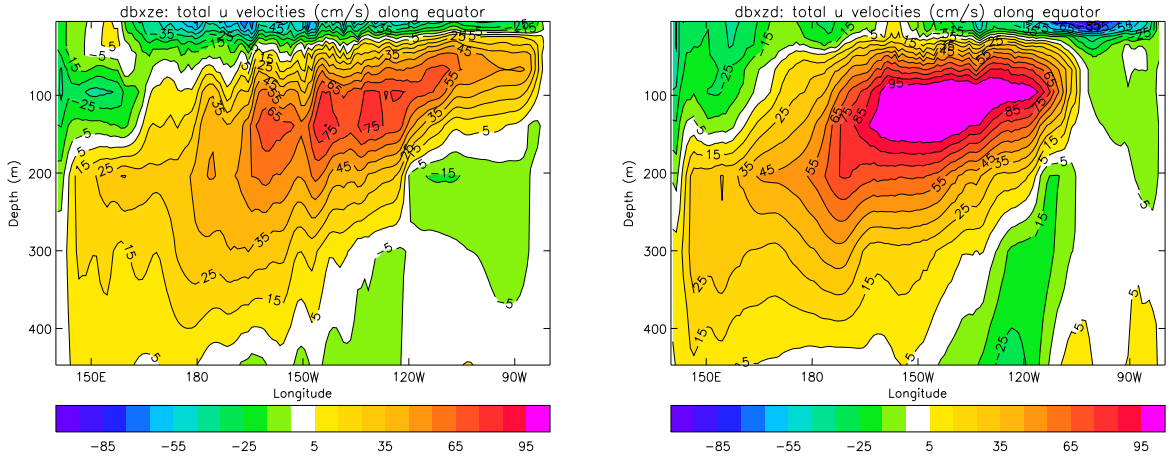

Figure 7.4.12: Instantaneous (1st May 1996) u velocities (cm/s) cross section along the equator: (a) no assimilation (b) normal assimilation

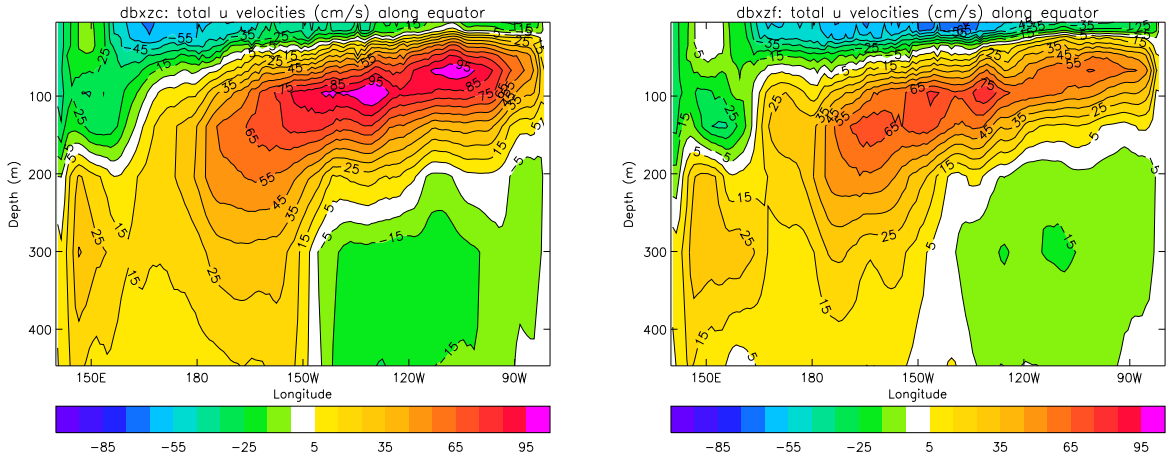

Figure 7.4.13: Instantaneous (1st May 1996) u velocities (cm/s) cross section along the equator: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

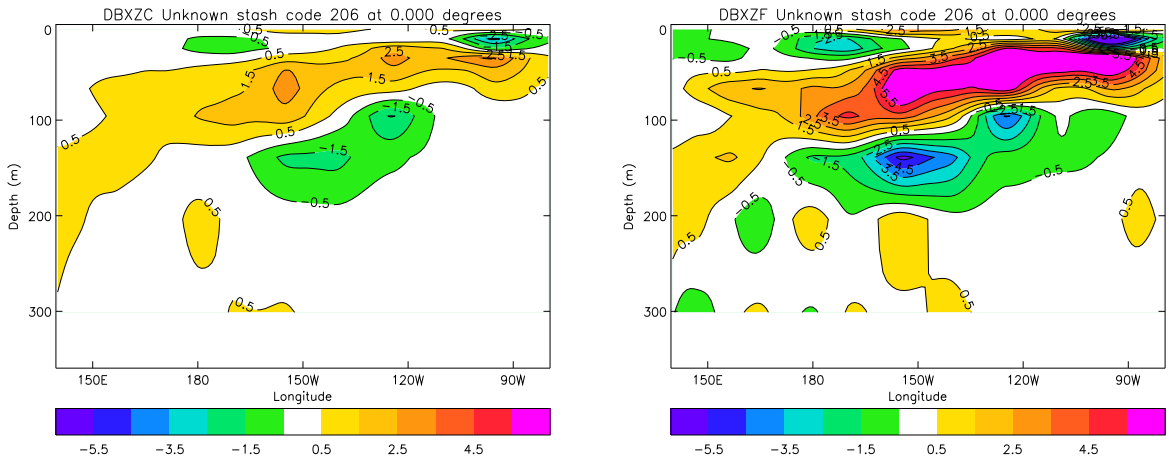

Figure 7.4.14: Annual mean (year 2) compensating temperature field (^{0}C) cross section along the equator: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

7.5 Conclusions

In this chapter, we first of all described the ocean component of the FOAM system which is used operationally at the Met. Office. As can be seen from this description, the model contains many parameters to which the resulting flows are sensitive. Perhaps the most important parameterisations in the equatorial Pacific are those for the mixing of the wind stresses in the vertical. Errors in these parameterisations can cause the model to have significant systematic errors. There are also many other possible sources of systematic model error. One important aspect of the model is the numerical scheme used for advection. In the experiments of this chapter, this scheme is a simple centred-difference scheme which is known to produce artificial grid point noise. Some experiments were also carried out using the more

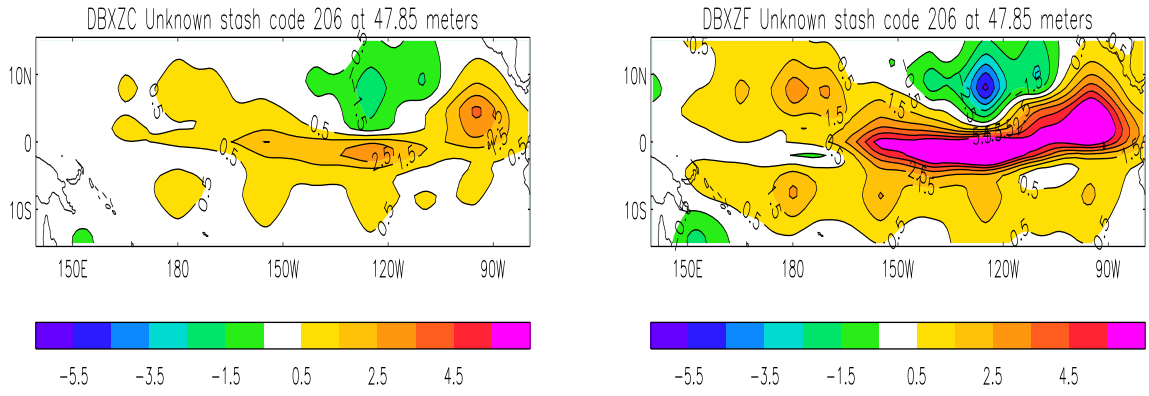

Figure 7.4.15: Annual mean (year 2) compensating temperature field (^{0}C) at about 50m depth for region $\pm 15^0$N/S, 140^0E-80^0W: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

sophisticated QUICK advection scheme but are not shown here because the impacts from the pressure correction method were qualitatively similar.

The results of experiments with and without the normal AC data assimilation scheme were presented. These illustrated the problems of assimilating data into a numerical model of the equatorial Pacific in the presence of systematic errors in the wind forcing. They showed that the model is driven out of balance, which leads to very large vertical velocities. This is due to the data assimilation continually putting in large temperature increments, attempting to bring the model towards the observations, whilst the wind stresses force the model to a different equilibrium. The reason for performing the data assimilation is to produce the best estimate of the initial conditions for a forecast. A forecast made from the initial conditions at the end of these

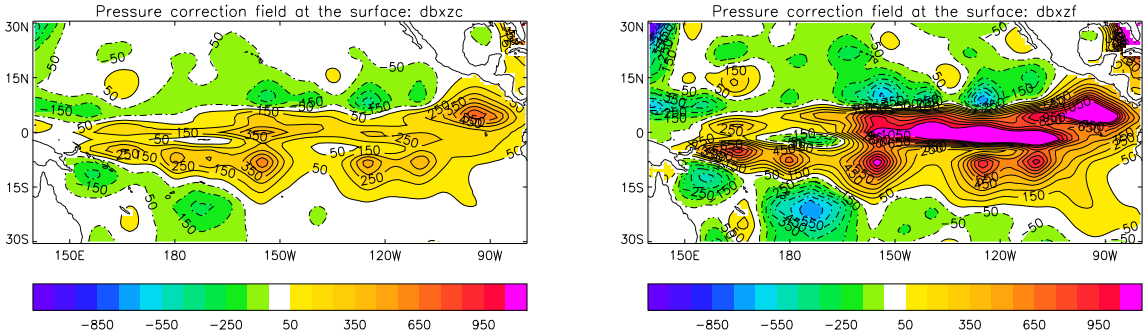

Figure 7.4.16: Annual mean (year 2) pressure correction field (dPa) at the surface for region $\pm 15^0$N/S, 140^0E-80^0W: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

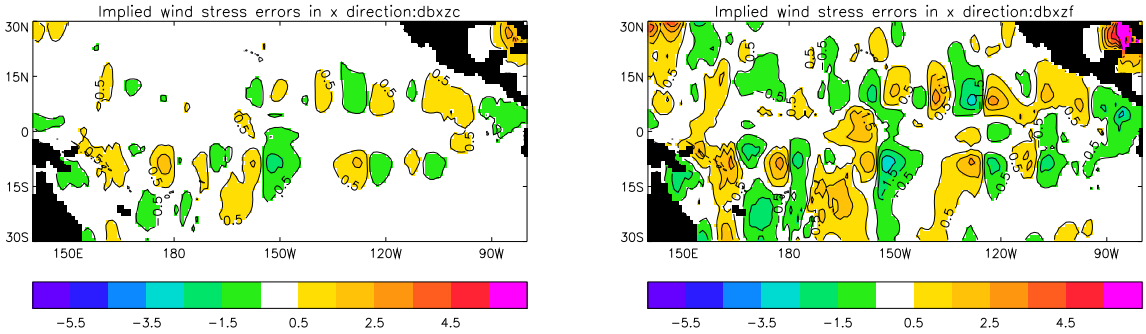

Figure 7.4.17: Implied annual mean (year 2) systematic errors in τ^x at the surface for region $\pm 15^0$N/S, 140^0E-80^0W: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

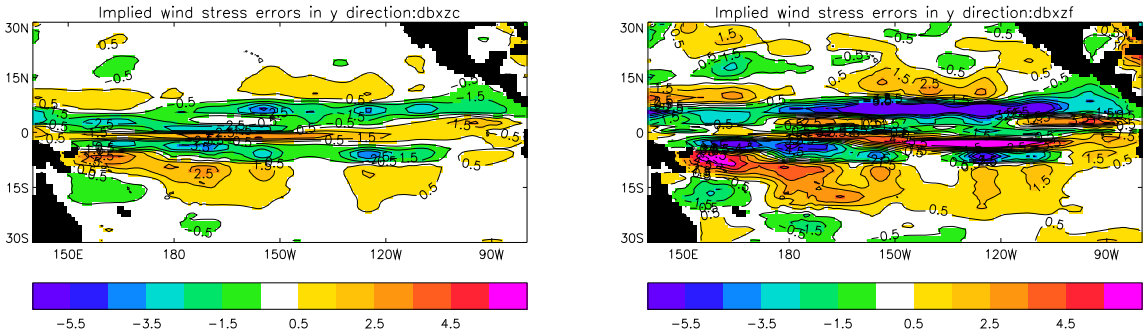

Figure 7.4.18: Implied annual mean (year 2) systematic errors in τ^y at the surface for region $\pm 15^0$N/S, 140^0E-80^0W: (a) $\gamma = -0.1$ (b) $\gamma = -0.3$

experiments would not produce very accurate results.

The pressure correction method was then applied to produce a more balanced and accurate estimate of the state of the ocean. The results of the experiments with this method showed that by using the temperature observations, the pressure correction method can alleviate the problems caused by the systematic errors in the wind forcing. The vertical velocities and temperature increments were shown to be smaller than with the normal data assimilation, which implies that the model is now more balanced. Also, statistics were presented which show the model's temperatures to be closer to the observations than with the normal data assimilation scheme. However, the observations also contain errors and so these statistics do not necessarily imply that the model is closer to the true state of the ocean. It would therefore

be useful to verify the results against some independent observations.

As well as producing a more balanced ocean state, the pressure correction method gives us some insight into the origins of the systematic model error through the structure of the compensating temperature field. These fields contain a significant depth dependent structure. It therefore seems likely that both the the wind stresses themselves and the parameterisation of mixing in the vertical contain systematic errors. Another important point when implementing this method is the question of what to do with the compensating temperature field during a forecast. Perhaps the best thing would be to ramp the field down to zero over some time period. However, more work is needed on these last two points.

An important source of observational information, the satellite altimeter, has not been discussed so far in this chapter. This instrument provides a wide and frequent coverage of information on the height of the sea surface, [1]. A method for including this data in the assimilation scheme to infer subsurface information is given in [17]. It is shown in [57] that altimeter data is very useful but that subsurface observations are still required to produce good analyses. Using the pressure correction method whilst assimilating both temperature data and altimeter data is therefore another problem which needs to be addressed.

Chapter 8

Conclusions

The foundations of physical oceanography which are relevant to the work in this thesis are reviewed in Chapter 2. We first described the standard approximations used when formulating the equations for an ocean general circulation model. We also indicated the main area of study for this thesis, that is, the tropical Pacific, and why we are interested in this area. The problems associated with estimating and forecasting the ocean state in this region were also discussed, including those associated with wind forcing. This discussion motivated the main theme of the thesis, that is to obtain a better understanding of the flow in the equatorial Pacific by using observations.

The general data assimilation problem is formulated in a discrete framework in Chapter 3, where we aim to use observations, together with a numerical model, to produce an estimate of the state of a system. Some theoretical results are given from Control Theory which include some results for the convergence of observers. These observers are designed to achieve the same results as data assimilation. We therefore related the theory of observers to

a general sequential data assimilation method in the case of a linear model and gave conditions for stability and convergence. We also showed how the theory can be extended to non-linear problems under certain circumstances.

Chapter 4 contains an overview of the more popular data assimilation methods. Initially, the variational cost function used in some of the data assimilation methods is formulated in a statistical framework. Here, a weighted sum of the squares of the differences between the state and a background state and the differences between the state and the observations is minimised. A description is then given of the main sequential data assimilation methods, including the optimal method over a period of time, the Kalman filter. It is shown how the Kalman filter can be extended to non-linear problems and how it is possible to simplify the Kalman filter using various methods. A brief overview of some of the 4D data assimilation methods is also given for completeness.

Some results were presented which give examples of how four sequential data assimilation methods perform when implemented on two ordinary differential equation models, a linear damped oscillating system and the non-linear, chaotic Lorenz equations. These experiments show that the Kalman filter is the best method in the linear model. In the nonlinear model, the extended Kalman filter produces the most accurate results when the observations are closely spaced in time. When the observations are taken less frequently, the extended Kalman filter produces similar results to those of Optimal Interpolation.

The theory given in Chapters 3 and 4 is based on the assumption that the forecast model is perfect. This is rarely the case in practice, however,

as Chapter 2 indicates in the oceanographical context. We therefore showed in Chapter 5 how the method of state augmentation can be used to account for systematic model errors in the linear case. The theory given in the two preceding chapters is then extended to give conditions for the stability of the data assimilation process and the convergence of the analysed state and systematic model error to their true values. To implement this method, a model is needed for the evolution of the systematic model error. This model is an important part of the technique and some possible choices are given, the final decision depending on the particular problem to which the method is to be applied. We then showed how the method could be applied to the sequential data assimilation methods and gave expressions for the calculation of the weighting matrices.

Some experimental results are presented in this chapter which aim to account for systematic model errors. A constant bias is included in the models for the damped oscillating system and the Lorenz equations, and it is shown that this error can be accounted for by a constant bias correction. Also, the model parameters are altered in the damped oscillating system and it is shown that these can be accounted for by using the state model equations to evolve the systematic model error variables. We also showed that, in the Heat equation, the constant bias correction method gives an accurate correction for incorrectly specified boundary conditions.

In Chapter 6, we investigated a specific type of systematic model error in the two-dimensional linear Shallow Water equations. It is shown in a continuous framework that systematic errors in the wind forcing of the equations are not taken into account when using normal data assimilation. We also ex-

amined the use of systematic model error correction in this case, which leads to a new method that can account for certain types of systematic errors in the wind forcing whilst being easy to implement and computationally inexpensive: the pressure correction method. The convergence of this method was analysed in a continuous framework, and it was shown that the analysed solution should converge to the true solution for the types of waves expected in the equatorial Pacific.

A numerical model of the linear Shallow Water equations was formulated in the domain of the equatorial Pacific. This was used to verify the theoretical results where there are systematic errors in the wind forcing. These experiments show that using Optimal Interpolation does improve the accuracy of the pressure variable during the assimilation, but the velocities are worse than the solution without any data assimilation. The results when using the bias correction and pressure correction methods were then compared. The main conclusion from these results is that the lack of information about the gain matrix in the analysis of the bias, and the small values of α, mean that the bias correction method does not perform well. The form of the pressure correction in the momentum equations implies that the pressure correction method produces better results than the normal data assimilation scheme. These experiments also give an indication of which values of γ should be used in the pressure correction method.

After producing a method which should account for systematic errors in the wind forcing of a model of the equatorial Pacific, we implemented and tested the method in an operational model of the ocean, as described in Chapter 7. The model used is FOAM, the operational Met. Office ocean model.

We first described the FOAM ocean model component and Analysis Correction data assimilation scheme. This description showed the importance of the many parameterisations that are included in the model, especially those for the mixing of the wind stress in the vertical. Results were shown of some experiments with and without data assimilation. These illustrate the problems caused by incorrect wind stresses. The vertical velocities and temperature increments are very large when the data assimilation is performed. This is explained to be due to the systematic errors in the wind stresses and their parameterisation. We next described how the pressure correction method is implemented when assimilating temperature observations. The results of the experiments using the pressure correction method for two years of integration were presented. These show promising results in that the vertical velocities and temperature increments in the model have both been reduced. Also, the model temperature is shown to be closer to the temperature observations.

Open questions and further work

The experimental testing of the pressure correction method given in Chapter 7 could be extended. The first thing to do would be to obtain some independent observations of temperature with which to verify our results more rigorously. Another way of testing the results would be to observe how the forecasts of the various experiments evolve after the assimilation period. This would lead to another area of future work which is to explore how best to predict the compensating temperature field in the forecast. Also, the experiments could be performed again but in an El Niño or La Niña year to see how the changes in the surface wind stresses affect the impact of the pressure

correction method. More knowledge could be obtained about the main types of systematic model errors by looking in more detail at the compensating temperature field. The structure and magnitude of this field should give us some insight into how to alter the parameterisation of the mixing of the wind stresses in the vertical.

Another open question is how well the pressure correction method would perform in a different region of the world's oceans. For instance, it is difficult to obtain an accurate estimate of the place in which the Gulf stream of the North Atlantic separates from the coast of North America. An investigation into the types of error responsible for this could lead to a form of bias correction which will produce a more accurate estimate of this separation point.

Bibliography

[1] Anderson, D.L.T., Sheinbaum, J., Haines, K. Data assimilation in ocean models. *Rep. Prog. Phys.*, **59**, 1996, 1209-1266.

[2] Barnett, S., Cameron, R.G. *Introduction to mathematical control theory.* OUP, 1975.

[3] Bell, M.J. FOAM data assimilation scheme. Part II: Description of methods. *FOAM Documentation Paper.* The Met. Office, London Road, Bracknell, 1991, FOAM/4/1/2.

[4] Bell, M. J., Gregorious, L. S. An assimilation of historical observations of temperature profiles into an ocean model. *Int. WOCE Newsletter*, **30**, 36-39, 1998.

[5] Bell, M. J., Forbes, R.M, Hines, A. Assessment of the FOAM global data assimilation system for real-time operational ocean forecasting. *J. Mar. Syst.*, **24** 2000, 249-275.

[6] Bell, W.W. *Special functions for scientists and engineers.* D. Van Nostrand Company ltd., 1968.

[7] Bergthorsson, P., Doos, B.R. Numerical weather map analysis. *Tellus*, **7**, 1955, 329-340.

[8] Bratseth, A.M. Statistical interpolation by means of successive corrections. *Tellus*, **38A**, 1986, 439-447.

[9] Bryan, K, Cox, M.D. A numerical investigation of the oceanic general circulation. *Tellus*, **19**, 1967, 54-80.

[10] Bryan, K. A numerical study of the circulation of the world ocean. *J. Comp. Phys.*, **4** 1969, 347-376.

[11] Burger, G., Cane, M.A. Interactive Kalman filtering. *J. of Geoph. Res.*, **99**, 1994, 8015-8031.

[12] Carton, J.A., Giese, B.S., Cao, X., Miller, L. Impact of altimeter, thermistor and expendable bathythermograph data on retrospective analyses of the tropical Pacific ocean. *J. Geophys. Res.*, **101**, 1996, 14,147-14,159.

[13] Chassignet, E.P., Verron, J. (Eds.) *Ocean modelling and parameterization.* Kluwer Academic Publishers, 1998.

[14] Cohn,S.E., Dee, D.P. Observability of discretized partial differential equations. *SIAM J. Numer. Anal.*, **25**, 1988, 586-617.

[15] Cohn, S.E., Sivakumaran, N.S., Todling, R. A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 1994, 2838-2867.

[16] Cohn, S.E., Todling, R. Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Jpn.*, **74**, 1996, 63-75.

[17] Cooper, M., Haines, K. Altimetric assimilation with water property conservation. *J. Geophys. Res.*, **101**, 1996, 1059-1077.

[18] Courtier, P. Dual formulation of four-dimensional variational assimilation. *Quart. J. R. Met. Soc.*, **123**, 1997, 2449-2461.

[19] Cox, M.D. A primitive equation, 3-dimensional model of the ocean. *GFDL Ocean Group Tech. Rep. No. 1*, Geophysical Fluid Dynamics Laboratory, 1984, 143.

[20] Cressman, G. An optimal objective analysis system. *Mon. Weather Rev.*, **87**, 1959, 367-374.

[21] Daley, R. *Atmospheric data analysis.* Cambridge University Press., 1991.

[22] Daley, R. The effect of serially correlated observation and model error on atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 1992, 164-177.

[23] Davies, H.C., Turner, R.E. Updating prediction models by dynamical relaxation: an examination of the technique. *Quart. J. R. M. Soc.*, **103**, 1977, 225-245.

[24] Dee, D.P., Da Silva, A.M. Data assimilation in the presence of forecast bias. *Quart. J. R. Met. Soc.*, **124**, 1998, 269-295.

[25] Derber, J.C. A variational continuous data assimilation technique. *Mon. Wea. Rev.*, **117**, 1989, 2437-2446.

[26] Evensen, G., Van Leeuwen, P.J. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasi-geostrophic model. *Mon. Wea. Rev.*, **124**, 1996, 85-96.

[27] Evensen, G. Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.*, **125**, 1997, 1342-1354.

[28] Février, S., Frankignoul, C., Sirven, J., Davey, M., Delecluse, P., Ineson, S., Macías, J., Sennéchael, N., Stephenson, D.B. A multivariate intercomparison between three oceanic GCMs using observed current and thermocline depth anomalies in the tropical Pacific during 1982-1992. *J. Mar. Syst.*, **24**, 2000, 249-275.

[29] Friedland, B. Treatment of bias in recursive filtering. *IEEE Trans. Auto. Contr.*, **AC-14**, 1969, 359-367.

[30] Gauthier, P. Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model. *Tellus*, **44A**, 1992, 2-17.

[31] Gelb, A. (ed.) *Applied optimal estimation.* MIT Press, 1974.

[32] Ghil, M. Meteorological data assimilation for oceanographers. Part I: description and theoretical framework. *Dyn. Atmos. Oceans*, **13**, 1989, 171-218.

[33] Ghil, M., Malanotte-Rizzoli, P. Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 1991, 141-266.

[34] Gill, A.E. Some simple solutions for heat-induced tropical circulation. *Quart. J. R. Met. Soc.*, **106**, 1980, 447-462.

[35] Gill, A.E. *Atmosphere-Ocean Dynamics.* Academic Press, 1982.

[36] Goldenburg, S.B., O'Brien, J.J. Time and space variability of tropical Pacific wind stress. *Mon. Wea. Rev.*, **109**, 1981, 1190-1207.

[37] Griffith, A.K., Nichols, N.K. Accounting for model error in data assimilation using adjoint methods. *Compututational Differentiation: Techniques, Applications and Tools*, (Eds. M. Berz, C. Bischof, G.Corliss and A. Greiwank), SIAM, Philadelphia, 1996, 195-204.

[38] Griffith, A.K. Data assimilation in numerical weather prediction using control theory. PhD Thesis, The University of Reading, Department of Mathematics, 1997.

[39] Griffith, A.K., Nichols, N.K. Adjoint methods for treating model error in data assimilation. *Numerical Methods for Fluid Dynamics*, (Ed. M.J. Baines), ICFD, Oxford University Computing Laboratory, 1998, 335-344.

[40] Hannachi, A., Haines, K. Convergence of data assimilation by periodic updating of simple Hamiltonian and dissipative systems. *Tellus*, **50A**, 1998, 58-75.

[41] Hayes, S.P., Mangum, L.J., Picaut, J., Sumi, A., Takeuchi, K. TOGA-TAO: a moored buoy array for real-time measurements in the tropical Pacific Ocean. *Bull. Am. Meteor. Soc.*, **72**, 1991, 339-347.

[42] Hellerman, S., Rosenstein, M. Normal monthly wind stress over the world ocean with error estimates. *J. Phys. Oceanogr.*, **13**, 1983, 1093-1104.

[43] Holland, W.R. The role of mesoscale eddies in the general circulation of the ocean, Numerical experiments using a wind-driven quasi-geostrophic model. *J. Phys. Oceanogr.*, **8**, 1978, 363-392.

[44] Ignagni, M.B. Separate-bias Kalman estimator with bias state noise. *IEEE Trans. Auto. Contr.*, **AC-35**, 1990, 338-341.

[45] Jazwinski, A.H. *Stochastic processes and filtering theory.* Academic Press, 1970.

[46] Ji, M., Leetmaa, A., Derber, J. An ocean analysis system for seasonal to interannual climate studies. *Mon. Wea. Rev.*, **123**, 1995, 460-481.

[47] Ji, M., Smith, T.M. Ocean model response to temperature data assimilation and varying surface wind stress: intercomparison and implications for climate forecast. *Mon. Wea. Rev.*, **123**, 1995, 1811-1821.

[48] Johnson, G.C., Moore, D.W. The Pacific subsurface countercurrents and an inertial model. *J. Phys. Ocean.*, **27**, 1997, 2448-2459.

[49] Kalman, R.E. A new approach to linear filtering and prediction problems. *Trans. of the ASME ser. D, J. Basic Eng.*, **82**, 1960, 35-44.

[50] Kalman, R.E., Bucy, R.S. New results in linear filtering and prediction theory. *Trans. of the ASME ser. D, J. Basic Eng.*, **83**, 1961, 95-108.

[51] Kantha, L.H., Clayson, C.A. *Numerical models of oceans and oceanic processes.* Academic Press, 2000.

[52] Kautsky, J., Nichols, N.K., Van Dooren, P. Robust pole assignment in linear feedback. *Int. J. Control*, **41**, 1985, 1129-1155.

[53] Lebedev, N.N. *Special functions and their applications.* Prentice-Hall, inc., 1965.

[54] Le Dimet, F-X, Talagrand, O. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, **38A**, 1986, 97-110.

[55] Levitus, S., Burgett, R., Boyer, T. World ocean atlas 1994. Volume 3: Salinity and Volume 4: Temperature. *NOAH Atlas NESDIS 3 & 4*.

[56] Lewis, J.M., Derber, J.C. The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 1985, 309-322.

[57] Long, R.B., Thacker, W.C. Data assimilation into a numerical equatorial ocean model. II: assimilation experiments. *Dyn. Atmos. Oceans*, **13**, 1989, 413-439.

[58] Lorenc, A.C. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, **109**, 1981, 701-721.

[59] Lorenc, A.C. Analysis methods for numerical weather prediction. *Quart. J. R. Met. Soc.*, **112**, 1986, 1177-1194.

[60] Lorenc, A.C. Optimal nonlinear objective analysis. *Quart. J. R. Met. Soc.*, **114**, 1988, 205-240.

[61] Lorenc, A.C., Bell, R.S., MacPherson, B. The met. office analysis correction data assimilation scheme. *Quart. J. R. Met. Soc.*, **117**, 1991, 59-89.

[62] Lorenz, E.N. Deterministic non-periodic flow. *J. Atmos. Sci.*, **20**, 1963, 455-475.

[63] Martin, M.J., Nichols, N.K., Bell, M.J. Treatment of systematic errors in sequential data assimilation. *The Met. Office, Ocean Applications Tech. Note no. 21*, 1999.

[64] Mendel, J.M. Extension of Friedland's bias filtering technique to a class of nonlinear systems. *IEEE Trans. Auto. Contr.*, **AC21**, 1976, 296-298.

[65] Miller, R.N., Ghil, M., Gauthiez, F. Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sc.*, **51**, 1994, 1037-1056.

[66] Miller, R.N., Busalacchi, A.J., Hackert, E.C. Sea surface topography fields of the Pacific from data assimilation. *J. Geophys. Res.*, **100**, 1995, 13,389-13,425.

[67] Morton, K.W., Mayers, D.F. *Numerical solution of partial differential equations*. Cambridge University Press, 1994.

[68] Ogata, K. *Discrete-time control systems.* Prentice Hall, 1987.

[69] Provost, C., Salmon, R. A variational method for inverting hydrographic data. *J. Mar. Res.*, **44**, 1986, 1-34.

[70] Rutherford, I.D. Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sc.*, **29**, 1972, 809-815.

[71] Sasaki, Y. An objective analysis based on the variational method. *J. Meteorol. Soc. Jpn.*, **36**, 1958, 77-88.

[72] Sasaki, Y. Some basic formulisms on numerical variational analysis. *Mon. Wea. Rev.*, **98**, 1970, 875-883.

[73] Schiff, L.I. *Quantum mechanics.* McGraw-Hill Book Company, 1968.

[74] Silverman, L.M., Meadows, H.E. Controllability and observability in time-variable linear systems. *SIAM J. Control*, **5**, 1967, 64-73.

[75] Silverman, L.M., Anderson, B.D.O. Controllability, observability and stability of linear systems. *SIAM J. Control*, **6**, 1968, 121-130.

[76] Sparrow, C.*The Lorenz equations: bifurcations, chaos, and strange attractors.* Springer-Verlag, 1982.

[77] Stockdale, T.N. Coupled ocean-atmosphere forecasts in the presence of climate drift. *Mon. Wea. Rev.*, **125**, 1997, 809-818.

[78] Tacker, E.C., Lee, C.C. Linear filtering in the presence of time-varying bias. *IEEE Trans. Auto. Contr.*, **AC-17**, 1972, 828-829.

[79] Talagrand, O. On the mathematics of data assimilation. *Tellus*, **33**, 1981, 321-339.

[80] Todling, R., Cohn, S.E. Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.*, **122**, 1994, 2530-2557.

[81] Veronis, G. An approximate theoretical analysis of the equatorial undercurrent. *Deep-Sea Res.*, **6**, 1960, 318-327.

[82] Weisburg, R.H., Qiao, L. Equatorial upwelling in the central Pacific estimated from moored velocity profilers. *J. Phys. Ocean.*, **30**, 2000, 105-123.

[83] Weiss, L. Controllability, realization and stability of discrete-time systems. *SIAM J. Control*, **10**, 1972, 230-251.

[84] Yu, L., O'Brien, J.J. Variational assimilation of the wind stress drag coefficient and the eddy viscosity profile. *J. Phys. Oceanogr.*, **21**, 1991, 709-719.

[85] Zhou, D.H., Sun, Y.X., Xi, Y.G., Zhang, Z.J. Extension of Friedland's separate-bias estimation to randomly time-varying bias of nonlinear systems. *IEEE Trans. Auto. Contr.*, **AC-38**, 1993, 1270-1273.

Appendix A

Derivation of the forced shallow water equation

In this section we describe the derivation of equation (6.3.5) from equations (6.3.1 -6.3.4). First of all, differentiate equation (6.3.2) with respect to x and equation (6.3.1) with respect to y, and use equation (6.3.3) to obtain

$$\frac{\partial}{\partial t}(\frac{\partial \hat{v}}{\partial x} - \frac{\partial \hat{u}}{\partial y}) + \beta \hat{v} - \frac{f}{H}(\frac{\partial}{\partial t} + \epsilon)\hat{p} = \frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y}. \tag{A.0.1}$$

To eliminate $\hat{u}$, take the x derivative of equation (A.0.1) and use equation (6.3.3) again, giving

$$\frac{\partial}{\partial t}(\nabla^2 \hat{v}) + \beta \frac{\partial \hat{v}}{\partial x} + \frac{1}{H}(\frac{\partial}{\partial t} + \epsilon)(\frac{\partial^2}{\partial y \partial t} - f\frac{\partial}{\partial x})\hat{p} = \frac{\partial}{\partial x}(\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y}), \tag{A.0.2}$$

where $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. We require another equation in $\hat{v}$ and $\hat{p}$ and obtain this by taking f multiplied by the first derivative with respect to time of equation (6.3.1), and subtracting the second derivative with respect to time

of equation (6.3.2). This gives us

$$\frac{\partial}{\partial t}(\frac{\partial^2 \hat{v}}{\partial t^2} + f^2 \hat{v}) + g(\frac{\partial}{\partial t} + \gamma)(\frac{\partial^2}{\partial y \partial t} - f\frac{\partial}{\partial x})\hat{p} = \frac{\partial}{\partial t}(\frac{\partial \tau^y}{\partial t} - f\tau^x). \qquad \text{(A.0.3)}$$

To eliminate $\hat{p}$ we add the derivative with respect to time of equation (A.0.2) to γ multiplied by equation (A.0.2) and multiply this by gH. We then subtract from this the derivative with respect to time of equation (A.0.3) and ϵ multiplied by equation (A.0.3) to obtain

$$\begin{aligned} c^2(\frac{\partial}{\partial t} \quad + \quad \gamma)(\frac{\partial}{\partial t}(\nabla^2 \hat{v}) + \beta\frac{\partial \hat{v}}{\partial x}) - \frac{\partial}{\partial t}(\frac{\partial}{\partial t} + \epsilon)(\frac{\partial^2 \hat{v}}{\partial t^2} + f^2 \hat{v}) \\ = c^2(\frac{\partial}{\partial t} \quad + \quad \gamma)\frac{\partial}{\partial x}(\frac{\partial \tau^y}{\partial x} - \frac{\partial \tau^x}{\partial y}) - \frac{\partial}{\partial t}(\frac{\partial}{\partial t} + \epsilon)(\frac{\partial \tau^y}{\partial t} - f\tau^x), \end{aligned} \qquad \text{(A.0.4)}$$

where $c^2 = gH$.

Appendix B

Properties of two polynomials

B.1 Hermite polynomials

In this section, we state some of the properties of the Hermite polynomials used in Sections 2.4.1 and 6.3. Much of this section is obtained from [53], [6] and [73].

We want to find the polynomials which satisfy the equation

$$\frac{d^2J(\zeta)}{d\zeta^2} - 2\zeta\frac{dJ(\zeta)}{d\zeta} + 2nJ(\zeta) = 0, \tag{B.1.1}$$

and are of the form

$$J(\zeta) = \sum_{r=0}^{\infty} c_r^- \zeta^r. \tag{B.1.2}$$

We write the recursion relation derived in Section 6.3 as

$$c_r^- = -\frac{(r+1)(r+2)}{2(n-r)}c_{r+2}^-. \tag{B.1.3}$$

Substituting (B.1.3) into (B.1.2) gives the following expression:

$$J(\zeta) \;=\; c_n^-(\zeta^n - \frac{n(n-1)}{2.2}\zeta^{n-2} + \ldots$$

$$+(-1)^n \frac{n(n-1)(n-2)\ldots(n-2r+1)}{2^r.2.4.6\ldots 2r}\zeta^{n-2r}+\ldots),$$

$$= \quad c_n^- \sum_{r=0}^{[\frac{1}{2}n]} \frac{(-1)^r}{2^{2r}} \frac{n!}{r!(n-2r)!}\zeta^{n-2r}, \tag{B.1.4}$$

where $[\frac{1}{2}n]$ denotes the largest integer $\leq 1/2n$. The standard coefficient is chosen to be $c_n^- = 2^n$ giving,

$$H_n(\zeta) = \sum_{r=0}^{[\frac{1}{2}n]} (-1)^n \frac{n!}{r!(n-2r)!}(2\zeta)^{n-2r}. \tag{B.1.5}$$

Theorem B.1 *It is possible to write the Hermite polynomials in terms of a generating function:*

$$w(\zeta,t) = e^{2\zeta t - t^2} = \sum_{n=0}^{\infty} H_n(\zeta)\frac{t^n}{n!}. \tag{B.1.6}$$

Proof B.1

$$\begin{aligned} e^{2t\zeta - t^2} &= e^{2t\zeta}e^{-t^2} \\ &= \sum_{r=0}^{\infty} \frac{(2t\zeta)^r}{r!} \sum_{s=0} \infty \frac{(-t^2)^s}{s!} \\ &= \sum_{r,s=0}^{\infty} (-1)^s \frac{(2\zeta)^r}{r!s!} t^{r+2s}. \end{aligned} \tag{B.1.7}$$

For the value $r+2s=n$. the coefficient of t^n is

$$(-1)^s \frac{(2\zeta)^{n-2s}}{s!(n-2s)!}. \tag{B.1.8}$$

Summing gives the coefficients of t^n as

$$\sum_{s=0}^{[\frac{1}{2}n]} (-1)^s \frac{(2\zeta)^{n-2s}}{s!(n-2s)!} = \frac{1}{n!}H_n(\zeta). \tag{B.1.9}$$

Theorem B.2 *Another way of defining the Hermite polynomials is*

$$H_n(\zeta) = (-1)^n e^{\zeta^2}\frac{d^n}{d\zeta^n}e^{-\zeta^2}. \tag{B.1.10}$$

Proof B.2 *The Taylor series expansion of the generating function is given by*

$$w(\zeta,t) = e^{2t\zeta - t^2} = \sum_{n=0}^{\infty} \left[\frac{\partial^n w}{\partial t^n}\right]_{t=0} \frac{t^n}{n!}. \tag{B.1.11}$$

We can therefore write H_n *as*

$$\begin{aligned} H_n(\zeta) &= \left[\frac{\partial^n w}{\partial t^n}\right]_{t=0} \\ &= \left[e^{\zeta^2-(\zeta-t)^2}\right]_{t=0} \\ &= e^{\zeta^2}\left[\frac{\partial^n}{\partial t^n} e^{-(\zeta-t)^2}\right]_{t=0}. \end{aligned} \tag{B.1.12}$$

We also have that

$$\frac{\partial^n}{\partial s^n} f(x-s) = (-1)^n \frac{\partial^n}{\partial x^n} f(x-s), \tag{B.1.13}$$

which gives

$$H_n(\zeta) = (-1)^n e^{\zeta^2} \frac{d^n}{d\zeta^n} e^{-\zeta^2}. \tag{B.1.14}$$

It can be proved by differentiating both sides of the generating function with respect to ζ that

$$H_n'(\zeta) = 2nH_{n-1}(\zeta), \qquad n \geq 1; \qquad H_0'(\zeta) = 0. \tag{B.1.15}$$

By differentiating both sides of the generating function with respect to t, is can be shown that

$$H_{n+1}(\zeta) = 2\zeta H_n(\zeta) - 2nH_{n-1}(\zeta), \qquad n \geq 1; \qquad H_1(\zeta) = 2\zeta H_0(\zeta). \tag{B.1.16}$$

It is possible to combine these two equations to obtain

$$H_n''(\zeta) - 2\zeta H_n' + 2nH_n = 0. \tag{B.1.17}$$

This shows that the Hermite polynomials do satisfy the original equation, B.1.1.

The first few Hermite polynomials are given as

$$\begin{aligned} H_0(\zeta) &= 1, \\ H_1(\zeta) &= 2\zeta, \\ H_2(\zeta) &= 4\zeta^2 - 2, \\ H_3(\zeta) &= 8\zeta^3 - 12\zeta, \\ H_4(\zeta) &= 16\zeta^4 - 48\zeta^2 + 12. \end{aligned}$$

B.2 I_n polynomials

The polynomials I_n introduced in Section 6.3 were said to satisfy the differential equation

$$\frac{d^2J(\zeta)}{d\zeta^2} + 2\zeta\frac{dJ(\zeta)}{d\zeta} - 2nJ(\zeta) = 0. \tag{B.2.1}$$

Again we look for polynomial solutions of the form

$$J(\zeta) = \sum_{r=0}^{\infty} c_r^+ \zeta^r. \tag{B.2.2}$$

This leads us to the recursion relation

$$c_r^+ = \frac{(r+2)(r+1)}{2(n-r)} c_{r+2}^+. \tag{B.2.3}$$

Subsituting this into equation (B.2.2) , we obtain

$$J(\zeta) = c_n^+ \sum_{r=0}^{[\frac{1}{2}n]} \frac{n!\zeta^{n-2r}}{2^{2r}r!(n-2r)!}. \tag{B.2.4}$$

If we choose $c_n^+ = 2^n$, then this becomes

$$I_n(\zeta) = \sum_{r=0}^{[\frac{1}{2}n]} \frac{n!}{r!(n-2r)!}(2\zeta)^{n-2r}. \tag{B.2.5}$$

Theorem B.3 *The polynomials $I_n(\zeta)$ can be written in terms of a generating function as*

$$w(\zeta, t) = e^{2t\zeta+t^2} = \sum_{n=0}^{\infty} I_n(\zeta)\frac{t^n}{n!}. \tag{B.2.6}$$

Proof B.3

$$e^{2t\zeta+t^2} = e^{2t\zeta}e^{t^2} = \sum_{r,s=0}^{\infty} \frac{(2\zeta)^r t^{r+2s}}{r!s!}. \tag{B.2.7}$$

For the value $r + 2s = n$, the coefficient of t^n is

$$\frac{(2\zeta)^{n-2s}}{s!(n-2s)!}. \tag{B.2.8}$$

Summing gives the coefficients in t^n as

$$\sum_{s=0}^{[\frac{1}{2}n]} \frac{1}{s!(n-2s)!}(2\zeta)^{n-2s} = \frac{1}{n!}I_n(\zeta). \tag{B.2.9}$$

Theorem B.4 *Another way of defining the polynomials $I_n(\zeta)$ is as follows:*

$$I_n(\zeta) = e^{-\zeta^2}\frac{d^n}{d\zeta^n}e^{\zeta^2}. \tag{B.2.10}$$

Proof B.4 *The Taylor series expansion of the generating function gives*

$$w(\zeta, t) = \sum_{n=0}^{\infty} \left[\frac{\partial^n w}{\partial t^n}\right]_{t=0} \frac{t^n}{n!}. \tag{B.2.11}$$

Therefore

$$\begin{aligned} I_n(\zeta) &= \left[\frac{\partial^n w}{\partial t^n}\right]_{t=0} \\ &= \left[\frac{\partial^n}{\partial t^n}e^{-\zeta^2+(\zeta+t)^2}\right]_{t=0} \\ &= e^{-\zeta^2}\left[\frac{\partial^n}{\partial t^n}e^{(\zeta+t)^2}\right]_{t=0}. \end{aligned} \tag{B.2.12}$$

We also have that

$$\frac{\partial^n}{\partial t^n} f(x+t) = \frac{\partial^n}{\partial x^n} f(x+t), \tag{B.2.13}$$

so that

$$\begin{aligned} I_n(\zeta) &= e^{-\zeta^2} \left[\frac{\partial^n}{\partial \zeta^n} e^{(\zeta+t)^2} \right]_{t=0} \\ &= e^{-\zeta^2} \frac{d^n}{d\zeta^n} e^{\zeta^2}. \end{aligned} \tag{B.2.14}$$

It can be shown by differentiating both sides of the generating function by ζ that

$$I_n'(\zeta) = 2n I_{n-1}(\zeta), \qquad n \geq 1; \qquad I_0'(\zeta) = 0. \tag{B.2.15}$$

Differentiating both sides of the generating function by t gives

$$I_{n+1}(\zeta) = 2\zeta I_n(\zeta) + 2n I_{n-1}(\zeta), \qquad n \geq 1; \qquad I_1(\zeta) = 2\zeta I_0(\zeta). \tag{B.2.16}$$

Combining these two equations we obtain

$$I_n''(\zeta) + 2\zeta I_n'(\zeta) - 2n I_n(\zeta) = 0 \tag{B.2.17}$$

This shows that the polynomials I_n do satisfy the equation (B.2.1).

The first few I_n polynomials are given as

$$\begin{aligned} I_0(\zeta) &= 1, \\ I_1(\zeta) &= 2\zeta, \\ I_2(\zeta) &= 4\zeta^2 + 2, \\ I_3(\zeta) &= 8\zeta^3 + 12\zeta, \\ I_4(\zeta) &= 16\zeta^4 + 48\zeta^2 + 12. \end{aligned}$$