THE UNIVERSITY OF READING
DEPARTMENT OF MATHEMATICS

The Application of PV-based Control Variable Transformations in Variational Data Assimilation

by

David Katz

Thesis submitted for the degree of
Doctor of Philosophy

September 2007

Declaration

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

David Katz

Abstract

Data assimilation is the process of finding the best estimate of the current state of a system. In numerical weather prediction (NWP) this system is the atmosphere and oceans. In most operational weather forecasting centres variational data assimilation is performed using a different set of variables from the actual model variables. The transformation of variables simplifies the problem by assuming that the errors in the transformed variables are uncorrelated. The validity of this hypothesis is key to the accuracy of the data assimilation. Recently a potential vorticity (PV) based set of variables has been proposed. These new variables are thought to exploit more accurately important dynamical properties of the atmosphere. Here we present new results, obtained with a simplified 1-D shallow water model, comparing the PV-based variables to the vorticity-based variables currently used at operational weather forecasting centres, including the Met Office. The validity of the fundamental assumption that the errors in the transformed variables are uncorrelated is tested in a variety of dynamical regimes. The results show that the errors in the PV-based variables are uncorrelated across all regimes tested. This is not the case, however, for the vorticity-based variables. This suggests that the PV-based control variables imply a better representation of the background errors than the current vorticity-based variables. Finally, initial data assimilation results are presented. Experiments are run in high and low Burger regimes and analysis increments compared when the vorticity and PV-based control variable transforms used in the assimilation.

Acknowledgments

Firstly, I would like to thank my supervisors Prof. Nancy Nichols and Dr. Amos Lawless for all the support they have given me, both personal and academic, to complete this thesis. I would also like to thank my Met Office supervisor Prof. Mike Cullen whose knowledge of the subject was inspiring. In addition, over the years Dr. Ross Bannister has gone out of his way to help me, something that I greatly appreciate. I could not have completed the thesis without the insight, knowledge and personal support I have been given. I was very lucky to have experts in a wide variety of fields around me, all of whom also happened to be extremely nice people!

I would also like to thank Lucy for putting up with all of this (and me) for so long. A great deal has happened since I started the doctorate and you supported me through it all, I could not have done it without you.

Finally, I would like to thank the Engineering and Physical Science Research Council (EPSRC) and the Met Office for the funding I received. The funding was in the form of a Cooperative Award in Science and Engineering (CASE) studentship with sponsorship from the Met Office. Without this funding this research would not have been possible.

I dedicate the thesis to my father and grandmother Winifred.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Data assimilation is used in numerical weather prediction (NWP) to find the initial conditions for model forecasts. By combining observational data, statistical data, knowledge of atmospheric dynamics and a previous short forecast the best estimate, or *analysis*, of the state of the atmosphere is found. Due to the chaotic nature of the governing equations errors in the initial conditions may grow rapidly in the forecast [42] and thus proper definition of initial conditions forms a vital part of NWP. With approximately 10^7 variables this is a huge problem and special methods need to be found to make the problem practical to solve.
Data assimilation is formulated statistically so ideally should work with statistically independent random variables. In operational weather prediction centres around the world the data assimilation is usually performed using a different set of variables to the model variables [16], [40], [41], [47]. These variables are the *control variables* and the choice of these is key to the data assimilation system performance. The transformation of variables simplifies the problem by assuming that errors in the new variables are uncorrelated. One way that is thought to do this accurately is by using control variables whose evolution can be considered independent. A *balanced / unbalanced* partitioning of control variables is usually made. Here an attempt is made to separate the slow large-scale balanced mode and the fast unbalanced modes as it is thought there is little or no dynamical interaction between these flows and so their errors are uncorrelated.

The use of control variables in data assimilation was first introduced in [47]. Here balance between mass and momentum is introduced by combining the balanced parts of mass and momentum fields into a single variable. In this multivariate formulation, increments to this single balanced variable produce consistent balanced increments in both the mass and momentum fields.

All operational centres use control variables that are based on the ideas of [47] and are essentially vorticity-based (one exception is the data assimilation system used for the High Resolution Limited Area Model
(HIRLAM) [25], which is mass-based, but this has similar limitations). The definitions of balance used in all these transforms does not represent the correct separation of variables in all flow regimes. Surprisingly, until now relatively little work has focused on addressing these deficiencies. Recently, however, a new set of control variables has been proposed [11] that should be valid across all regimes. The new variables use a conserved quantity, the potential vorticity (PV), to capture the balanced motion.

Initial theoretical and numerical results using the PV-based control variables are promising, though not conclusive. In [67] the PV-based transform is derived for the 2D Shallow Water equations on a sphere and numerical solutions are compared to the vorticity-based variables in specific dynamical regimes. These results are in agreement with the theory though the transforms were not implemented into a data assimilation system. In [11] the PV-based transform is tested in the European Centre for Medium Range Weather Forecasting's (ECMWF) operational data assimilation system. Some promising results were obtained. However, serious numerical issues were encountered as a result of the vertical grid in the ECMWF model. To resolve these problems aspects of the transform had to be compromised, which was not ideal.

In this study we extend the body of work on the PV-based control variable transform. We test the validity of the fundamental assumption that the control variables are uncorrelated. The statistical results we obtain are new and in agreement with those predicted by the theory. The results provide further

justification for the use of the PV-based variables over the current vorticity-based implementations. Finally, initial data assimilation experiments are run to compare each transform. Specifically we aim to address the following questions:

1. How accurate is the fundamental assumption that the control variables are uncorrelated?

   The assumption that the control variables are uncorrelated is made in all the most advanced data assimilation systems and yet its accuracy has not been fully investigated.

2. Are there any approximations to the PV-based transform that can be used? What affect does this have on the data assimilation?

   For the PV-based transform to be implemented operationally it must be efficient. We therefore derive an approximated form that makes the transform more attractive from a practical perspective. We then consider the impact this approximation has on the correlation of the control variables.

3. Finally, does a more accurate representation of balance and unbalanced dynamics in the control variable transform influence the analysis produced by the data assimilation?

   It remains unclear whether a better representation of balance in the control variable transform will actually be evident in the analysis produced by the data assimilation.

The thesis is structured as follows. **Chapter 2** introduces the dynamical aspects of the work. We use the Shallow Water Equations (SWEs) to demonstrate important dynamical properties that are vital to the developmemt of this research. We introduce the concept of balance and find solutions to the linearised equations through a normal mode analysis. Crucially, the normal modes of the linearised equations are linked to balanced and unbalanced parts of the flow. Most importantly, in this linear system, the balanced mode is characterised by the linearised potential vorticity (PV), a conserved quantity for the SWEs.

Chapter 3 introduces the concepts of data assimilation. There are many types of data assimilation methods and this work concentrates on one of the most advanced methods, four-dimensional variational data assimilation (4D VAR). This method is used at the Met Office and the European Centre for Medium Range Weather Forecasting (ECMWF) in an incremental formulation. Even with advancements in numerical methods and computing power there are still many practical difficulties in implementing the incremental 4D VAR algorithm operationally. One of these issues is specifying the necessary background / forecast error statistics. This is a problem for several reasons:

- The true state of the atmosphere is never known exactly and therefore we are unable to calculate the errors in the background guess or the forecasts.
- The amount of data required to specify all the necessary error covariance information is $10^7 \times 10^7$ and a matrix of this size is impossible to store in computer memory. We note that in this work the number of variables is greatly reduced.

To proceed further with this method we model these error statistics using control variable transformations. The control variable transforms that are implemented operationally are described and their limitations discussed. This motivates our work as the PV-based transform is designed to overcome these limitations. However, this is yet to be verified.

Chapter 4 introduces the 1D form of the Shallow Water Equations and these equations are used for the remainder of the work. We describe the discrete model and present a variety of experiments used to verify the model is coded correctly and to demonstrate its behaviour. We use the ideas introduced in Chapter 2 to interpret these results. Before the incremental 4D VAR can be implemented we require a linear approximation of the 1D SWE model and its adjoint. Both these models are developed and tested using standard methods. Once these models are verified they are implemented into the data assimilation.

Chapter 5 derives the vorticity-based and PV-based control variable transforms for the 1D SWEs. We test the numerical implementation of the transforms and derive mathematical conditions necessary for the transforms to be solved. The properties of the control variables in different dynamical regimes are then demonstrated by qualitatively comparing the fields. We obtain results that are in agreement with the theory presented in Chapter 2.

Chapter 6 contains the main results of this work. From these new results we are able to answer Questions 1 and 2, drawing conclusions regarding the effectiveness of the each transform. We test the validity of the fundamental assumption that the control variables are uncorrelated in a variety of dynamical regimes. The results show that the errors in the PV-based variables are uncorrelated across all regimes tested. This is not the case, however, for the vorticity-based variables. This suggests that the PV-based control variables imply a better representation of the background errors than the current vorticity-based variables. To address Question 2 the consequences of using an approximated form of the PV-based transform is investigated. We consider the impact of this approximation on the correlations between the control variables. Finally, we use our statistical method to derive auto-correlations for each control variable. The auto-covariances are a vital component of the data assimilation as they provide a length scale for each variable that indicates how far background information should be spread. The length scales produced are then interpreted to assess the success of each transform in capturing the balanced and unbalanced dynamics. We demonstrate that the auto-covariances are highly dependent on regime and therefore this should be accounted for in the assimilation system.

Chapter 7 considers simple single observation experiments where the observation is taken at the initial time $t = 0$. In this simple scenario the results can be analysed more easily. We first consider the theoretical influence that each control variable transform has on the analysis produced by the assimilation. This is achieved using a modified version of the best linear unbiased estimate (BLUE), which is adapted to include the control variable transform. From this analysis

much can be learnt regarding the influence of the vorticity and PV-based transforms on the data assimilation and we are able to identify fundamental differences between the transforms. To complete the specification of the data assimilation system that we develop we chose to model the auto-covariances using simple Gaussian correlation functions. Finally, simple assimilation results are presented when each control variable transform is used in our data assimilation system. The results demonstrate the theory that is presented in this work.

Chapter 8 draws conclusions from these results regarding the effectiveness of the PV-based transforms. We then make suggestions for possible further work in this area.

We now introduce the dynamical background to the project and discuss concepts of balance and dynamic regimes that are key to this research. We do this in the framework of the 2D shallow water equations (SWEs).

Chapter 2

Dynamical Background

The shallow water equations (SWEs) are often used as a test case for atmospheric research. They approximate the full governing equations used in numerical weather prediction (NWP) whilst being capable of describing important aspects of atmospheric and oceanic motions [49]. They lead to a simplified set of equations that has the vertical coordinate removed. We use the SWEs to introduce concepts key to this research and a further simplified version of the SWEs is used later in this work.

2.1 Shallow Water Equations

The properties of the governing equations used in NWP are extremely complicated and will not be examined here, but a detailed analysis is provided in [49] and [24]. Progress is often made by identifying small parameters and using these parameters to derive simpler forms of the equations, which can be analysed. The SWEs are one such system of approximate equations. The link between the governing system of equations and the SWEs is demonstrated in certain key dynamical properties that are introduced later in this chapter.

The SWEs describe a layer of shallow, rotating, inviscid and incompressible fluid, as shown in figure 2.1. The height of the surface above the reference level $z = 0$ is $h + \widetilde{H}$, where $h(x, y, t)$ is the fluid depth and $\widetilde{H}(x, y)$ is the height of the orography.

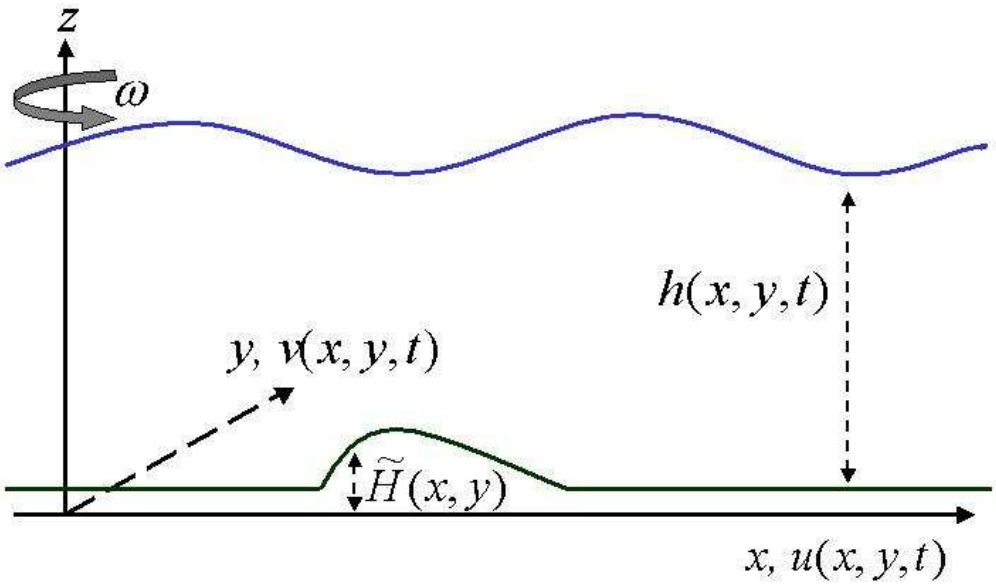

Figure 2.1: Diagram for the Shallow Water Equations.

The acceleration due to gravity, g, acts perpendicular to the reference level $z = 0$ and f is the coriolis parameter. The rotation rate ω is constant and choosen to be about the z axis and so the coriolis parameter is given by $f = 2\omega$. The velocity has components u and v in the horizontal x and y directions respectively. The SWEs are then given by the non-linear partial differential equations

$$x\text{-momentum:} \quad \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv = -g\frac{\partial(h+\widetilde{H})}{\partial x} \tag{2.1}$$

$$y\text{-momentum:} \quad \frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu = -g\frac{\partial(h+\widetilde{H})}{\partial y} \tag{2.2}$$

$$\text{Continuity:} \quad \frac{\partial h}{\partial t} + \frac{\partial(hu)}{\partial x} + \frac{\partial(hv)}{\partial y} = 0. \tag{2.3}$$

The fundamental assumption in deriving the SWEs that the fluid is shallow. This implies that the aspect ratio,

$$\frac{H}{L} << 1, \tag{2.4}$$

where H is the characteristic height scale and L the characteristic horizontal scale.

The aspect ratio is also small for large-scale atmospheric motions where typical horizontal scales are much larger than the vertical scales. We now analyse some of the dynamical properties of the SWEs.

2.1.1 The Potential Vorticity

To derive a conserved quantity for the SWEs we start by eliminating h from equations (2.1) and (2.2) by cross differentiation of (2.1) by y and (2.2) by x. Then $\partial(2.2)/\partial x - \partial(2.1)/\partial y$ gives

$$\frac{D\zeta}{Dt} = -(\zeta + f)\mathcal{D}, \tag{2.5}$$

where the vertical component of the *relative vorticity* is given by

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \tag{2.6}$$

the divergence $\mathcal{D}$ by

$$\mathcal{D} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \tag{2.7}$$

and the material derivative $\frac{D}{Dt}$ is defined by

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}.$$

We can then write equation (2.3) as

$$\frac{Dh}{Dt} + h\mathcal{D} = 0 \tag{2.8}$$

and substitute $\mathcal{D}$ in (2.5) giving

$$\frac{D\zeta}{Dt} = \frac{\zeta + f}{h}\frac{Dh}{Dt}.$$

With a constant f this can be written as

$$\frac{D}{Dt}\left(\frac{\zeta + f}{h}\right) = 0. \tag{2.9}$$

Therefore following the motion of each fluid column the quantity

$$q = \frac{\zeta + f}{h} \tag{2.10}$$

is conserved. Therefore if the depth h increases, the *absolute vorticity*, $\zeta + f$, must also increase proportionally. This conserved quantity q is the potential vorticity (PV). The conservation of PV is a powerful dynamical constraint. In NWP models that are used operationally, similar quantities can be derived which are conserved by the model dynamics. Accurately representing the initial PV and transporting it in an NWP model is vital to obtaining accurate forecasts ([33] page 55).
The PV is key to this research. In section 2.3.3 we will see that one of the normal modes of the SWEs, linearised about a resting state, is characterised by the linearised form of the PV. Now we introduce the concept of geostrophic balance and derive the geostrophic approximation. To do this we first identify several key parameters that help to characterise the behaviour of the equations.

2.1.2 Key parameters

We non dimensionalise SWEs by scaling the momentum equations (2.1) to (2.3) and writing them in terms of non dimensional variables given by

$$x = Lx^*,$$

$$y = Ly^*,$$

$$u = Uu^*,$$

$$v = Uv^*,$$

$$t = Tt^*,$$

and

$$h(x, y, t) + \widetilde{H}(x, y) = Hh^*,$$

where L, U, T and H are characteristic length, velocity, time and surface height scales chosen so that the magnitude of the non-dimensional variables are order unity. The scaled momentum equations are then

$$\frac{U}{T}\frac{\partial u^*}{\partial t} + \frac{U^2}{L}\left(u^*\frac{\partial u^*}{\partial x} + v^*\frac{\partial u^*}{\partial y}\right) - fUv^* = -\frac{gH}{L}\frac{\partial h^*}{\partial x} \tag{2.11}$$

$$\frac{U}{T}\frac{\partial v^*}{\partial t} + \frac{U^2}{L}\left(u^*\frac{\partial v^*}{\partial x} + v^*\frac{\partial v^*}{\partial y}\right) + fUu^* = -\frac{gH}{L}\frac{\partial h^*}{\partial y}. \tag{2.12}$$

We now assume that the time scale is equal to the advective time L/U, i.e.

$$T = \frac{L}{U}. \tag{2.13}$$

Then, dividing equations (2.11) and (2.12) by fU, we obtain

$$R_o\left(\frac{\partial u^*}{\partial t} + u^*\frac{\partial u^*}{\partial x} + v^*\frac{\partial u^*}{\partial y}\right) - v^* = -\frac{B_u^2}{R_o}\frac{\partial h^*}{\partial x} \tag{2.14}$$

$$R_o\left(\frac{\partial v^*}{\partial t} + u^*\frac{\partial v^*}{\partial x} + v^*\frac{\partial v^*}{\partial y}\right) + u^* = -\frac{B_u^2}{R_o}\frac{\partial h^*}{\partial y}, \tag{2.15}$$

where the Rossby number R_o is defined by

$$R_o = \frac{U}{fL} \tag{2.16}$$

and the Burger number B_u by

$$B_u = \frac{\sqrt{gH}}{fL}. \tag{2.17}$$

The parameters R_o and B_u are dimensionless. The Rossby number is a measure the significance of rotation in the flow [49] and is ratio of the inertial time scale, $\tau_1 = f^{-1}$ to the advective time scale $\tau_2 = L/U$ [12]. If we take characteristic values, $f = 0.0001s^{-1}$, $U = 10ms^{-1}$, $L = 10^6m$ relevant to large-scale, mid-latitude, motion of the Earth's atmosphere this gives $R_o = 0.1$. This type of motion is therefore characterised by small Rossby number. We note that, if the SWEs are solved on a sphere where f is variable, as we approach the equator $f \to 0$ and the Rossby number becomes infinite. Motion in the mid-latitudes and equatorial regions is therefore very different. This work is concerned mainly with mid-latitude, large scale motion defined by small R_0.

The Burger number defined in equation (2.17) is the ratio of the Rossby number, defined in equation (2.16), and the Froude number

$$F_r = \frac{U}{\sqrt{gH}}. \tag{2.18}$$

The Froude number is the ratio of velocity with gravity wave speed,

$$c_g = \sqrt{gH}. \tag{2.19}$$

In most deep atmosphere cases the Froude number is small, i.e. the advection velocity is much less than the gravity wave speed (though stationary gravity waves are sometimes generated around orographic features, in this case the characteristic height would be taken to be the orographic height).
An important length scale is

$$L_r = \frac{\sqrt{gH}}{f} \tag{2.20}$$

where L_r is the *Rossby radius of deformation* for a layer of depth H. The Rossby radius of deformation is the distance a gravity wave travels in one inertial period. We can see that the Burger number is also related to the Rossby radius. It is in fact the ratio of the horizontal scale of motion to the Rossby radius of deformation. In summary we have

$$B_u = \frac{\sqrt{gH}}{fL} = \frac{R_o}{F_r} = \frac{L_r}{L}$$

and when $L = L_r$ we have $R_o = F_r$. Therefore the Rossby radius of deformation is also the scale at which the Froude number and Rossby number are equal and rotation effects become important.
In the atmosphere the Burger number can vary greatly; for example, the Burger number approaches infinity at the equator as f approaches zero, but is close to zero when typical scales of motion are very large relative to the Rossby radius. This can happen for large horizontal flows or almost stationary gravity waves. For example, large weather systems or flows over mountains.
If it is assumed that B_u is order 1 and $R_o << 1$ then the advective terms in (2.14) and (2.15) are small in comparison to the Coriolis terms. Therefore, in order for u^* and v^* to be non-zero, the pressure gradient term must *balance* the Coriolis term. These ideas are now discussed in more detail.

2.2 Balance

Geostrophic balance is a fundamental concept in this work. Later in this chapter we will see that the linearised SWEs support two types of motion; slow Rossby

waves and fast inertial-gravity waves. These types of motion are also present in the atmosphere. It is suggested in [12] that we may think of the fast inertial-gravity waves as a form of meteorological noise. The justification for this is that the majority of synoptic and planetary scales of the atmosphere are dominated by the advective time scale, and inertial-gravity waves are normally a minor part of the flow [12]. Flows where the motion is dominated by the advective time scale, and therefore have $R_o << 1$, are said to be *balanced*.

Historically in NWP attempts to capture this balanced motion by a reduced set of equations filtering inertial-gravity waves. This reduced set of equations constitutes a balanced model. There is a great deal of research concerning balanced models in NWP and applications within the field, for example [3], [63] and hierarchies of balance conditions for the SWEs are examined in [18]. These are not discussed here though similar ideas are exploited later to define the control variable transform. We now demonstrate simple first-order balance in the SWEs. Firstly, this is done with a scale analysis and then formally through an asymptotic expansion in the Rossby number.

2.2.1 Balance in the SWEs

If we apply scaling to the governing equations of NWP that correspond to large-scale motions we may derive the geostrophic approximation. Here the momentum balance for the horizontal velocity reduces to a balance between the horizontal pressure gradient and the horizontal component of the Coriolis acceleration [49]. The accuracy of this approximation is demonstrated in the free atmosphere where observations of the wind field are very close to geostrophic in the midlatitudes. The wind coincides with the isobars, lines of constant pressure, which are perpendicular to the pressure gradient.

For the SWEs in a regime where $R_o << 1$ and $B_u \approx 1$ equations (2.14) and (2.15) approximate to a geostrophic balance relationship given by

$$v^* = \frac{B_u^2}{R_o}\frac{\partial h^*}{\partial x^*} \tag{2.21}$$

and

$$u^* = -\frac{B_u^2}{R_o}\frac{\partial h^*}{\partial y^*}, \tag{2.22}$$

for the non-dimensional variables defined in section 2.1.2. Written in dimensional form this relationship is given by

$$fv = g\frac{\partial(h+\widetilde{H})}{\partial x} \tag{2.23}$$

and

$$fu = -\frac{\partial(h+\widetilde{H})}{\partial y}. \tag{2.24}$$

This regime is relevant to large scale, mid-latitude motions in the atmosphere. Here the horizontal velocities are simply related to gradients of the depth of the fluid.

We may formally demonstrate the balance in the SWEs through an asymptotic expansion in the Rossby number. For this analysis we follow [49] and write equations (2.1) to (2.3) in terms of the departure η of the free surface from its level at rest. So the height of the free surface $h(x,y,t)+\widetilde{H}(x,y)$ is given by

$$h(x,y,t)+\widetilde{H}(x,y) = H+\eta(x,y,t),$$

where the constant H is the height of the free surface at rest, and the depth of the fluid, h, in terms of η is given by

$$h(x,y,t) = H+\eta(x,y,t)-\widetilde{H}(x,y).$$

Therefore equations (2.1) to (2.3) become

$$\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}-fv = -g\frac{\partial \eta}{\partial x} \tag{2.25}$$

$$\frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+fu = -g\frac{\partial \eta}{\partial y} \tag{2.26}$$

$$\frac{\partial \eta}{\partial t}+u\frac{\partial}{\partial x}(\eta-\widetilde{H})+v\frac{\partial}{\partial y}(\eta-\widetilde{H})+\left(H+\eta-\widetilde{H}\right)\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)=0. \tag{2.27}$$

We now scale equations (2.25) to (2.27) and write them in terms of non dimensional variables x^*, y^*, u^*, v^*, t^* and η^* where
$x = Lx^*, y = Ly^*, u = Uu^*, v = Uv^*, t = Tt^*$ as before and let

$$\eta = N_o \eta^*.$$

Here L, U, T and N_0 are characteristic length, velocity, time and surface height departure scales chosen so that the magnitude of the non dimensional variables are order *unity*. The scaled equations are then

$$\frac{U}{T}\frac{\partial u}{\partial t} + \frac{U^2}{L}\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right) - fUv = -\frac{gN_0}{L}\frac{\partial \eta}{\partial x} \tag{2.28}$$

$$\frac{U}{T}\frac{\partial v}{\partial t} + \frac{U^2}{L}\left(u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right) + fUu = -\frac{gN_0}{L}\frac{\partial \eta}{\partial y} \tag{2.29}$$

$$\begin{aligned}\frac{N_0}{T}\frac{\partial \eta}{\partial t} &+ \frac{U}{L}\left(u\frac{\partial}{\partial x}(N_0\eta - H) + v\frac{\partial}{\partial y}(N_0\eta - H)\right) \\ &+ \frac{U}{L}(H + N_0\eta - \widetilde{H})\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0,\end{aligned} \tag{2.30}$$

where the superscripts have been dropped from the non-dimensional variables. Now we look for motions where the Rossby number, R_o, is small, i.e.

$$R_o = \frac{U}{fL} = \epsilon << 1.$$

We also assume that

$$\epsilon_T = \frac{1}{fT} << 1,$$

so that the time scale T is much greater than the inertial time $1/f$. We then *choose* N_0 such that

$$fU = gN_0/L,$$

and therefore the right hand sides of (2.28) and (2.29) are the same order as the Coriolis terms in the equations allowing non-trivial solutions. This is a fair assumption if we consider typical values for large-scale atmospheric motions in the mid-latitudes. Taking characteristic values; $f = 0.0001s^{-1}$, $U = 10ms^{-1}$, $L = 10^6 m$

and $g = 10ms^2$, gives $N_0 = 100m$. This is a reasonable scale of departure from the atmospheric depth H, which is of order $10^5 m$. Therefore we have

$$N_0 = \frac{fUL}{g} = \epsilon \frac{f^2 L^2}{g}$$

and

$$h = H\left(1 + \epsilon B_u^2 \eta - \frac{\widetilde{H}}{H}\right),$$

where B_u is the Burger number defined in equation (2.17). Dividing momentum equations (2.28) and (2.29) by fU and equation (2.30) by HU/L gives

$$\epsilon_T \frac{\partial u}{\partial t} + \epsilon\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right) - v = -\frac{\partial \eta}{\partial x} \tag{2.31}$$

$$\epsilon_T \frac{\partial v}{\partial t} + \epsilon\left(u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right) + u = -\frac{\partial \eta}{\partial y} \tag{2.32}$$

$$\begin{aligned}\epsilon_T B_u^2 \frac{\partial \eta}{\partial t} + \epsilon B_u^2\left(u\frac{\partial \eta}{\partial x} + v\frac{\partial \eta}{\partial y}\right) - u\frac{\partial}{\partial x}\left(\frac{\widetilde{H}}{H}\right) - v\frac{\partial}{\partial y}\left(\frac{\widetilde{H}}{H}\right) \\ +(1 + \epsilon B_u^2 \eta - \frac{\widetilde{H}}{H})\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0,\end{aligned} \tag{2.33}$$

We now assume that

$$B_u = O(1),$$

and that the time scale T is given by the advective time L/U, as in equation (2.13), therefore

$$\epsilon_T = \epsilon.$$

We are now in a position to examine the orders of magnitude of the terms in the equations. This allows us to determine relationships between terms of equal magnitude. Considering each variable as a function of x, y, t and ϵ we can expand in a power series as follows

$$u(x, y, t, \epsilon) = u_0(x, y, t) + \epsilon u_1(x, y, t) + \ldots$$

$$v(x, y, t, \epsilon) = v_0(x, y, t) + \epsilon v_1(x, y, t) + \ldots$$

$$\eta(x, y, t, \epsilon) = \eta_0(x, t) + \epsilon \eta_1(x, t) + \ldots,$$

where expansion functions u_k, v_k, η_k are $O(1)$ and independent of ϵ. We then substitute into equations (2.31) to (2.33). Since ϵ is arbitrary, terms of like order in ϵ must *balance*.

The $O(1)$ terms are given by

$$v_0 = \frac{\partial \eta_0}{\partial x} \tag{2.34}$$

and

$$u_0 = -\frac{\partial \eta_0}{\partial y}, \tag{2.35}$$

which is the geostrophic relationship, given by equations (2.23) and (2.24), but in a non-dimensional form. The first order, linear balance equations (2.34) and (2.35) imply that,

$$\frac{\partial u_0}{\partial x} + \frac{\partial v_0}{\partial y} = 0 \tag{2.36}$$

and so the $O(1)$ terms are non-divergent.

Assuming that $\widetilde{H}/H = O(\epsilon)$, which is reasonable since a characteristic orographic height would be much less than the atmospheric depth, we are unable to determine the fields u_0, v_0 and η_0 from just the $O(1)$ terms. This is because the $O(1)$ terms in (2.33) just give us equation (2.36) again and no additional information.

We now look to the $O(\epsilon)$ terms in the asymptotic expansion

$$\frac{\partial u_0}{\partial t} + u_0\frac{\partial u_0}{\partial x} + v_0\frac{\partial u_0}{\partial y} - v_1 = -\frac{\partial \eta_1}{\partial x} \tag{2.37}$$

$$\frac{\partial v_0}{\partial t} + u_0\frac{\partial v_0}{\partial x} + v_0\frac{\partial v_0}{\partial y} + u_1 = -\frac{\partial \eta_1}{\partial y} \tag{2.38}$$

$$B_u\left(\frac{\partial \eta_0}{\partial t} + u_0\frac{\partial \eta_0}{\partial x} + v_0\frac{\partial \eta_0}{\partial y}\right) - u_0\frac{\partial \eta_b}{\partial x} - v_0\frac{\partial \eta_b}{\partial y} + \left(\frac{\partial u_1}{\partial x} + \frac{\partial v_1}{\partial y}\right) = 0, \tag{2.39}$$

where we have also assumed that the orographic height is order ϵ smaller than the fluid depth and we write $\widetilde{H}(x,y)/H$ in terms of a new function $\eta_b(x,y)$ given by

$$\frac{\widetilde{H}}{H} = \epsilon\eta_b(x,y),$$

with $\eta_b = O(1)$. Here we see that the $O(\epsilon)$ velocities u_1 and v_1 are not in geostrophic balance with the $O(\epsilon)$ pressure terms. In fact, departures from this

balance are caused by the material derivative of the $O(1)$ velocities, which are geostrophic.

If we continue with this analysis a closed system can be obtained by eliminating η_1 from equations (2.37) and (2.38). This leads to the quasigeostropic scaling of the SWEs. This scaling of the SWEs can be used to investigate many types of atmospheric motions and is an example of a balanced model [63]. Further analysis of quasigeostrophic motion is not required here but it can be found in [49] and [12].

2.2.2 Summary

In this section we defined the SWEs and demonstrated several important properties that make the SWEs a suitable test bed for atmospheric research. We derived a conserved quantity, the PV. We then defined several important dimensionless parameters; the Rossby, Froude and the Burger numbers. We then used a crude order-of-magnitude argument to demonstrate balance in the SWEs. This was then formally reinforced through an asymptotic expansion in the Rossby number. We found that the order 1 terms in the asymptotic expansion are in geostrophic balance. Geostrophic balance is a simple, first order, balance approximation that we use throughout this work. Higher order balance conditions are examined in detail in [18].

We now turn the our attention to the solutions of a linearised form of SWEs. The link between geostrophic balance and the slow solution of the linearised SWEs is then demonstrated.

2.3 Analysis of the SWEs

We now consider the behaviour of the SWEs for small amplitude motions. This section draws on [9], [12], [48] and [49]. We see how the linearised equations are solved through a normal mode analysis, resulting in decoupled system of equations in new dependent variables. We then extend this analysis in a way similar to [9] and [48] to develop an idea of what the decoupled variables represent in physical

space. We find that the normal modes are related to the linearised PV, the geostrophic departure and the divergence.

2.3.1 The linearised SWEs

To analyse the kind of motions possible in the SWEs we examine solutions of a linearised form of equations (2.1) to (2.3). This is done by assuming only small amplitude oscillations about a state of rest.

The thickness of the undisturbed fluid layer at rest is given by $\bar{h}$, which is constant in space and time, and we let $f = f_0$, a constant, this is referred to as the f-plane assumption. We then linearise equations (2.1) to (2.3) about a state of rest $\bar{u} = 0, \bar{v} = 0$

$$u = \bar{u} + u' = u'$$

$$v = \bar{v} + v' = v'$$

$$h = \bar{h} + h',$$

where the primed variables are perturbations. Substituting into equations (2.1) to (2.3) and ignoring products of u', v' and h', we obtain

$$\frac{\partial u'}{\partial t} + g\frac{\partial \phi}{\partial x} - f_0 v' = 0 \tag{2.40}$$

$$\frac{\partial v'}{\partial t} + g\frac{\partial \phi}{\partial y} + f_0 u' = 0 \tag{2.41}$$

$$\frac{\partial \phi}{\partial t} + \bar{\phi}\left(\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y}\right) = 0, \tag{2.42}$$

where $\bar{\phi} = g\bar{h}$ and $\phi = gh'$ is the deviation from $\bar{\phi}$.

We also note, using the PV equation (2.9), that under this linearisation the linearised PV, q', is given by

$$q' = \frac{1}{\bar{h}}\left(\zeta' - \frac{\phi f_0}{\bar{\phi}}\right), \tag{2.43}$$

where ζ' is a perturbation to the relative vorticity defined in equation (2.6), and $q = \bar{q} + q'$ with $\bar{q} = f/\bar{h}$. We note that all of the PV linearisations we use in this work are found by linearising the PV equation directly, as we show later in section

2.4.2. An alternative approach would be to linearise the model equations before deriving the analogue of linearised PV for the linearised equations.
We choose to follow [12] and write the above equations in terms of the streamfunction ψ and velocity potential χ using the Helmholtz theorem. This gives

$$u' = -\frac{\partial \psi}{\partial y} + \frac{\partial \chi}{\partial x}, \tag{2.44}$$

and

$$v' = \frac{\partial \psi}{\partial x} + \frac{\partial \chi}{\partial y}, \tag{2.45}$$

where the *streamfunction*, ψ, is defined by

$$\nabla^2 \psi = \zeta \tag{2.46}$$

and the *velocity potential*, χ, by

$$\nabla^2 \chi = \mathcal{D}, \tag{2.47}$$

with $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the horizontal Laplacian and the vorticity ζ and divergence $\mathcal{D}$ are given by equations (2.6) and (2.7) respectively.
We derive an equation for the vorticity by calculating $\frac{\partial (2.41)}{\partial x} - \frac{\partial (2.40)}{\partial y}$ giving

$$\frac{\partial}{\partial t}\nabla^2 \psi + f_0 \nabla^2 \chi = 0. \tag{2.48}$$

An equation for the divergence is found by taking $\frac{\partial (2.40)}{\partial x} + \frac{\partial (2.41)}{\partial y}$ to obtain

$$\frac{\partial}{\partial t}\nabla^2 \chi - f_0 \nabla^2 \psi + \nabla^2 \phi = 0. \tag{2.49}$$

We can also write equation (2.42) in terms of the divergence,

$$\frac{\partial \phi}{\partial t} + \bar{\phi} \nabla^2 \chi = 0. \tag{2.50}$$

We now look for solutions of the linear system of equations defined by (2.48) to (2.50). We do this by finding the normal modes of the system. This results in a decoupled system of transformed equations that can easily be solved.

2.3.2 Solutions Of The Linearised SWEs

We now find the normal modes of the linear system given by equations (2.48) to (2.50). We do this primarily to demonstrate the link between the normal modes and physically meaningful variables that we use later in this work to define our control variables. As well as this the normal mode analysis also helps us to understand the behaviour of the SWEs.

To find the normal modes of equations (2.48) to (2.50) we follow [12] and assume the following dependence of ψ, χ and ϕ on x and y,

$$\psi(x,y,t) = \hat{\psi}_n^m(t)e^{i\left(\frac{(mx+ny)}{a}\right)} \tag{2.51}$$

$$\chi(x,y,t) = i\hat{\chi}_n^m(t)e^{i\left(\frac{(mx+ny)}{a}\right)} \tag{2.52}$$

$$\phi(x,y,t) = f_0\sqrt{K}\hat{\phi}_n^m(t)e^{i\left(\frac{(mx+ny)}{a}\right)}, \tag{2.53}$$

where constant

$$K = \frac{(m^2+n^2)\bar{\phi}}{a^2 f_0^2}, \tag{2.54}$$

(m,n) are the wavenumbers and a is the radius of the earth. This gives a periodic domain of length $2\pi a$ in both the x and y directions. We note that the constant K defined above can be written as

$$K = \frac{L_r^2}{L^2} = B_u^2, \tag{2.55}$$

where $L_r = \sqrt{\bar{\phi}}/f_0$ is the Rossby radius, L is the length scale for wavenumbers m,n given by $L = \sqrt{(a^2/n^2+m^2)}$, and B_u is the Burger number defined in equation (2.17).

Substituting (2.51) to (2.53) into equations (2.48) to (2.50) we obtain

$$\frac{d\hat{\psi}}{dt} + if_0\hat{\chi} = 0, \tag{2.56}$$

$$i\frac{d\hat{\chi}}{dt} - f_0\hat{\psi} + f_0\sqrt{K}\hat{\phi} = 0, \tag{2.57}$$

$$\frac{d\hat{\phi}}{dt} - if_0\sqrt{K}\hat{\chi} = 0, \tag{2.58}$$

where $\hat{\psi} = \hat{\psi}_n^m$, $\hat{\chi} = \hat{\chi}_n^m$, and $\hat{\phi} = \hat{\phi}_n^m$. We now assume a temporal dependence of the form

$$\hat{\psi}(t) = \Psi e^{-if_o\sigma t}, \tag{2.59}$$

$$\hat{\chi}(t) = X e^{-if_o\sigma t}, \tag{2.60}$$

$$\hat{\phi}(t) = \Phi e^{-if_o\sigma t}, \tag{2.61}$$

where σ is a non-dimensional frequency and Ψ, X and Φ are constants. Therefore substituting (2.51) to (2.53) into equations (2.48) to (2.50) we obtain

$$(L - \sigma I) \begin{pmatrix} \Psi \\ X \\ \Phi \end{pmatrix} = 0, \tag{2.62}$$

for each horizontal scale (m, n) where I is the identity and

$$L = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -\sqrt{K} \\ 0 & -\sqrt{K} & 0 \end{pmatrix}. \tag{2.63}$$

We see that the form of (2.51) to (2.53) is chosen to make L real and symmetric. By diagonalising L we are able to decouple the system into three indepenently evolving equations in time that can easily be solved. We therefore solve an eigenvalue problem with eigenvalue σ. The matrix L has three eigenvectors and three eigenvalues for a given (m, n). The three eigenvalues are found by setting the determinant of $L - \sigma I$ equal to zero. If this is calculated we obtain a cubic frequency equation for σ given by

$$\sigma^3 - \sigma K - \sigma = 0,$$

and therefore

$$\sigma = 0, \pm\sqrt{1 + K}.$$

The eigenvectors are found by substituting each eigenvalue back into

$$(L - \sigma I)\,\mathbf{e} = 0,$$

where $\mathbf{e}$ is the eigenvector of eigenvalue σ. Each eigenvector is defined up to a multiplicative constant. We can therefore fix one element to be 1 say and determine the remaining two elements.

To determine the eigenvector of the first eigenvalue $\sigma = 0$ we solve

$$(L - \sigma I)\,\mathbf{e} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -\sqrt{K} \\ 0 & -\sqrt{K} & 0 \end{pmatrix} \begin{pmatrix} R_\psi \\ R_\chi \\ R_\phi \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{2.64}$$

for R_ψ, R_χ, and R_ϕ, where R_ψ, R_χ, and R_ϕ are the ψ, χ, and ϕ elements of the eigenvector $\mathbf{e}$. This is done by setting $R_\phi = 1$ and obtaining $R_\psi = \sqrt{K}$ and $R_\chi = 0$. Therefore the eigenvector for the eigenvalue $\sigma = 0$ is

$$\begin{pmatrix} R_\psi \\ R_\chi \\ R_\phi \end{pmatrix} = \frac{1}{\sqrt{1+K}} \begin{pmatrix} \sqrt{K} \\ 0 \\ 1 \end{pmatrix}. \tag{2.65}$$

This eigenvector is called the geostrophic, or Rossby mode because it describes a stationary, but non-trivial, geostrophic solution. If we were to use a non-constant f, or include orography, this solution would no longer be stationary, and would become the Rossby wave solution.

We note that the normalisation factor, $1/\sqrt{1+K}$, in (2.65) is chosen as in [12], making the magnitude 1. If the magnitude of each eigenvector is 1 then the matrix formed by the eigenvectors of L is orthonormal.

It is possible to demonstrate that this mode is geostrophic by substituting the eigenvector into the assumed temporal dependence (2.59) and (2.61) and we see that this mode implies

$$\Psi = \sqrt{K}\Phi.$$

In original model variables, using (2.59), (2.61), (2.51) and (2.53), this gives

$$f_0\psi = \phi.$$

Since the geostrophic mode is non divergent, i.e. $X = 0$, we have from the Helmholtz theorem, equations (2.44) and (2.45),

$$u' = -\frac{\partial \psi}{\partial y} = -\frac{1}{f_0}\frac{\partial \phi}{\partial y} \tag{2.66}$$

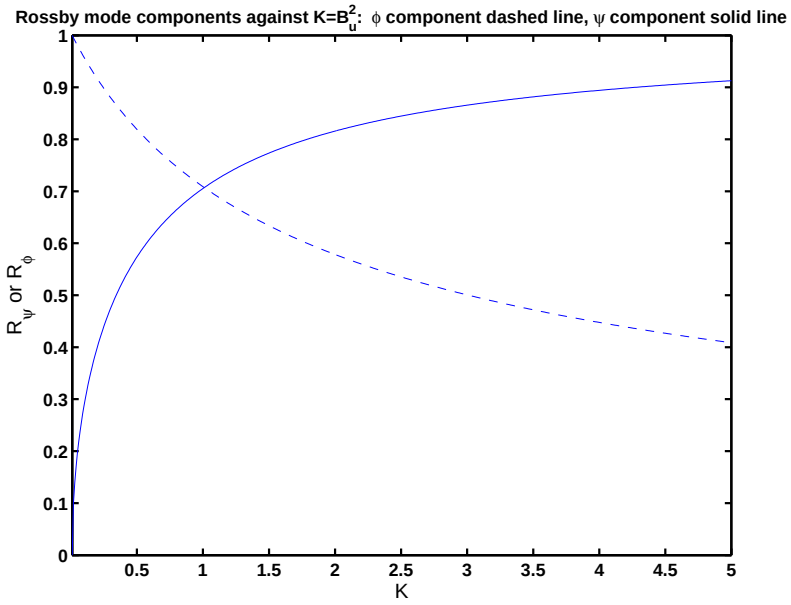

Figure 2.2: Plot of the Rossby mode components, R_ψ (solid line) and R_ϕ (dashed line), against K.

and

$$v' = \frac{\partial \psi}{\partial x} = \frac{1}{f_0}\frac{\partial \phi}{\partial x}, \tag{2.67}$$

which is the geostrophic relation, equations (2.23) and (2.24), derived previously. If we plot the Rossby mode components against K we are able to see how the relative contributions of each component changes with K. A plot of the Rossby mode components against K is shown in figure 2.2. Here we see there are 3 cases; $K << 1$, $K = 1$ and $K >> 1$. For $K = B_u^2 = 1$, using equation (2.55), we see that the contributions of R_ψ and R_ϕ to the Rossby mode are equal. For $K = B_u << 1$ the mode is dominated by the R_ϕ and $K = B_u >> 1$ it is dominated by R_ψ.
We now find the eigenvector associated with the positive eigenvalue $\sqrt{1+K}$. It is found by setting the ψ element of the to 1 in a similar as we did for the Rossby mode. The eigenvector is given by

$$\begin{pmatrix} G^1_\psi \\ G^1_\chi \\ G^1_\phi \end{pmatrix} = \frac{1}{\sqrt{2(1+K)}} \begin{pmatrix} 1 \\ \sqrt{1+K} \\ -\sqrt{K} \end{pmatrix}, \tag{2.68}$$

where G^1_ψ, G^1_χ, and G^1_ϕ are the ψ, χ, and ϕ elements of the eigenvector.
For the negative eigenvalue $\sigma = -\sqrt{1+K}$ we have

$$\begin{pmatrix} G^2_\psi \\ G^2_\chi \\ G^2_\phi \end{pmatrix} = \frac{1}{\sqrt{2(1+K)}} \begin{pmatrix} 1 \\ -\sqrt{1+K} \\ -\sqrt{K} \end{pmatrix}, \tag{2.69}$$

where G^2_ψ, G^2_χ, and G^2_ϕ are the ψ, χ, and ϕ elements of the eigenvector of the negative eigenvalue. These two modes are the inertial-gravity modes.
If we calculate the dimensional frequency of the inertial-gravity modes by multiplying the non-dimensional frequency σ by f_0 we obtain

$$f_o\sigma = \sqrt{f_0^2 + \frac{(m^2+n^2)g\bar{h}}{a^2}}.$$

This frequency is made up of two terms. The first term describes is the inertial frequency and the second is the non-rotating gravity wave frequency. The two terms are equal if

$$\frac{1}{\sqrt{(n^2+m^2)/a^2}} = \frac{\sqrt{g\bar{h}}}{f_0} \equiv L_r,$$

where L_r is the Rossby radius of deformation defined in equation (2.20). The non-rotating gravity wave speed is given by

$$c_g = \sqrt{g\bar{h}}. \tag{2.70}$$

To decouple the system we expand $\hat{\psi}(t), \hat{\chi}(t)$, and $\hat{\phi}(t)$ in the eigenfunctions (2.65), (2.68) and (2.69) and define $r(t), g_1(t)$, and $g_2(t)$ as the expansion coefficients for the Rossby and inertial-gravity modes. We then obtain

$$\begin{pmatrix} \hat{\psi}(t) \\ \hat{\chi}(t) \\ \hat{\phi}(t) \end{pmatrix} = E \begin{pmatrix} r(t) \\ g_1(t) \\ g_2(t) \end{pmatrix} \tag{2.71}$$

where

$$E = \begin{pmatrix} R_\psi & G^1_\psi & G^2_\psi \\ R_\chi & G^1_\chi & G^2_\chi \\ R_\phi & G^1_\phi & G^2_\phi \end{pmatrix} \tag{2.72}$$

is a matrix whose columns are the eigenvectors of L. The matrix L, defined by (2.63), is real and symmetric and so has an orthonormal basis of eigenvectors. The eigenvectors (2.65), (2.68) and (2.69) are normalised so that there magnitude is 1 and therefore E is orthogonal. We can then write

$$\begin{pmatrix} r(t) \\ g_1(t) \\ g_2(t) \end{pmatrix} = E^{-1} \begin{pmatrix} \hat{\psi}(t) \\ \hat{\chi}(t) \\ \hat{\phi}(t) \end{pmatrix} = E^T \begin{pmatrix} \hat{\psi}(t) \\ \hat{\chi}(t) \\ \hat{\phi}(t) \end{pmatrix}. \tag{2.73}$$

We can now decouple the system. We first write equations (2.56) to (2.58) in the form

$$\begin{pmatrix} \frac{d\hat{\psi}}{dt} \\ \frac{d\hat{\chi}}{dt} \\ \frac{d\hat{\phi}}{dt} \end{pmatrix} = -if_0 L \begin{pmatrix} \hat{\psi} \\ \hat{\chi} \\ \hat{\phi} \end{pmatrix}, \tag{2.74}$$

which in terms of $r(t), g_1(t)$, and $g_2(t)$ is

$$\begin{pmatrix} \frac{dr}{dt} \\ \frac{dg_1}{dt} \\ \frac{dg_2}{dt} \end{pmatrix} = -if_0 \Lambda \begin{pmatrix} r \\ g_1 \\ g_2 \end{pmatrix}, \tag{2.75}$$

where $\Lambda = E^{-1}LE$ and

$$\Lambda = \begin{pmatrix} \sigma_R & 0 & 0 \\ 0 & \sigma_G^1 & 0 \\ 0 & 0 & \sigma_G^2 \end{pmatrix},$$

a diagonal matrix of eigenvalues. The system given by (2.75) is therefore decoupled and reduces to

$$\frac{dr}{dt} = -i\sigma_R f_0 r(t),$$
$$\frac{dg_1}{dt} = -i\sigma_G^1 f_0 g_1(t)$$

and

$$\frac{dg_2}{dt} = -i\sigma_G^2 f_0 g_2(t)$$

in terms of new dependent variables $r(t), g_1(t)$ and $g_2(t)$. The equations have simple solutions

$$r(t) = r(0)exp(-i\sigma_R f_0 t) = r(0),$$

$$g_1(t) = g_1(0)exp(-i\sigma_G^1 f_0 t)$$

and

$$g_2(t) = g_2(0)exp(-i\sigma_G^2 f_0 t).$$

Before we give physical meaning to the new dependent variables $r(t), g_1(t)$, and $g_2(t)$ we first find the solutions in terms of $\hat{\psi}(t), \hat{\chi}(t)$, and $\hat{\phi}(t)$.
To find the solution for $\hat{\psi}(t)$ we use equation (2.71) and write $\hat{\psi}(t)$ in terms of $r(t), g_1(t)$, and $g_2(t)$, giving

$$\begin{aligned}\hat{\psi}(t) &= R_\psi r(t) + G_\psi^1 g_1(t) + G_\psi^2 g_2(t) \\ &= R_\psi r(0) + G_\psi^1 g_1(0)exp(-i\sigma_G^1 f_0 t) + G_\psi^2 g_2(0)exp(-i\sigma_G^2 f_0 t)(t).\end{aligned}$$

Now we use equation (2.73) to write $r(0), g_1(0)$, and $g_2(0)$ in terms of original variables $\hat{\psi}, \hat{\chi}$, and $\hat{\phi}$. We then obtain, after some manipulation,

$$\hat{\psi}(t) = \frac{K\hat{\psi}_0 + \sqrt{K}\hat{\phi}_0}{1+K} + \frac{\hat{\psi}_0 - \sqrt{K}\hat{\phi}_0}{1+K}\cos\alpha t - \frac{i\hat{\chi}_0}{\sqrt{1+K}}\sin\alpha t, \tag{2.76}$$

where $\alpha = f_0\sqrt{1+K}$ and $\hat{\psi}_0, \hat{\chi}_0$ and $\hat{\phi}_0$ are the initial $\hat{\psi}, \hat{\chi}$ and $\hat{\phi}$ fields. We can follow the same steps and find the solutions for $\hat{\chi}(t)$ and $\hat{\phi}(t)$. These are given by

$$i\hat{\chi}(t) = \frac{\hat{\psi}_0 - \sqrt{K}\hat{\phi}_0}{1+K}\sin\alpha t + i\hat{\chi}_0\cos\alpha t, \tag{2.77}$$

and

$$\sqrt{K}\hat{\phi}(t) = \frac{K\hat{\psi}_0 + \sqrt{K}\hat{\phi}_0}{1+K} - \frac{K(\hat{\psi}_0 - \sqrt{K}\hat{\phi}_0)}{1+K}\cos\alpha t + \frac{iK\hat{\chi}_0}{\sqrt{1+K}}\sin\alpha t. \tag{2.78}$$

Geostrophic Adjustment and the Rossby Radius of Deformation

We notice that the solutions (2.76), (2.77) and (2.78) have a stationary component and a time-dependent component. The stationary solution is given by

$$\hat{\psi}(t) = \sqrt{K}\hat{\phi}(t) = \frac{K\hat{\psi}_0 + \sqrt{K}\hat{\phi}_0}{1+K} \tag{2.79}$$

and $\hat{\chi}(t) = 0$. This is the geostrophic solution. This stationary, geostrophically balanced solution is dependent on the initial perturbation and also the Rossby radius. To understand this dependence we recall equation (2.54) and write

$$K = \frac{(m^2+n^2)\bar{\phi}}{a^2 f_0^2} = \frac{L_r^2}{L^2} = B_u^2, \tag{2.80}$$

where $L_r = \sqrt{\bar{\phi}}/f_0$ is the Rossby radius, L is a characteristic length scale with $L = \sqrt{(a^2/n^2 + m^2)}$ and B_u is the Burger number defined in equation (2.17). Writing (2.79) in terms of L_r then gives

$$\hat{\psi}(t) = \frac{L_r}{L}\hat{\phi}(t) = \frac{\hat{\psi}_0 + (L/L_r)\hat{\phi}_0}{1 + (L^2/L_r^2)}. \tag{2.81}$$

Here the time-dependent solution is not dispersed, but in cases where it is this process is known as geostrophic adjustment and in [24] the process is shown to have a characteristic time scale less than or equal to f^{-1}.

There are three cases we can consider for equation (2.81), as was demonstrated for equation (2.65); $B_u << 1$, $B_u = 1$ and $B_u >> 1$. For $B_u >> 1$ we have

$$\hat{\psi}(t) = B_u\hat{\phi}(t) \approx \hat{\psi}_0, \tag{2.82}$$

and the stationary geostrophic component is equal to the initial ψ perturbation. For $B_u << 1$ we have

$$\hat{\psi}(t) = B_u\hat{\phi}(t) \approx B_u\hat{\phi}_0, \tag{2.83}$$

and the stationary component is equal to the initial ϕ perturbation. When $B_u = 1$

$$\hat{\psi}(t) = B_u\hat{\phi}(t) = \frac{\hat{\psi}_0 + \hat{\phi}_0}{2}, \tag{2.84}$$

so the stationary component is the average of both the initial ψ and ϕ perturbations. These results are in agreement with the behaviour of the Rossby mode observed in figure 2.2.

2.3.3 Normal Modes In Physical Space

Now that we have found solutions to the SWEs we would like to understand what the new dependent variables $r(t), g_1(t)$, and $g_2(t)$ represent in physically. To do this we can use equation (2.73) and the orthogonality condition $E^{-1} = E^T$. We write (2.73) as

$$\begin{pmatrix} r(t) \\ g_1(t) \\ g_2(t) \end{pmatrix} = \begin{pmatrix} R_\psi & R_\chi & R_\phi \\ G^1_\psi & G^1_\chi & G^1_\phi \\ G^2_\psi & G^2_\chi & G^2_\phi \end{pmatrix} \begin{pmatrix} \hat{\psi}(t) \\ \hat{\chi}(t) \\ \hat{\phi}(t) \end{pmatrix}, \tag{2.85}$$

we can then project onto the Rossby mode, equation (2.65), by setting the inertial-gravity components to zero and calculating

$$y = \begin{pmatrix} R_\psi & R_\chi & R_\phi \end{pmatrix} \begin{pmatrix} \hat{\psi}(t) \\ \hat{\chi}(t) \\ \hat{\phi}(t) \end{pmatrix}. \tag{2.86}$$

To see what this represents in physical space we multiply y by a constant

$$C_1 = \frac{\sqrt{K}\sqrt{1+K}f_0^2}{\bar{h}\bar{\phi}}$$

giving

$$C_1 y = -\frac{1}{\bar{h}}\left(\frac{m^2+n^2}{a^2}\hat{\psi} + \frac{f_0^2\sqrt{K}}{\bar{\phi}}\hat{\phi}\right).$$

We recognise that, in the original perturbation variables this is the linearised PV q' given by (2.43). Thus we find that not only is the Rossby mode geostrophically balance but it is also characterised by the linearised PV. This association of the PV and the geostrophically balanced flow is exploited later to define a set of balanced control variables.

To interpret what the remaining two variables $g_1(t)$ and $g_2(t)$ represent in physical space we take linear combinations of the inertial-gravity eigenvectors. We see that

$$C_2 \begin{pmatrix} (G^1_\psi + G^2_\psi) & (G^1_\chi + G^2_\chi) & (G^1_\phi + G^2_\phi) \end{pmatrix} \begin{pmatrix} \hat{\psi}(t) \\ \hat{\chi}(t) \\ \hat{\phi}(t) \end{pmatrix} = -\left(\frac{m^2+n^2}{a^2}\right)(\hat{\psi} - \sqrt{K}\hat{\phi}),$$

where the constant $C_2 = -f_0\sqrt{2(1+K)}\frac{m^2+n^2}{2a^2}$, and we recognise this quantity, in original perturbation variables, as the geostrophic departure

$${}_a\zeta = f_0\nabla^2\psi - \nabla^2\phi. \tag{2.87}$$

Also

$$C_3 \begin{pmatrix} (G^2_\psi - G^1_\psi) & (G^2_\chi - G^1_\chi) & (G^2_\phi - G^1_\phi) \end{pmatrix} \begin{pmatrix} \hat{\psi}(t) \\ \hat{\chi}(t) \\ \hat{\phi}(t) \end{pmatrix} = -\left(\frac{m^2+n^2}{a^2}\right)\hat{\chi},$$

where the constant $C_3 = -\frac{m^2+n^2}{2\sqrt{1+K}a^2}$, is the divergence $\mathcal{D}$ in original perturbation variables.

An important extension of these ideas is finding the balanced component of any given set of data. This is achieved by projecting the data onto the basis of eigenfunctions. Then all but the Rossby mode is set to zero and we project back into physical variables, giving only the balanced component. This is the Rossby adjustment problem and it has important applications in NWP, see for example [61].

In [18] the normal modes of the SWEs are used to define a mapping from physical space to the linearised PV, divergence and the geostrophic departure (or ageostrophic vorticity in [18]). The SWEs are then written in terms of the linearised PV, divergence and the geostrophic departure and the accuracy of various balance approximations are examined numerically. A similar mapping to that of [18] is used later to define a new set of variables based on the normal modes of the SWEs.

2.3.4 Summary

In analysing the solutions to this linearised version of the SWEs we have found two types of motion; one mode is the slow, balanced Rossby mode, the other two, fast, unbalanced inertial-gravity modes. We have demonstrated the link between the balanced geostrophic solution and the PV. The unbalanced solutions are related to the divergence and geostrophic departure. Later in this work we define control variable transformations that attempt to exploit these relationships.

We now examine how relative contributions to the linearised PV, which characterises the balanced solution, may change with the Burger number.

2.4 Potential Vorticity and Burger Regimes

We have seen the relationship between the PV and the balanced Rossby mode. We now consider the behaviour of vorticity and height fields that satisfy both the

linearised PV equation (2.43) and the linear balance equation defined in the next section. Here the linearisation state is non-trival. We show how the ratio of absolute vorticity and height changes with Burger number. We start by defining the linear balance equation.

2.4.1 The Linear Balance Equation

The *linear balance equation* (LBE) can be derived from the geostrophic mode found in the linear analysis in section 2.3.2. We combine the geostrophic, or Rossby mode equations (2.66) and (2.67) by taking $\partial(2.67)/\partial x - \partial(2.66)/\partial y$ to give a *linear balance equation*

$$f_0\zeta = \nabla^2\phi, \tag{2.88}$$

or, using the streamfunction ψ

$$f_0\nabla^2\psi = \nabla^2\phi, \tag{2.89}$$

where the vorticity ζ is given by

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2\psi$$

and the geopotential height ϕ by

$$\phi = gh.$$

The LBE forms a fundamental part of this research and is used throughout this work in the form of equation (2.89).

2.4.2 Contributions to the Linearised PV

We now look at how the relative contributions of height and vorticity to the linearised PV change with the Burger number. First, we derive an equation for the linearised PV q'. We linearise u, v and h about a varying linearisation state to obtain

$$u(x,y,t) = \bar{u}(x,y,t) + u'(x,y,t),$$

$$v(x,y,t) = \bar{v}(x,y,t) + v'(x,y,t)$$

and

$$h(x,y,t) = \bar{h}(x,y,t) + h'(x,y,t).$$

Then we have

$$\bar{\zeta} = \frac{\partial \bar{v}}{\partial x} - \frac{\partial \bar{u}}{\partial y} \qquad\qquad \zeta' = \frac{\partial v'}{\partial x} - \frac{\partial u'}{\partial y},$$

and

$$\bar{q} = \frac{\bar{\zeta} + f_0}{\bar{h}}. \tag{2.90}$$

Therefore

$$q = \bar{q} + q' = \frac{\bar{\zeta} + \zeta' + f_0}{\bar{h} + h'},$$

and neglecting products of perturbations gives

$$\bar{h}\bar{q} + \bar{h}q' + h'\bar{q} = \bar{\zeta} + \zeta' + f_0.$$

Using equation (2.90) we obtain

$$\bar{h}q' + h'\bar{q} = \zeta' \qquad \Rightarrow \qquad q' = \frac{1}{\bar{h}}\left(\zeta' - h'\bar{q}\right),$$

which is similar to equation (2.43). This can be written as

$$\frac{q'}{\bar{q}} = \frac{\zeta'}{\bar{\zeta} + f_0} - \frac{h'}{\bar{h}}. \tag{2.91}$$

We now assume that increments in vorticity, ζ', and height, h', satisfy the scaled linearised PV equation (2.91) and the LBE (2.89). It is then possible to show, following [67], that

$$\frac{q'}{\bar{q}} = -N\frac{h'}{\bar{h}}, \tag{2.92}$$

and

$$\left(1 - \frac{1}{N}\right)\frac{q'}{\bar{q}} = \frac{\zeta'}{\bar{\zeta} + f_0}, \tag{2.93}$$

where

$$N = 1 + \frac{f_0 B_u^2}{\bar{\zeta} + f_0}.$$

The Burger number, B_u, is always positive and so is N. If we consider a fixed $q'/\bar{q}$ and let $N >> 1$ we can see from equation (2.92) that the scaled height increments $h'/\bar{h}$ will not contribute much to the scaled PV increments. However, equation (2.93) tells us that the scaled PV increments will be sensitive to changes in ζ', and

$$\frac{q'}{\bar{q}} \approx \frac{\zeta'}{\bar{\zeta} + f_0}.$$

The value of N will be large for large B_u, and therefore in this regime the PV increments will be well approximated by vorticity.

Following [67] we may also obtain the following relationships

$$\left(1 - \frac{1}{P}\right) \frac{q'}{\bar{q}} = -\frac{h'}{\bar{h}}, \tag{2.94}$$

and

$$\frac{q'}{\bar{q}} = P \frac{\zeta'}{\bar{\zeta} + f_0}, \tag{2.95}$$

where

$$P = 1 + \frac{\bar{\zeta} + f_0}{f_0 B_u^2}.$$

Therefore for fixed $q'/\bar{q}$ and $P << 1$ we have

$$\frac{q'}{\bar{q}} \approx \frac{h'}{\bar{h}}.$$

Hence in a low B_u regime height increments are a good approximation to the scale PV increment.

We have shown in section 2.3.3 that the linearised PV characterises the balanced flow. The analysis above tells us that in a high Burger regime we can approximate the linearised PV, and so the balanced flow, well using vorticity increments. In a low Burger regime it is the height increments that approximate the balance.

2.5 Summary

In this chapter we presented the SWEs and demonstrated some of their many important properties. We started by deriving a conserved quantity, the PV. We then defined three key parameters R_o, B_u and F_r using the non-dimensional form

of the equations. The concept of balance was introduced and we demonstrated that, in the limit of small Rossby number, a first order approximation of the SWEs is given by the geostrophic balance relationship.

The SWEs were then linearised. We found that the linear system supports two types of motion. One is the balanced Rossby mode, which is geostrophic and non-divergent. The two other modes are unbalanced inertial-gravity modes. By decoupling the linear equations into new dependent variables we are able to show that the slow mode is characterised by the linearised form of the PV. The remaining two fast inertial gravity modes are related to the divergence and the geostrophic departure.

We then examined the behaviour of increments in height and absolute vorticity that satisfy both LBE and the linearised PV. We demonstrated that in a high Burger regimes the linearised PV, or the balance flow, is rotational and approximated well by the vorticity. On the other hand in low Burger regimes the balanced flow is characterised by height increments.

In the next chapter we discuss how these properties may be exploited in data assimilation.

Chapter 3

Data Assimilation and the Control Variable Transform

In this chapter we introduce data assimilation and the control variable transform. We aim to provide motivation for, and set in the context of the wider field, the work of the subsequent chapters.

We start with a brief overview of data assimilation. Our attention is then focused on the data assimilation method of interest; four-dimensional variational data assimilation (4D VAR). We discuss the practical issues surrounding the implementation of the algorithm, specifically the inherent problems involved in defining the background error covariance matrix. The control variable transform is central to handling this matrix. We discuss the current implementations of the control variable transform in operational centres around the world and highlight their potential limitations. The PV-based control variable transform is then presented as a possible alternative that should overcome these limitations. We discuss several studies of the PV-based transform, [11], [67] and [68], that, whilst not being conclusive, have provided positive results. The work of this project aims to extend these results and examine further the possible benefits of the PV-based transform.

In the final section we discuss various methods of generating the statistics required to implement the 4D VAR data assimilation system. This includes how

auto-covariance structures are generated and can be modelled in the data assimilation.

3.1 Data Assimilation

Data assimilation is the process of finding the best estimate of the current state of a system. In numerical weather prediction (NWP) this system is the atmosphere and oceans. The 'best estimate' of the current state is referred to as the *analysis* and it is used as the initial conditions for the model forecast. Due to the chaotic nature of the governing equations any errors in the initial conditions may grow rapidly in the forecast [41]. Thus data assimilation forms a vital part of NWP. However, observations are sparse. Typically there may be 10^7 model variables and only 10^6 observations, and additional knowledge of the atmosphere is required. This additional information is known as the *background* and usually comes from a previous short forecast. The data assimilation therefore combines observations, statistical data, knowledge of atmospheric dynamics and a previous short forecast to find the best initial conditions for the model forecast. Recently there have been many advancements in numerical methods, observational networks and computer processing power but it is clear that without a sophisticated data assimilation system these benefits cannot be fully realised.

There are many types of data assimilation methods, for example; optimal interpolation, various Kalman filter based methods, 3D VAR and 4D VAR. Here we will concentrate on 4D VAR as this is currently the most advanced method and is in operational use in an incremental formulation at the Met Office and the ECMWF [53]. There are a variety of practical and theoretical issues surrounding 4D VAR and its implementation, for instance the determination and representation of background error statistics, and it is therefore the focus of much research.

We now present 4D VAR and the incremental 4D VAR algorithm, which is used later in this work. The theory of the control variable transform is then introduced in the context of 4D VAR.

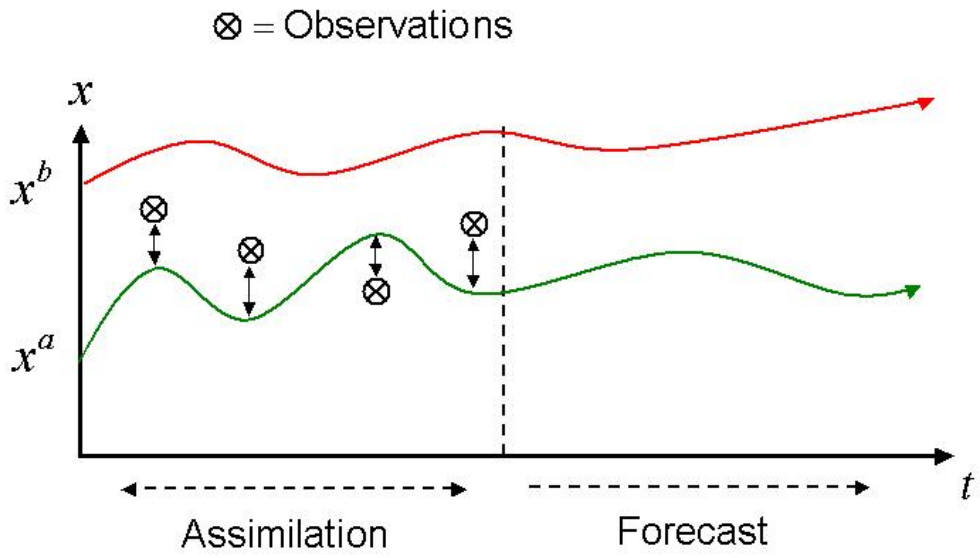

Figure 3.1: 4D VAR diagram

3.1.1 4D VAR

4D VAR data assimilation allows observations to be distributed in time as well as space. One of the earliest proposals of this type of method is given in [57]. This method, however, has a state vector containing all model fields at all times and is prohibitively large to implement practically. In [38] the state vector is reduced to just the model fields at the initial time. This method became known as 4D VAR and is described in [5], [12], [15], [33] and [39].

The objective of the 4D VAR is to find the model state $\mathbf{x}_0$ at time $t = t_0$ that minimises the cost function,

$$J[\mathbf{x}_0] = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{n}(\mathcal{H}_i[\mathbf{x}_i] - \mathbf{y}_i^o)^T \mathbf{R}_i^{-1}(\mathcal{H}_i[\mathbf{x}_i] - \mathbf{y}_i^o), \quad (3.1)$$

with the constraint

$$\mathbf{x}_i = \mathcal{M}(t_i, t_0, \mathbf{x}_0),$$

where $\mathcal{M}(t_i, t_0, \mathbf{x}_0)$ is the non-linear model evolved to time t_i, $i = 1, \ldots, n$, $\mathbf{x}^b$ is

the background field found from a previous short forecast, $\mathbf{y}_i^o$ are the observations at time $t = t_i$ and $\mathcal{H}_i$ is the observation operator that maps model space to observational space. The background error covariance matrix is defined by $\mathbf{B}$, which is of size approximately $O(10^7 \times 10^7)$, and $\mathbf{R}_i$ is the observation error covariance matrix of size $O(10^6 \times 10^6)$ in full operational systems.
A schematic diagram of 4D VAR is shown in figure 3.1. Here we see that the 4D VAR combines observational data available at various times in the data assimilation to improve the initial background guess, $\mathbf{x}^b$. Departures, or innovations, of the background from the observations are calculated at the observation times by evolving the background state at time $t = 0s$ to the observation times. The innovations are then used to find the best least-squares fit between the background and the observations, given the specified error statistics. The analysis state $\mathbf{x}^a$ at time $t = 0$ has a trajectory through the assimilation window that is closer to the observations than the initial guess. A forecast is then run from the end of the assimilation window. When a new set of observations is available the process is repeated using the current forecast as a background state. The potentially non-linear model $\mathcal{M}$ and operator $\mathcal{H}$ means that 4D VAR is a non-linear least squares minimisation problem. Due to the size of the problem 4D VAR has only recently been implemented operationally and it was not until an incremental formulation of the problem was developed in [7] that the method became practical. The incremental formulation is now described.

3.1.2 Incremental 4D VAR

Incremental 4D VAR minimises a series of approximate convex quadratic cost functions [7],

$$\begin{aligned}\widetilde{J}^{(k)}[\mathbf{x}_0'^{(k)}] = \frac{1}{2}(\mathbf{x}_0'^{(k)} - \mathbf{x}'^b)^T\mathbf{B}^{-1}(\mathbf{x}_0'^{(k)} - \mathbf{x}'^b) \\ +\frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\mathbf{x}_i'^{(k)} - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{x}_i'^{(k)} - \mathbf{d}_i),\end{aligned} \tag{3.2}$$

where k is the iteration count and $\mathbf{H}_i$ is the linearised observation operator. Here the increment $\mathbf{x}_i'^{(k)} = \mathbf{M}(t_i, t_0, \mathbf{x}^{(k)})\mathbf{x}_0'^{(k)}$ where $\mathbf{M}(t_i, t_0, \mathbf{x}^{(k)}) \equiv \mathbf{M}_i$ denotes the

linear evolution from t_0 to t_i of the linear model about the current guess trajectory. The model $\mathbf{M}$ is a linear approximation to the non-linear model $\mathcal{M}$, one example of a linear model that might be used is the tangent linear model. This is discussed in section 4.4.1. The background increment $\mathbf{x}'^b$ is given by

$$\mathbf{x}'^b = \mathbf{x}^b - \mathbf{x}_0^{(k)} \tag{3.3}$$

and the innovation vector $\mathbf{d}_i$ by

$$\mathbf{d}_i = \mathbf{y}_i^o - \mathcal{H}_i[\mathbf{x}_i^{(k)}]. \tag{3.4}$$

Even with the incremental formulation of the 4D VAR problem there are still many practical problems to overcome. One of the main difficulties is the specification of the background error covariance matrix. Operationally $\mathbf{B}$ is too large to be used directly and must be modelled in some way. This is done using the control variable transform.

3.2 Control Variable Transforms in Incremental 4D VAR

The modelling of the background error covariance matrix is achieved by transforming from model variables to new *control variables* to perform the data assimilation. When represented as control variables background errors for each control variable are then *assumed* to be *uncorrelated* with each other. In other words the control variable transform is assumed to remove the multi-variate component of the background error covariance, or the cross-covariance. Thus the transformed background error covariance matrix becomes a block diagonal matrix containing only the univariate covariance, or auto-covariance, for each individual control variable. This greatly reduces the size of the problem. We note that operationally further transformations are then applied to remove the auto-covariance. Effectively the problem of modelling and storing the $\mathbf{B}$ matrix has been shifted to defining the control variable transform.

The transform from control variable increments, $\mathbf{z}'$, to model variable increments, $\mathbf{x}'$, is known as the U-transform,

$$\mathbf{x}' = \mathbf{U}\mathbf{z}', \tag{3.5}$$

and its inverse,

$$\mathbf{z}' = \mathbf{T}\mathbf{x}', \tag{3.6}$$

is known as the T-transform. Here $\mathbf{z}'$ are the control variable increments and $\mathbf{x}'$ the model variable increments. Substituting into the incremental cost function (3.2) we obtain,

$$\begin{aligned}\widetilde{J}^{(k)}[\mathbf{z}_0'^{(k)}] = \frac{1}{2}(\mathbf{z}_0'^{(k)} - \mathbf{z}'^b)^T\mathbf{U}^T\mathbf{B}^{-1}\mathbf{U}(\mathbf{z}_0'^{(k)} - \mathbf{z}'^b) \\ +\frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i(\mathbf{M}_i\mathbf{U}\mathbf{z}_0'^{(k)}) - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{H}_i(\mathbf{M}_i\mathbf{U}\mathbf{z}_0'^{(k)}) - \mathbf{d}_i),\end{aligned} \tag{3.7}$$

where $\mathbf{M}_i\mathbf{U}\mathbf{z}_0'^{(k)}$ represents the control variable increment at time t_0 for iteration k transformed to model space and evolved by the linear model to the observation time t_i. If we now assume that the errors in the control variables are uncorrelated we imply that $\mathbf{U}$ is such that

$$\mathbf{U}^T\mathbf{B}^{-1}\mathbf{U} = \mathbf{\Lambda}^{-1}, \tag{3.8}$$

where $\mathbf{\Lambda}$ is a block diagonal matrix specifying the only univariate component, or auto-covariance, of the background error covariance for each control variable. The matrix $\mathbf{\Lambda}$ block diagonal since the cross-covariance between control variables is assumed to be removed by the transform. We see later that further transformations can then be applied to diagonalise $\mathbf{\Lambda}$ though this is not implemented in this work. A much simpler form of (3.2) is then obtained

$$\begin{aligned}\widetilde{J}^{(k)}[\mathbf{z}_0'^{(k)}] = \frac{1}{2}(\mathbf{z}_0'^{(k)} - \mathbf{z}'^b)^T\mathbf{\Lambda}^{-1}(\mathbf{z}_0'^{(k)} - \mathbf{z}'^b) \\ +\frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i(\mathbf{M}_i\mathbf{U}\mathbf{z}_0'^{(k)}) - \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{H}_i(\mathbf{M}_i\mathbf{U}\mathbf{z}_0'^{(k)}) - \mathbf{d}_i).\end{aligned} \tag{3.9}$$

The assumption that $\mathbf{U}$ satisfies equation (3.8) implies that the background error covariance matrix for the orginal model variable background errors is of the form

$$\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T. \tag{3.10}$$

Having now presented the theory of the control variable transform as it applies in the incremental 4D VAR cost function, we defer discussion of the choice of control variables and the details of the transform design until section 3.3. The implementation of the incremental 4D VAR algorithm is now described.

3.2.1 The Incremental 4D VAR algorithm

The incremental 4D VAR algorithm involves an inner and outer iteration. The inner iteration minimises the approximate quadratic cost function, equation (3.9). One iteration of the incremental 4D VAR algorithm described below is known as the *outer loop*. The iterative algorithm proceeds as follows, where k is the iteration number:

1. Define the current guess $\mathbf{x}_0^{(k)}$ with $\mathbf{x}_0^{(0)} = \mathbf{x}^b$.

2. Run the non-linear model with initial conditions $x_0^{(k)}$ to obtain $\mathbf{x}_i$ for each time step t_i.

3. Calculate innovation vectors for each observation

$$\mathbf{d}_i^{(k)} = \mathbf{y}_i^o - \mathcal{H}_i[\mathbf{x}_i^{(k)}].$$

4. Calculate the background increment

$$\mathbf{x}_b'^{(k)} = \mathbf{x}^b - \mathbf{x}_0^{(k)}.$$

5. Transform the background increment

$$\mathbf{z}_b'^{(k)} = \mathbf{T}^{(k)}\mathbf{x}_b'^{(k)},$$

   where $\mathbf{T}^{(k)}$ is the $\mathbf{T}$ transform on outer iteration k.

6. Start the inner loop minimisation: Find $\mathbf{z}_0'^{(k)}$ that minimises (3.9).

7. Transform solution of inner loop minimisation

$$\mathbf{x}_0'^{(k)} = \mathbf{U}^{(k)}\mathbf{z}_0'^{(k)},$$

   where $\mathbf{U}^{(k)}$ is the $\mathbf{U}$ transform on outer iteration k.

8. Update the current guess

$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \mathbf{x}_0'^{(k)}.$$

9. Repeat outer loop (steps 2 to 8) until desired convergence is reached.

In the above algorithm we use the T- and the U-transform on each outer iteration (steps 5 and 7). We can avoid using the T-transform provided the control variable transform is *static*, that is to say, the transform does not depend on the current linearisation state $\mathbf{x}_i^{(k)}$. If this is the case then the increment $\mathbf{z}_0'^{(k+1)}$ can be calculated directly from the values of control variables increments on the previous outer loop, without the need for the T-transform.
To demonstrate this we start by assuming that we have $\mathbf{z}_0'^{(k)}$ and $\mathbf{z}_b'^{(k)}$ and want to find $\mathbf{z}_b'^{(k+1)}$. We also know that

$$\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \mathbf{x}_0'^{(k)} = \mathbf{x}_0^{(k)} + \mathbf{U}^{(k)}\mathbf{z}_0'^{(k)} \tag{3.11}$$

and

$$\mathbf{x}_b'^{(k)} = \mathbf{x}^b - \mathbf{x}_0^{(k)} = \mathbf{U}^{(k)}\mathbf{z}_b'^{(k)}, \tag{3.12}$$

since $\mathbf{z}_b'^{(k)} = \mathbf{T}^{(k)}\mathbf{x}_b'^{(k)}$. Then the background increment for the next iteration is given by

$$\begin{aligned}
\mathbf{z}_b'^{(k+1)} &= \mathbf{T}^{(k+1)}\left(\mathbf{x}^b - \mathbf{x}_0^{(k+1)}\right) \\
&= \mathbf{T}^{(k+1)}\left(\mathbf{x}^b - \mathbf{x}_0^{(k)} - \mathbf{U}^{(k)}\mathbf{z}_0'^{(k)}\right) && \text{using equation (3.11)} \\
&= \mathbf{T}^{(k+1)}\left(\mathbf{U}^{(k)}\mathbf{z}_b'^{(k)} - \mathbf{U}^{(k)}\mathbf{z}_0'^{(k)}\right) && \text{using equation (3.12)} \\
&= \mathbf{T}^{(k+1)}\left(\mathbf{U}^{(k)}\mathbf{z}_b'^{(k)}\right) - \mathbf{T}^{(k+1)}\left(\mathbf{U}^{(k)}\mathbf{z}_0'^{(k)}\right) \\
&= \mathbf{z}_b'^{(k)} - \mathbf{z}_0'^{(k)} \quad \text{provided} \quad \mathbf{T}^{(k+1)}\mathbf{U}^{(k)} = \mathbf{I}.
\end{aligned}$$

In the case of a static control variable transform, which does not depend on the linearisation state,

$$\mathbf{T}^{(k+1)} = \mathbf{T}^{(k)} = \ldots = \mathbf{T}^{(0)} = \mathbf{T}$$

$$\mathbf{U}^{(k+1)} = \mathbf{U}^{(k)} = \ldots = \mathbf{U}^{(0)} = \mathbf{U}$$

and so

$$\mathbf{T}^{(k+1)}\mathbf{U}^{(k)} = \mathbf{T}\mathbf{U} = \mathbf{I},$$

eliminating the need for the T-transform in the incremental 4D VAR algorithm. We now describe the details of the inner loop minimisation.

3.2.2 The Inner Loop

The inner loop minimisation, step 6, can be solved by various methods including quasi-Newton and conjugate gradient methods. In most NWP applications conjugate gradient methods are used since they are the best compromise between convergence properties and memory requirements for high dimensional problems [4]. For these reasons we also choose a conjugate gradient method for the inner loop minimisation. Various conjugate gradient methods are compared in [43], of these the Beale restarted memoryless quasi-Newton method of [58] performed best. The routine is implemented in the CONMIN package [59] and has been used successfully in [54].

The routine requires the calculation of the cost function (3.9) and its gradient with respect to control variables at time $t = 0$. The gradient of the cost function with respect to the control variables at the start of the data assimilation window is given by

$$\nabla \widetilde{J}^{(k)}[\mathbf{z}_0'^{(k)}] = \mathbf{\Lambda}^{-1}(\mathbf{z}_0'^{(k)} - \mathbf{z}'^{b})$$
$$+\mathbf{U}^T \sum_{i=0}^{n} \mathbf{M}_1^T \ldots \mathbf{M}_i^T \mathbf{H}_i^T R_i^{-1}(\mathbf{H}_i(\mathbf{M}_i \mathbf{U}\mathbf{z}_0'^{(k)}) - \mathbf{d}_i), \tag{3.13}$$

where $\mathbf{M}_i^T$ is the adjoint of the linear model $\mathbf{M}(t_i, t_0, \mathbf{x}^{(k)})$ in (3.9) and $\mathbf{d}_i$ is the *innovation* vector at time $t = t_i$ with $\mathbf{d}_i = \mathbf{y}_i^o - \mathcal{H}_i[\mathbf{x}_i^{(k)}]$, as defined previously. The adjoint model $\mathbf{M}_i^T$ propagates gradient information from time t_i to time t_{i-1}, allowing us to compute the gradient with respect to the control variables at the initial time.

The inner loop cost function is found by calculating (3.9), which is written in terms of control variables. The background penalty (the first term) is trivial to

compute in these variables, but the observation term (the second term) requires use of the U-transform and the linear model, as the observation operators act on model-space increments at the appropriate times.

The gradient of the incremental cost function at time t_0, equations (3.13) is then calculated using the adjoint model. The gradient at time t_i is propagated backwards to time t_{i-1}, then the gradient contribution at t_{i-1} is calculated, and the process is repeated until time t_0. This gives the value of the gradient with respect to *model variables* at time $t = 0$. We then apply the adjoint of the U-transform, $\mathbf{U}^T$, to obtain the gradient with respect to control variables. We then include the background gradient.

The U-transform and its adjoint are used once every inner iteration and so must be computationally efficient. One way that it is possible to save resources in the incremental 4D VAR is to perform the inner loop at a lower resolution than the outer loop. In fact the inner loop minimisation in the Met Office and the ECMWF's data assimilation system is performed at a lower resolution [53].

3.2.3 Summary

In this section we introduced the concepts of data assimilation. We then focused our attention on one of the most advanced data assimilation methods, 4D VAR. We discussed the practical issues surrounding the implementation of the method. 4D VAR is actually solved in an incremental form, but even with this simplification we still need to model the background error covariance matrix. This is done using the control variable transform. The transformation to a new set of variables simplifies the problem by assuming the new variables are uncorrelated. We then described the incremental 4D VAR algorithm. From this it is clear that the U-transform must be computationally efficient. This is because we actually require the U-transform and its adjoint once every inner iteration and once every outer loop. The T-transform is only required at the beginning of each outer loop and, in fact, we demonstrated that if the transforms are static then background

increments can be calculated from the previous increments. Therefore, for a static transform, the T-transform is not required at all in the algorithm. We note, however, that the T-transform is still needed when calculating background error statistics for the control variables. This is discussed in section 3.4.
In this section the control variable transform was introduced very generally. However, the success of the data assimilation is highly dependent of the choice of the control variables since they are assumed to be uncorrelated. In the following section we discuss the possible choices of control variables.

3.3 Choice of Control Variable

So far we have introduced the theory of the control variable transform but have not discussed the basis on which the transform can be constructed. It is clear that a good choice of control variables is essential to ensure that errors in these variables are uncorrelated, or nearly uncorrelated. The importance is highlighted when only a single observation, located at a point j, is assimilated at time $t = 0s$. In this simple scenario it can be shown that the analysis increment

$$\mathbf{x}^a - \mathbf{x}^b \propto \mathbf{B}\mathbf{e}_j^T = \mathbf{U}\Lambda\mathbf{U}^T\mathbf{e}_j^T$$

where $\mathbf{x}^a$ is the model state at time $t = 0$ that minimises the cost function (3.1) and $\mathbf{e}_j^T$ is the jth unit vector. The quantity $\mathbf{B}\mathbf{e}_j^T$ is the jth column of the background error covariance matrix $\mathbf{B}$ sometimes refered to as a *structure function*. This method is used later to test the control variable transforms.
We use dynamical properties to attempt to identify possible control variables. In section 2.3.2 two kinds of atmospheric motion were identified as normal modes; one of these is unbalanced, the inertial-gravity waves, the other, geostrophically balanced, the Rossby waves. It is thought that a good choice of control variables will involve capturing the balanced and unbalanced modes in separate control variables since we assume that there is little or no interaction between these flows. In the linear case the modes evolve independently and therefore they would be an

obvious choice of control variable. In the non-linear case the degree of this interaction may depend in some way on the degree of the non-linearity.
It is also desirable for the data assimilation to generate increments that in some way respect important dynamical properties of the system, an example of this being the balance relationship. Introducing these relationships in the assimilation can be achieved through the control variable transform.
We may therefore think of the control variable transform as having a dual purpose:

1. It is a necessity required to simplify the problem.

2. It is a means to introduce important dynamical relationships into the data assimilation.

With this in mind we start this section describing the modelling of the background error covariance matrix. We make the distinction between the univariate auto-covariance and multi-variate cross-covariance of the control variables background errors; the latter of these of particular interest in this research. The current choice of control variables that are used operationally and their shortcomings are also discussed.

3.3.1 Modelling the Background Error Covariance Matrix

The background error covariance matrix is vital to the success of the data assimilation [20]. Ironically, we will never know the true background errors since we never have access to the truth. Even if we did, we would still be forced to model them since the matrix is too large to be stored in memory. Given this impossible task we must, however, attempt to represent the background errors realistically in the data assimilation.
Background errors are usually split into two contributions; block diagonal elements that represent the auto-covariance between grid points corresponding to particular control variables (the univariate component of $\mathbf{B}$), and cross-covariance between

grid points corresponding to different control variables (the multivariate component of $\mathbf{B}$) [64]. Operationally the control variable transform is used to remove both the cross-covariance and the auto-covariance [16], [40]. The theory we presented in section 3.2 describes a transform that only removes the cross-covariance. The removal of the covariance is generally made through a series of transformations. At the Met Office this is written as [62]

$$\mathbf{z} = \mathbf{T}_h \mathbf{T}_v \mathbf{T} \mathbf{x} \tag{3.14}$$

$$\mathbf{x} = \mathbf{U} \mathbf{U}_v \mathbf{U}_h \mathbf{z}, \tag{3.15}$$

The transformations $\mathbf{T}$ and $\mathbf{U}$ are referred to as the *parameter transforms* and are assumed to remove the cross-covariance of the variables, this theory was introduced in section 3.2. Subsequent transforms $\mathbf{T}_h$ and $\mathbf{T}_v$ model the horizontal and vertical auto-covariance. The Met Office's covariance model is discussed in detail in [30]. Applying this sequence of transforms results is a transformed background error covariance matrix equal to the identity matrix. We note that in this work we are concerned only with the parameter transform and the validity of the assumption that the cross-covariance is removed. The transform at the Met Office therefore implies a $\mathbf{B}$ matrix of the form

$$\mathbf{B} = \mathbf{U} \mathbf{U}_v \mathbf{U}_h \mathbf{U}_h^T \mathbf{U}_v^T \mathbf{U}^T, \tag{3.16}$$

where we assume that the errors in the control variables are completely uncorrelated, i.e. $\mathbf{U}^T \mathbf{B}^{-1} \mathbf{U} = \mathbf{I}$. The transformations have simplified the data assimilation as the size of the problem is greatly reduced. By using the control variable transform it is not necessary to store the covariance information explicitly since it is implicit in the definition of $\mathbf{U}$.

We now discuss how the parameter transform has been derived in operational centres around the world.

3.3.2 The Parameter Transform In Some Leading Assimilation Schemes

In [13] Daley states that *the key to successful implementation of 3D VAR is an accurate specification of the multivariate forecast error covariance using appropriate linear relationships*. Here Daley specifically identifies the parameter transform as the vital component in the data assimilation. Much effort and research is focused on the generation and modelling auto-covariance and the representation of the cross-covariance by the parameter transform.
A separation into so-called balanced and unbalanced control variables is usually made and it is then assumed that these control variables are uncorrelated [47]. At present the control variable transforms are often based on the linear balance equation (given in section 2.4.1 for the SWEs) and assume that either the height or the vorticity is a totally balanced variable. That is to say, the variables are assumed to satisfy the linear balance equation, equation (2.89), exactly. This definition of balance is at best approximate and only valid in specific dynamical regimes. So, whilst relatively sophisticated modelling of auto-covariance is carried out, the fundamental assumption regarding the cross-covariance of control variables is potentially flawed. The shortcomings of these control variable transforms can be summarised as follows:

1. Height or vorticity is taken to be the totally balanced variable, which is a poor assumption in some regimes.

2. Linear balance is enforced for the 'balanced' variable and this is not always appropriate.

3. Sophisticated treatment of auto-covariance is made without giving much attention to point 1.

The use of a control variable transform in variational data assimilation was first introduced in [47]. Here the linear balance relationship is used partition the mass variable into balanced and unbalanced components. The balanced and unbalanced

variables are defined using the linear balance relationship in terms of the relative vorticity. The control variables are taken to be the vorticity, divergence and the unbalanced mass variable (as well as specific humidity). The implication is that the vorticity is a totally balanced variable and the divergence is totally unbalanced. It is then assumed that the balanced and unbalanced variables are uncorrelated. The control variable transform described in [47] is essentially vorticity-based, that is to say, the vorticity control variable is assumed to be totally balanced and satisfies the linear balance equation exactly. The vorticity-based control variable transform of [47] has subsequently been implemented in several operational data assimilation systems. These include the Met Office, the ECMWF, the Canadian Meteorological Centre's 3D VAR system [22] and in the CIRA/CSU 4D VAR system [70]. We now discuss how the parameter transform is defined in these leading assimilation schemes.

The Met Office

A very similar implementation of the control variable transform is operational at the Met Office [40], [41]. In the current Met Office data assimilation system the control variables are the stream function ψ, the velocity potential χ and an 'unbalanced', residual pressure P_u (as well as the relative humidity to account for moisture) [40]. Although the vorticity and the divergence are not used directly as a control variable (due to conditioning) the problem solved is the same and fundamental assumptions of [47] have been made; the vorticity is totally balanced and the divergence totally unbalanced. Then, using the linear balance equation a balanced pressure field is calculated and subtracted from the full pressure field to obtain the residual unbalanced pressure, in the same way as [47].

ECMWF

One of the most sophisticated control variable transforms is used in the ECMWF data assimilation system. Here assumed correlations between each variable are systematically removed [16]. The control variables are the vorticity, the

unbalanced divergence, unbalanced temperature and pressure (as well as specific humidity). The balance relationships were initially determined using statistical regression assuming the same structure as the linear balance equation. In [16] the statistical relationship is compared to that of the analytical linear balance equation and found to be very similar. Recently improvements to the balance operator have been made to include the non-linear balance equation and the quasi-geostrophic omega equation [20]. These equations are linearised about the background state thus keeping the inner loop minimisation quadratic and incorporating flow dependence in the transform. In addition to these improvements the transformation is more refined than those discussed previously since the divergence is not assumed to be totally unbalanced. The similarity, however, lies in the assumption that vorticity is totally balanced.

HIRLAM

One exception to the vorticity-based transform is used in the high resolution limited area model (HIRLAM) [25]. Here a balanced height / mass field and the ageostrophic unbalanced wind are used as a control variables. This is the opposite of the assumption made in the vorticity-based transform; in [25] the height, or mass, is considered to be totally balanced and a balanced geostrophic wind is defined using the linear balance equation. The balanced wind is subtracted from the full wind giving an ageostrophic wind. This method was chosen for simplicity in the limited area model. Using the ageostrophic wind has serious problems near the equator due division by f, which goes to zero at the equator, and neglecting the correlations between mass and ageostrophic wind [25]. Since the height is assumed to be totally balanced the control variables will only be a valid separation of balanced and unbalanced components in a low Burger regime. In a high Burger regime the height is not a good approximation to balance and therefore is not the appropriate balanced variable.

Ocean Data Assimilation

Historically ocean data assimilation systems generally assume the model variables are uncorrelated, for example [65]. Recently this deficiency has been addressed in [66] were a multivariate balance operator for ocean data assimilation is defined. This method uses a similar approach as [16]; one variable, the temperature, is taken in totality (in [16] the variable taken in totality is the vorticity) and assumed correlations are then systematically removed from one variable, and then the next using balance relationships. The linear balance relationship is used in the case of the horizontal velocities.

3.3.3 Summary

A defining factor of all the control variable transforms discussed is that they are based on the ideas of [47] and can be thought of as vorticity-based, or height/mass-based in the case of [25]. The assumption that the balanced and unbalanced components of the flow are uncorrelated forms the basis of these ideas. However, none of these transforms successfully capture the balanced flow in all regimes.

To attempt to address the problems above a PV-based control variable transform was recently developed in [11] that should be valid across all regimes. We now discuss the PV-based control variable transform.

3.3.4 The PV-based Transform

In [11] a new version of control variables is derived based on the PV, which characterises the balanced flow of the linearised system. The PV-based transform defines the control variables using the normal modes of the system. We note that if the system were linear the normal modes would be the ideal choice of control variable since they evolve independently. The balanced variable is the linearised PV q', as we found in section 2.3.3, the two unbalanced variables are the divergence and the geostrophic departure, which are linear combinations of the two

fast modes, which have zero PV. This is demonstrated in section 2.3.3.
It was found in [11] that using these variables directly resulted in serious numerical issues and so an equivalent version was tested using the balanced stream function, ψ_b, the unbalanced geopotential, ϕ_u and the divergence. We note that the rotational wind has been split here into balanced and unbalanced components, and not used in totality as in the vorticity-based control variable transform discussed previously. The divergence is not split in [11] but this can be added at a later stage. Unfortunately even with the revisions to the PV-based transform problems with the ECMWFs vertical discretisation were still encountered [11]. To deal with these issues it was necessary to make many compromises in the transform. The transform still performed satisfactorily; better forecasts measured against the analysis were obtained in the tropics and also positive improvements in the northern hemisphere. Negative aspects of the experiments could all be traced back to the numerical difficulties.
Another contribution to work on the PV-based transform has come from [67]. In [67] and [68] the PV-based transform is derived for the 2D SWEs. In this work the qualitative advantages of the transform are demonstrated. This is done by making a visual comparison of the vorticity and PV-based control variable fields in several regimes. The work demonstrated that the PV-based transform theoretically feasible and also that it may provide a better representation of balance than the vorticity-based transform. However, the PV-based transform was not implemented into a data assimilation scheme and the correlation of the PV-based variables was not investigated.
The results of both [11] and [67] are in agreement and give encouraging results regarding the PV-based transform. This, however, is not comprehensive since the following has not yet been addressed:

1. The PV-based control variables have not been conclusively tested in a data assimilation system that is manageable enough to be studied.

2. The assumption that the variables are uncorrelated has never been verified.

3. More fundamentally, the accuracy of the assumption that balanced and unbalanced flows are uncorrelated in the non-linear system has never been quantified.

With this in mind we now out line the aims of this project.

3.3.5 Project Aims

In this work we attempt to address the following questions:

1. Can we quantify the accuracy of the fundamental assumption that the balanced and unbalanced flows are uncorrelated?

   In section 3.3.2 we discussed the various methods of deriving the parameter transform. Each of the transforms only approximates the balanced and unbalanced modes of the linear system. In addition, they all are based on the assumption that the balanced and unbalanced flows are uncorrelated yet this remains to be validated. By testing the correlation of balanced and unbalanced PV-based variables, which are not approximate, we address this question.

2. How accurate is the current assumptions regarding the vorticity-based control variables?

   We know that there are potential problems with the vorticity-based control variables in low Burger regimes but we would like to quantify the impact of this on the correlations of the control variables.

3. Can we demonstrate that the PV-based variables give quantitatively better representation of the background statistics?

   The PV-based control variables must improve on the limitations of the vorticity-based transform. We therefore aim to quantify the difference.

4. How do the control variable transforms influence the analysis?

Improving the analysis is the ultimate goal of the PV-based transform. We thus aim to consider simple assimilation experiments using the vorticity and PV-based control variables. In this simple setting the analysis is relatively straightforward and much can be learnt regarding the impact of the transforms on the data assimilation. We aim to demonstrate the different influences of the vorticity and PV-based transforms on the analysis.

3.4 Auto-Correlations

Before data assimilation can be performed we must generate auto-covariances for each control variable. This involves first generating a population of model field perturbations by a choosen method, some of which are discussed in the following section. The population is then transformed to control variable perturbations using the appropriate T-transform, from which auto-covariances are computed. The auto-covariances represent the block diagonal components of the background error covariance matrix after only the parameter transform, discussed in section 3.3.2, has been applied. Operationally the auto-covariances must then be modelled by the control variable transform since they are too large to be used directly. In our system this is not the case.

Ideally the process of generating background error auto-covariances would be repeated before each assimilation. Operationally, however, it would not be practical and therefore a climatological approach is adopted; auto-covariances are generated in a calibration stage and then used in all subsequent assimilations. Zonally averaged quantities are used and the method produces a climatological representation of the background error auto-covariance. We choose to follow this approach in our work to reflect the implementation in a full operational system. We now discuss the methods for generating and modelling the auto-correlations.

3.4.1 Generating Auto-Correlations

In NWP we can never know the background error covariances since we never know the truth (if we knew the truth we would not need the data assimilation). Therefore we must find other ways to generate background statistics. There are two main approaches. One is to attempt to extract information regarding the background errors from innovations (differences between observations and the background state). The other method is to find a surrogate quantity that is thought to be similar to background errors but whose statistical characteristics are available [20].

The calculation of background covariances from innovations was used in [27]. This is done by assuming that observation and background errors are uncorrelated and further that the errors in the observations are spatially uncorrelated. These are relatively safe assumptions for certain types of observations. It is then possible to show that the covariances of the background errors are described by the innovation covariance. This method does have some practical problems. Firstly, background errors are generated in terms of observed quantities but in the 4D VAR cost function the background errors are required in terms of control variables. Further, the method can only be used in data-dense areas and only if observation errors are truly uncorrelated [19].

The second approach involves finding an alternative quantity that is thought to have similar statistics to background errors. There are several advanced methods that use these ideas; the so-called NMC method, the analysis ensemble method (AEM) and the quick cov's method (QCM).

The NMC method used in [47] assumes that background errors are similar to forecast errors. More specifically the method assumes that the spatial correlations of background errors are similar to the correlations of differences between 24 hour and 48 hour forecasts *valid at the same time.* Forecast times are taken 24 hours apart to remove the diurnal signal, which would otherwise dominate the statistics. This method is convenient to implement in an operational centre since the forecast

data required can be obtained from historical archives. However, the method does not perform well downstream of data sparse areas since the 24 hour forecast will be very similar to the 48 hour forecast. Also, the time interval of 24 hours is not appropriate since the background field is usually a 6 hour forecast. Therefore we might expect that any data obtained from the NMC method to be much more balanced than is actually the case for the background field, which has only 6 hours to geostrophically adjust before it is used [19].

Background statistics at the ECMWF are generated using the AEM [20]. Here the data assimilation is run for many cycles in each of which random perturbations are added to the observations. A set of perturbed background fields is then obtained and it is assumed that background errors are similar to the differences between pairs of the perturbed background fields. Again this method also has some drawbacks: It assumes that observational error statistics are accurately known. There is a potential of feedback since the data assimilation is cycled. Also, the statistics obtained are highly dependent on the current analysis and the system itself. That said, the AEM is considered to produce the best results [20].

It is worth noting that the two most advanced methods, the NMC and the AEM, both require an operational data assimilation system. This creates a paradox: We require statistical knowledge to set up the data assimilation, yet the most advanced methods require an operational system to generate the statistical data. Therefore, when implementing a data assimilation system it is necessary to use another method for initial generation of covariances. These must then be optimised using a more sophisticated method.

A method that has been used for generating initial statistical data is based on using forecast differences as an alternative quantity. In this sense it is similar to the NMC since it uses forecast differences as a proxy for background errors. We refer to this method as the quick-covs method (QCM). It was suggested in [51] as an initial step to setting up the data assimilation. Here we generate a single long forecast using our model. We then take forecast differences at regular intervals of time apart. This is done until we have a data set of multiple time differences.

Using these differences we may look at the correlations of the difference fields. We can examine the correlation between control variables by transforming the forecast time-differences using the relevant T transform. The QCM used in [51] takes 6 hour time-differences since the background state is a 6 hour forecast. The difference fields are then adjusted to remove the diurnal signal. The QCM method is used later in this work to test the correlation of the control variables.

3.4.2 Modelling The Auto-Covariance Matrices

Once the auto-covariances have been found they need to be modelled in the B matrix. This can be done in many ways, for example using diffusion operators in [64].

In the cost function it is the inverse matrix that appears and so modelling the auto-correlations directly would mean inverting a very large matrix. It is, however, possible to model the inverse matrix directly. In this work we use the method outlined in [31] following work in [2]. Here we directly define $\mathbf{B}^{-1}$ in such a way that $\mathbf{B}$ comprises of approximate Gaussian shaped structure functions with homogeneous correlation length L. The Gaussian correlation function, ρ_g, is defined by

$$\rho_g(x) = e^{-x^2/2L^2}. \tag{3.17}$$

The correlation function ρ_L for the inverse Laplace smoother is given by

$$\rho_L = \omega_0 + \omega_1 \left(L_{xx}\right)^2, \tag{3.18}$$

where L_{xx} is a second derivative matrix with periodic boundary conditions and coefficients $\omega_0 = 1/L$ and $\omega_1 = \omega_0(L^4/2)$ are chosen such that the inverse function approximates a Gaussian distribution with length scale L [31].

If we generate the $\mathbf{B}^{-1}$ matrix given by (3.18) and invert it we obtain the modelled $\mathbf{B}$ matrix. As a comparison we plot in figure 3.2 a column of the modelled $\mathbf{B}$ with the Gaussian function. Here we see the functions are very similar except at the start of the tails were the Laplace smoothing function is negative. This appears to

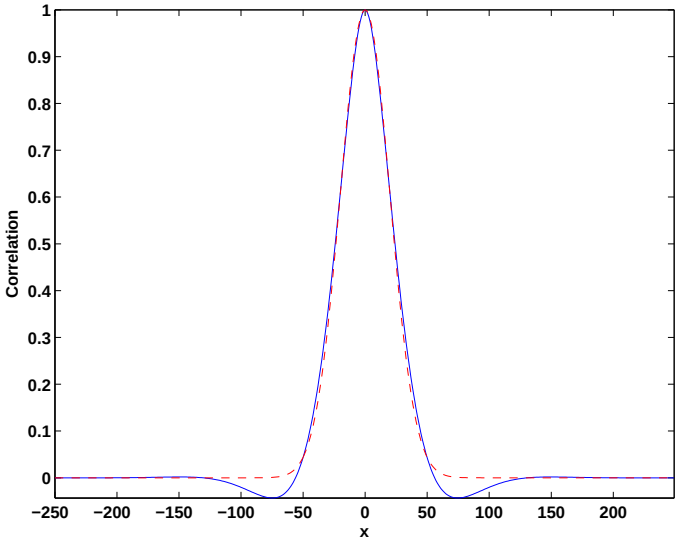

Figure 3.2: Plot shows Gaussian function (red dashed line) and Laplace-based smoothing correlation function (blue solid line) for a length scale $L = 20$.

be a good enough approximation for our purpose, which is to represent a spatial length scale for each control variable. We choose this method to model the auto-covariances generated later in this work.

3.5 Summary

In this chapter we introduced the concepts of data assimilation and focused on 4D VAR. We highlighted key issues evolving the implementation of such a method, specifically how the control variable transform is used to model background error covariances. We explained in detail the steps in the incremental 4D VAR algorithm. The current issues concerning the parameter transform were presented. We identified a deficiency in the definition of the balanced variables in all the current operational control variables; all current control variables are either vorticity-based or height/mass-based. These definitions of balance are not appropriate in all Burger regimes and therefore can potentially violate the assumption that they are uncorrelated. A PV-based control variable transform is then introduced, which addresses these issues. The current results relating to the

PV-based transform were discussed and suggest that there is potential for the transform to address the limitations of the current implementations, but this has not been fully examined in a data assimilation system.

In the final section we discussed how auto-covariances are generated operationally. We outlined the variety of methods available to generate auto-covariances and also a method that can be used to model them in the $\mathbf{B}$ matrix.

We now present the numerical model we will be using throughout the remainder of this work.

Chapter 4

Numerical Models

We now introduce the model used for the remainder of this research. We choose to use a simplified one-dimensional form of the shallow water equations (SWEs) that were introduced in section 2.1. These equations have the same key properties as the two-dimensional equations but in one-dimensional form.
The continuous model is then discretised using a semi-implicit, semi-Lagrangian scheme. We present the discrete equations and discuss their implementation. We highlight several aspects of the numerical model that require extra attention when deriving the linear models necessary for the incremental 4D VAR. The discrete model is then tested and we demonstrate some of the dynamical properties of the model that have been shown analytically in section 2.1. These include conservation of potential vorticity (PV) and geostrophic adjustment. We also demonstrate the changes in behaviour of the model as we change dynamical regimes.
Next the tangent linear model (TLM) and the adjoint model (AM) are derived from the non-linear model code. We discuss the various methods for deriving these models and then outline the method used. We now introduce the continuous model used for the remainder of this research.

4.1 1D Shallow Water Equations

The two-dimensional SWEs were introduced in section 2.1 and are given by equations (2.1), (2.2) and (2.3). We now simplify equations (2.1), (2.2) and (2.3) by assuming that velocities u and v and the depth h do not vary in the y direction. We write the velocity in the x direction as

$$U_c + u$$

where U_c is a constant forcing *mean* flow. We then solve for u, the departure from this constant flow. The 1D SWE model is then given by

$$\frac{\partial u}{\partial t} + (U_c + u)\frac{\partial u}{\partial x} - fv = -g\frac{\partial(h + \widetilde{H})}{\partial x}, \tag{4.1}$$

$$\frac{\partial v}{\partial t} + (U_c + u)\frac{\partial v}{\partial x} + fu = 0, \tag{4.2}$$

$$\frac{\partial h}{\partial t} + \frac{\partial h(U_c + u)}{\partial x} = 0. \tag{4.3}$$

Here $\widetilde{H} = \widetilde{H}(x)$ is the height of the orography, f the Coriolis parameter, taken to be constant, and g is the gravitational force. The model domain is periodic in the x direction. In equation 4.2 $\frac{\partial h}{\partial y}$ is choosen to balance the fU_c, which is why these terms are not present in the equation.

This special case of the SWEs has been applied to various mesoscale phenomena in the atmosphere for example flow over mountains [29]. This model was also used in [46] to examine the accuracy of a conservative finite difference scheme in capturing the formation of hydraulic jumps.

The model is chosen as it retains the key properties of the 2D SWEs whilst being significantly simpler to develop. We have a non-trivial first order balance relationship

$$fv - g\frac{\partial(h + \widetilde{H})}{\partial x} = 0. \tag{4.4}$$

This relationship is found through an asymptotic expansion in small Rossby number as in section 2.2. The balance equation is a fundamental component of both the vorticity and the PV-based transforms and is applied to increments in a

linearised form. We let $u(x,t) = \bar{u}(x,t) + u'(x,t)$, $v(x,t) = \bar{v}(x,t) + v'(x,t)$ and $h(x,t) = \bar{h}(x,t) + h'(x,t)$, where $\bar{u}(x,t)$, $\bar{v}(x,t)$ and $\bar{h}(x,t)$ are reference states. If we assume that the reference states satisfy the balance equation (4.4) to first order accuracy, then we obtain a corresponding first-order linear balance equation for the increments, given by

$$fv' - g\frac{\partial h'}{\partial x} = 0. \tag{4.5}$$

This can be written in terms of the streamfunction ψ and the geopotential ϕ as in equation (2.89), i.e.

$$f\nabla^2\psi' = \nabla^2\phi', \tag{4.6}$$

where the 1D streamfunction is given by

$$\nabla^2\psi' = \frac{\partial v'}{\partial x}. \tag{4.7}$$

The potential vorticity (PV) is conserved in a one-dimensional form of equation (2.10), i.e.

$$q = \frac{1}{h}\left(f + \zeta\right), \tag{4.8}$$

where the one-dimensional vorticity ζ is given by

$$\zeta = \frac{\partial v}{\partial x}. \tag{4.9}$$

It can also be shown, as in section 2.3 for the 2D case, that the 1D SWEs linearised about a simple reference state have three normal modes. The slow, or balanced mode again satisfies linear balance and is characterised by the linearised PV. The remaining two fast modes can be related to the geostrophic departure, ${}_a\zeta'$, defined by

$${}_a\zeta' = f\frac{\partial v'}{\partial x} - g\frac{\partial^2 h'}{\partial x^2}, \tag{4.10}$$

as in equation (2.87) for the 2D SWEs, and the divergence

$$\mathcal{D}' = \frac{\partial u'}{\partial x}, \tag{4.11}$$

in the same way as for the 2D equations shown in section 2.3. We now derive a discrete form of equations (4.1), (4.2) and (4.3). We use a semi-implicit, semi-Lagrangian scheme to solve the system.

4.2 The Discrete Model

A discrete form of the 1D SWEs is now derived using a two-time-level semi-implicit, semi-Lagrangian (SISL) scheme. The SISL scheme is chosen since it is used widely in NWP. This type of scheme is implemented at both the Met Office [14] and the ECMWF [28]. We desire that in our simplified model we have similar numerical properties to the models at these operational centres.
In the following section a brief description of the SISL scheme is given. This scheme is then applied to our 1D SWEs and the discrete equations derived. The model is an extension of work in [34] and [35].
After the non-linear model has been tested we turn our attention to deriving the Tangent Linear Model (TLM) and the Adjoint Model. These models are required for use in the incremental 4D VAR that was introduced in section 3.1.2.

4.2.1 The Semi-Lagrangian Scheme

Historically, early numerical weather prediction (NWP) models were based on Eulerian finite difference schemes. Here the time-step size Δt is limited by the speed of the fastest wave through the Courant-Friedrichs-Lewy, or CFL, condition given by

$$0 \leq \mu \leq 1 \quad \text{where} \quad \mu = c\Delta t/\Delta x, \tag{4.12}$$

where Δx is the spatial step size. The parameter μ is the *Courant number* and c is the speed of the fastest wave. The CFL condition (4.12) is a necessary stability condition for Eulerian finite difference schemes [17]. A Lagrangian scheme, on the other hand, does not have these restrictions but the grid will tend to become irregular and therefore some areas may be poorly resolved [60]. The semi-Lagrangian scheme gets the best of both worlds by taking a Lagrangian time step onto a new *regular* grid at the next time level. At each time step it is necessary to find where the points on the new regular grid (the *arrival points*) came from (the *departure points*).
The popularity of the SISL scheme in NWP is due to the fact that there is no CFL

stability restriction linked to the speed of the fastest wave and also removes stability restrictions due to advection speeds. Therefore increased time-steps can be achieved without increasing the spatial resolution [60]. This is particularly useful for latitude / longitude grids where the grid size decreases moving pole-ward.

4.2.2 The Discrete Model Equations

We now apply the scheme outlined in the previous section to our 1D SWEs. Firstly the 1D SWEs, equations (4.1) to (4.3), are written in terms of their full Lagrangian derivatives as follows,

$$\frac{Du}{Dt} + \phi_x + g\widetilde{H}_x - fv = 0, \tag{4.13}$$

$$\frac{Dv}{Dt} + fu = 0, \tag{4.14}$$

$$\frac{D\ln\phi}{Dt} + u_x = 0, \tag{4.15}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + (U_c + u)\frac{\partial}{\partial x} \tag{4.16}$$

and $\phi = gh$. This form of the equations is chosen as it is a more convenient when applying the SISL scheme. Also this formulation avoids growth in computational noise due to using time-extrapolated quantities when calculating the $u\frac{\partial h}{\partial x}$ term in (4.3) [60].

Applying a two-time-level semi-implicit, semi-Lagrangian scheme to the equations above gives the following time-discrete equations

$$\frac{u_a^{n+1} - u_d^n}{\Delta t} + \alpha_1 \left[\phi_x + g\widetilde{H}_x - fv\right]_a^{n+1} + (1-\alpha_1)\left[\phi_x + g\widetilde{H}_x - fv\right]_d^n = 0, \tag{4.17}$$

$$\frac{v_a^{n+1} - v_d^n}{\Delta t} + \alpha_2 \left[fu\right]_a^{n+1} + (1-\alpha_2)\left[fu\right]_d^n = 0, \tag{4.18}$$

$$\frac{\ln\phi_a^{n+1} - \ln\phi_d^n}{\Delta t} + \alpha_3 \left[u_x\right]_a^{n+1} + (1-\alpha_3)\left[u_x\right]_d^n = 0, \tag{4.19}$$

where the subscript x denotes the derivative with respect to x, superscript n is the value at time level n and Δt is the time step. The constants $\alpha_1, \alpha_2, \alpha_3$ are chosen

to meet the stability requirements of the scheme. This is discussed in section 4.2.3. The advection terms are time differences along trajectories and *all* other terms are time averages along trajectories. The Coriolis terms are treated in this way as this avoids the introduction of instability due to using extrapolated values to evaluate the Coriolis terms [60]. The subscripts a and d represent the arrival and departure points.
In the following sections we will discuss the properties of the numerical scheme and describe in more detail how the model is solved.

The Solution Algorithm

Equations (4.17) to (4.19) are solved on a staggered grid which is a 1D analogue of the Arakawa B-grid. The grid has ϕ on x_i points and u and v on $x_{i+1/2}$ points. The staggered grid is chosen due to prevent uncoupling of neighbouring points which can occur using a standard unstaggered grid. This uncoupling of neighbouring points leads to poor conservational properties [17], [33].
We now outline how equations (4.17) to (4.19) are solved. The method proceeds as follows:

1. Find the departure points $x_d = x_i - a_i$.

   We find the departure points $x_d = x_i - a_i$ at time t_n by calculating the displacement a_i using an iterative procedure defined in [56]

   $$a_m^{(k+1)} = \Delta t u(x_m - a_m^{(k)}, t_n) + \Delta t U_c, \tag{4.20}$$

   with $a^{(0)} = 0$. The velocity at the mid-point of the trajectory is found by extrapolation formula,

   $$u(x_i, \Delta t_{n+1/2}) = \frac{3}{2} u(x_i, t_n) - \frac{1}{2} u(x_i, \Delta t_{n-1}).$$

   The value of u at ϕ points is the average of the two neighbouring u values and the value of u at $x_i - a_i^{(k)}/2$ is found by linear interpolation. It is worth noting that the INT Fortran function is used in this code. The function takes

the integer value of a real number. This function is non-differentiable and therefore will have to be handled carefully when developing the Tangent Linear Model, this will be addressed later in the report. In [60] it is found that there is no value in performing this iteration more than twice. A sufficient condition for the convergence of this iteration is given by

$$\Delta t < [max(|u_x|, |u_y|, |v_x|, |v_y|]^{-1}, \tag{4.21}$$

for 2D flow [60]. Thus the maximum time step size is restricted by properties of the flow. For our 1D model this condition can be written as

$$max\,|u_x|\,\Delta t = max\,|u_{i+1} - u_i|\,\frac{\Delta t}{\Delta x} < 1. \tag{4.22}$$

We will show later that the 1D SWEs permit jumps in certain dynamical regimes and hence the condition may be violated. We can limit the potential for breach of this condition by our choice of Δt and Δx.

2. Find values of variables at departure points.

   We can assume that we know all values at all grid points at time n and therefore by interpolating to the departure points we may find all the terms at t_n in equations (4.17) to (4.19). We then write the equations with all known terms on the right hand side and unknown terms on the left,

$$\begin{aligned} u_a^{n+1} + \Delta t\alpha_1 \left[\phi_x + g\widetilde{H}_x - fv\right]_a^{n+1} &= u_d^n \\ -\Delta t(1-\alpha_1)\left[\phi_x + g\widetilde{H}_x - fv\right]_d^n & \end{aligned} \tag{4.23}$$

$$v_a^{n+1} + \Delta t\alpha_2\,[fu]_a^{n+1} = v_d^n - \Delta t(1-\alpha_2)\,[fu]_d^n \tag{4.24}$$

$$\begin{aligned} \ln\phi_a^{n+1} + \Delta t\alpha_3\,[u_x]_a^{n+1} &= \ln\phi_d^n \\ &- \Delta t(1-\alpha_3)\,[u_x]_d^n\,. \end{aligned} \tag{4.25}$$

   The right hand sides of each of the above equations are evaluated at each grid point using a centred difference for the spatial derivatives with,

$$\left(\frac{\partial\phi}{\partial x}\right)_u = \frac{\phi_{i+1} - \phi_i}{\Delta x},$$

and

$$\left(\frac{\partial u}{\partial x}\right)_{\phi} = \frac{u_{i+1} - u_i}{\Delta x},$$

with the subscript u and ϕ meaning evaluated at a u or ϕ point. Then a cubic-Lagrange interpolation formula is used to interpolate the values of the right hand side of (4.17) to (4.19) to the departure points. This type of interpolation is used as it a good compromise between cost and accuracy [60].

3. Solve elliptic equation for ϕ.

   We then can eliminate u and v and solve an elliptic equation for ϕ of the form,

$$-c\phi_{i+1}^{n+1} + 2c\phi_i^{n+1} - c\phi_{i-1}^{n+1} + \ln(\phi)_i^{n+1} = R_i, \tag{4.26}$$

   where c is a constant and R_i consists of known terms that are computed at the departure points. Equation (4.26) could be solved iteratively but since the tri-diagonal matrix formed by the left hand side is not strictly diagonally dominant and is singular the iteration is not well defined [34]. Therefore the equation is linearised by putting

$$\phi_i^{n+1} = \phi_i^n + \phi_i'.$$

   We then add ϕ_i'/ϕ_i^n to both sides. The iteration then becomes

$$\begin{aligned} &-c\phi_{i+1}'^{(m+1)} + (2c + \frac{1}{\phi_i^n})\phi_i'^{(m+1)} - c\phi_{i-1}'^{(m+1)} \\ &= R_i - \ln(\phi_i^n + \phi_i'^{(m)}) + \frac{\phi_i'^{(m)}}{\phi_i^n} + c\phi_{i+1}^n - 2c\phi_i^n + c\phi_{i-1}^n, \end{aligned} \tag{4.27}$$

   where m is the iteration count and $\phi_i'^{(0)} = 0$. At each iteration a direct solution is found. This is a fixed point iteration and will converge to the correct solution if the sufficient condition

$$|\phi_i'| < \frac{\phi_i^n}{2}, \tag{4.28}$$

   is met [34].

4. Find u and v at next time level.

   Now that we have ϕ^{n+1} we are able to find u^{n+1} and v^{n+1}, hence completing one time step.

4.2.3 Stability and Accuracy

We now briefly discuss time accuracy and stability of the scheme. We note that good spatial accuracy can be gained by good choice of interpolation.
It has been shown that a choice of $\alpha_i = 0.5$, $i = 1, 2, 3$ in the discrete equations (4.17) to (4.19) gives a second order accurate scheme in time [60]. For $\alpha < 0.5$, however, the scheme becomes unstable. Therefore to avoid any instability due to machine round-off $\alpha_i's$ are chosen to be a little above 0.5 as this will guarantee stability and give close to second order accuracy [55]. In our model we choose $\alpha_i = 0.6$ with $i = 1, 2, 3$.

4.3 The Behaviour of the Model

The model was implemented as outlined above and tested extensively. In the next section we demonstrate some of the dynamical properties shown analytically in chapter 2. First we describe the tests that were used to validate the model code.

4.3.1 Testing the Model

Our model is an extension of the model used in [34] and [35], where a 1D Shallow Water model with no rotation is considered, which does not include the velocity v. As an initial check forecasts from each of these models are compared under specific conditions. By setting

$$f = 0, \tag{4.29}$$

$$U_c = 0, \tag{4.30}$$

$$v_0 = 0, \tag{4.31}$$

in our model, where v_0 is the initial condition for v and U_c is the forcing mean flow, the models are identical. The two model forecasts were compared and the matched to the tolerance set in the elliptic solver; the maximum difference in the forecast fields produced by each model was found to be $O(10^{-10})$ indicating that the model code is correct in this case.

The next stage of testing involved comparing the results of [46] to those of our model. These tests involved running the model with:

1. Rotation, but without orography.
2. Orography, but without rotation.
3. Orography and rotation.

The results are in agreement with those of [46] taking into consideration the fact the SISL scheme used in our model is not capable of representing the formation of jumps as well as the numerical schemes used in [46].

Further testing is carried out by considering the dynamical properties of our model introduced in chapter 2. We start with the conservation of potential vorticity (PV).

4.3.2 Conservation Of Potential Vorticity

In section 2.1.1 we demonstrated that the PV is conserved in the 2D SWEs. In our model the 1D form of the PV is conserved, given by

$$q = \frac{1}{h}(\partial v/\partial x + f).$$

The next test examines how well our model actually conserves this quantity. Since the SISL numerical scheme has not been specifically designed to conserve the PV we would not expect this to be exact.

A plot of the q is shown in figure 4.1 for a long model integration. We see that the initial PV, the top plot, is advected by the flow from west to east. Initially the maximum PV is over the central orography, which is given by

$$\widetilde{H}(x) = H_c\left(1 - \frac{x^2}{a^2}\right) \quad \text{for} \quad 0 \leq |x| \leq a, \tag{4.32}$$

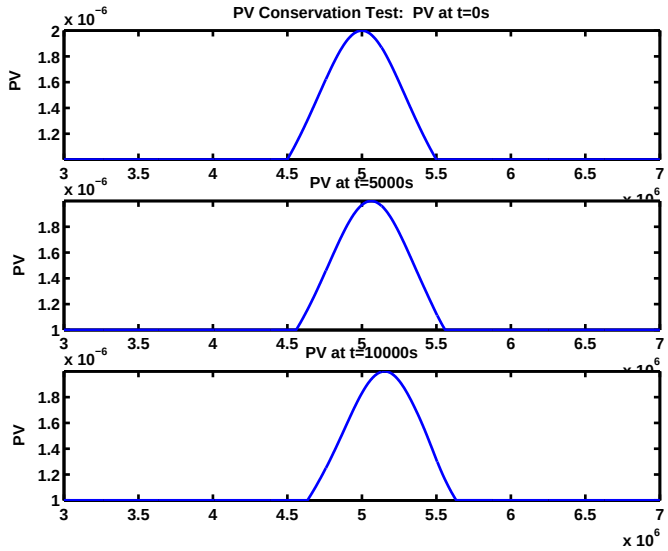

Figure 4.1: PV conservation: Top plot is the PV at the initial time $t = 0$, the central plot is the PV at time $t = 5000s$ and the bottom plot is the PV at time $t = 10,000s$.

where H_c is the maximum height of the orography and $a = 0.4$ is the width. As time moves on the PV is advected by the flow. The PV is conserved well though toward the end of the run the initial shape is slightly degraded. Table 4.1 shows the actual values of the PV at every 200 time steps. Here $\Delta t = 25s$ and, even after a period of nearly 3 hours, our model conserve this quantity to a high degree of accuracy.

Time step	Maximum PV	Total PV
0	0.0000020000000000	0.0084566366447693
200	0.0000019999982466	0.0084563832849261
400	0.0000019999999368	0.0084536772439323

Table 4.1: PV Conservation: Maximum and Total PV every $200s$.

4.3.3 Varying The Rossby Number

The Rossby number was defined previously by equation (2.16) as

$$R_0 = \frac{U}{fL}, \tag{4.33}$$

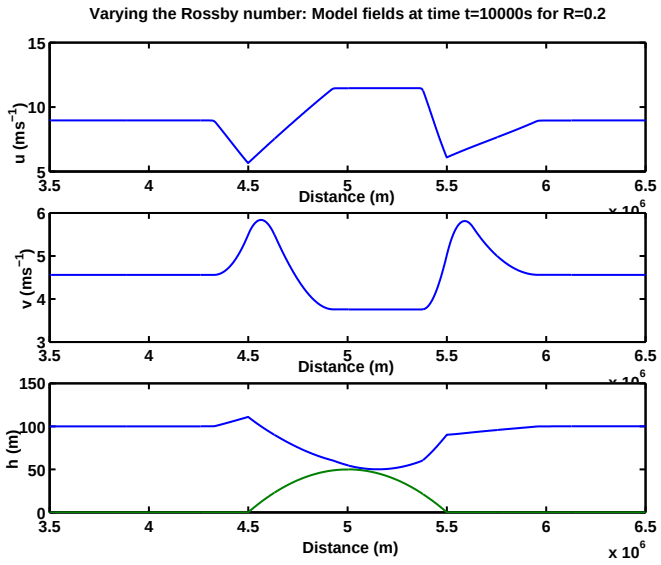

Figure 4.2: Varying the Rossby number: Model fields at time $t = 10000s$ for $R_0 = 0.2$. The top plot shows the u field, the middle plot is the v field and the bottom plot is the depth h, where the orography H is shown with the green line.

where U is a characteristic velocity and L is a characteristic length scale. This parameter is a measure of how important rotation is in the flow [49]. In the mid latitudes for atmospheric motions which carry weather systems $R_0 \approx 0.1$. Here we can expect the Earth's rotation to be important. In this regime it has been shown that the formation of jumps is prevented, or at least delayed, by the rotation [46]. As the Rossby number increases rotation becomes less dominant and when $R_0 \approx 1.0$ we reach a regime where the rotation can no longer prevent the formation of jumps.

In this experiment we look at how varying the Rossby number will affect the flow in our model. We consider $R_0 = 0.2$ and $R_0 = 2.0$. The Rossby number is varied by fixing f and U and varying the length scale L. The length scale that we take is the width of the orography, a, where the orography is given by (4.32) with $H_c = 50m$. Figures 4.2 and 4.3 show the results for the experiment with $R_0 = 0.2$ and $R_0 = 2.0$. They are as we would expect; rotation dominates in the first regime and the formation of a jump is prevented, in the second case rotation no longer

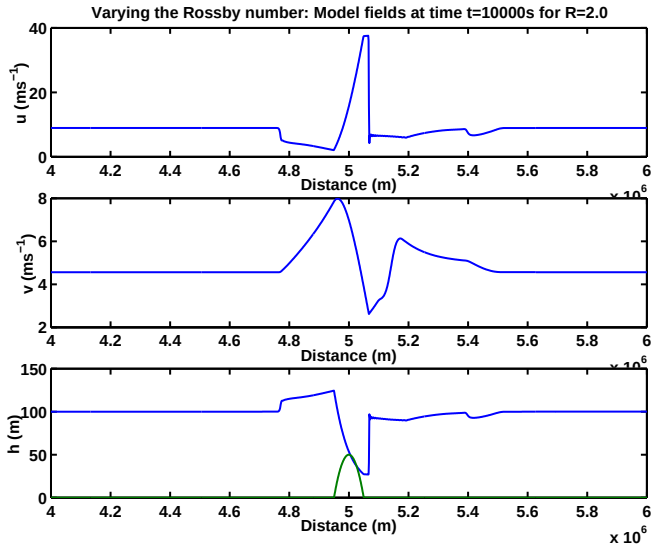

Figure 4.3: Varying the Rossby number: Model fields at time $t = 10000s$ for $R_0 = 2.0$. The top plot shows the u field, the middle plot is the v field and the bottom plot is the depth h, where the orography H is shown with the green line.

dominates and a jump starts to form.

Further examination of figure 4.2 shows that there is a time oscillation in u and v and two smooth gravity waves propagating away from the orography. The oscillation in the velocity fields is such that it is the 1D equivalent of an inertial oscillation. Inertial oscillations are long inertial-gravity waves with periods approximately equal to f [24]. This motion is rarely seen in the atmosphere as it requires the absences of a pressure gradient force. Usually this is observed in oceanic flows or in large lakes.

To verify that the inertial oscillation is a true solution of our model and, if so, to confirm its period $2\pi/f$, we consider a very simple case. We ignore orography and introduce no variation with x in u, v, and h in the initial conditions. Equations (4.1) to (4.3) therefore reduce simply to

$$\frac{\partial u}{\partial t} = fv, \tag{4.34}$$

$$\frac{\partial v}{\partial t} = -fu, \tag{4.35}$$

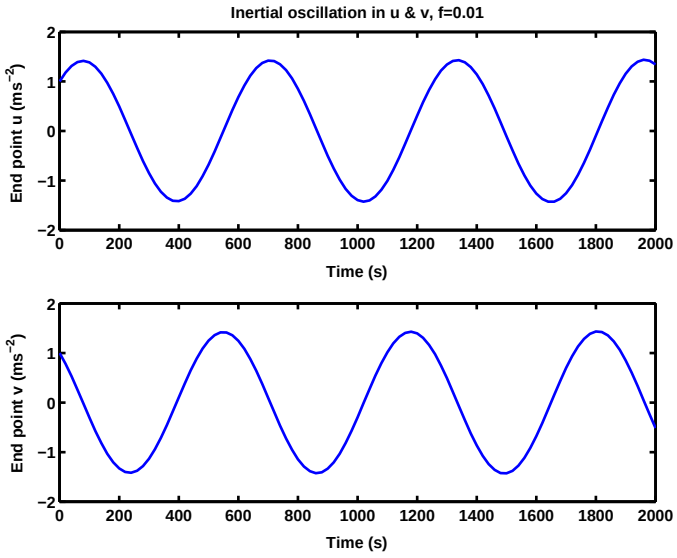

Figure 4.4: Inertial Oscillation for $f = 0.01$. The top plot is the value of u at a fixed point in the domain, the end point, plotted against time. In the bottom plot we have the same for v.

which can be combined to form the *harmonic oscillator* equations

$$\frac{\partial^2 v}{\partial t^2} + f^2 v = 0, \tag{4.36}$$

$$\frac{\partial^2 u}{\partial t^2} + f^2 u = 0. \tag{4.37}$$

Equations (4.36) and (4.37) have solutions of the form

$$u = V \sin(ft),$$

and

$$v = V \cos(ft),$$

with period $2\pi/f$. The inertial oscillation is therefore a solution of the model. We now show the numerical results of running this simple experiment. Figure 4.4 shows the value of u and v fields for a fixed point in space at each time step. There is no variation with x in u, v and h. We see that the period of the oscillation is a little above $600s$ and this compares well to $2\pi/f \approx 628$ with $f = 0.01$.
Returning to the experiment with $R_0 = 2.0$, the second regime, we see that rotation is no longer able to prevent the jump from forming. The SISL scheme

cannot be expected to represent the jump well and we see in figure 4.3 over and under shoots on the jump.

Now that we have verified that our model does accurately represent changes in flow regimes, we turn our attention to the geostrophic adjustment problem introduced in section 2.3.2.

4.3.4 Geostrophic Adjustment Experiments Using the 1D SWEs

In 2.3.2 we introduced geostrophic adjustment for the linearised 2D SWEs. We now demonstrate this process using our model. We start with an initially unbalanced state confined to the centre of the domain, the solid blue line in figures 4.5 and 4.6, and elsewhere all the fields are constant and in balance. We run two experiments: the first with the length scale of the unbalanced region is greater than the Rossby radius, i.e.

$$L_u > L_r,$$

where L_u is the unbalanced length scale and

$$L_r = \frac{\sqrt{\phi}}{f}$$

is the Rossby radius, defined previously by equation (2.20). The second experiment has

$$L_u > L_r.$$

Both the experiments are run without orography and $U = 0$. For small amplitude perturbations the model evolution should be close to linear and therefore should be similar to the solutions of the linear problem given in section 2.3.2.

In figures 4.5 and 4.6 the initial state is the solid blue line and the final state is the dashed red line. The departure from balance is given by

$$DLB = fv - \frac{\partial \phi}{\partial x}$$

and is the top plot in figures 4.5 and 4.6, the middle plot is ϕ and the bottom plot is ψ. In both the experiments $L_u = 500m$, but the Rossby radius is varied by

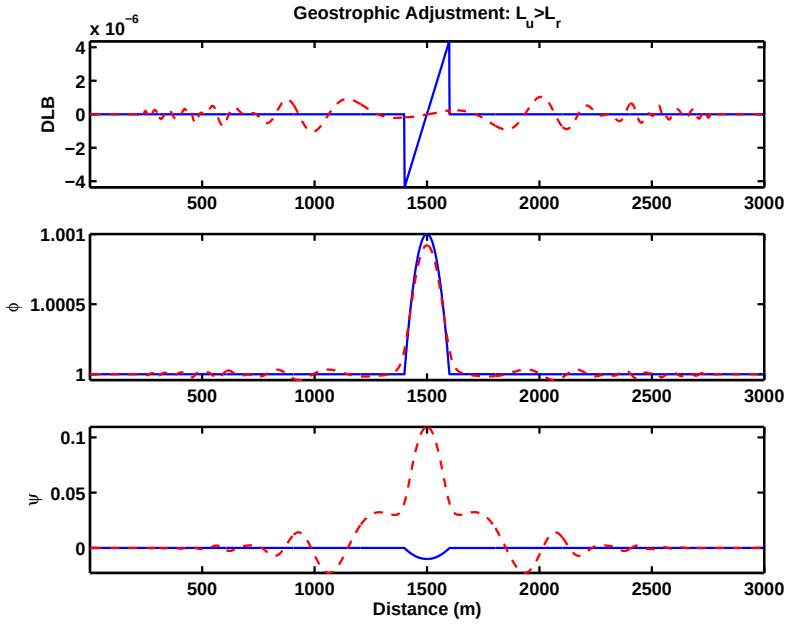

Figure 4.5: Geostrophic adjustment: $L_r < L_u$. The blue solid line indicates the initial fields, the red dashed line indicates the final fields. The top plot is the departure from balance, the middle plot is ϕ and the bottom plot is ψ.

changing the depth of the fluid so that $L_r = 100m$ in the first experiment and $L_r = 2000m$ in the second experiment.

In each of the experiments we see that the unbalanced perturbation is dispersed leaving behind an adjusted balanced state. This can be seen by looking at the DLB in the figures for both experiments. We see that initially at time $t = 0$ (the blue solid line) the DLB is non-zero in the centre of the domain. At the end of the experiment (dashed red line) the DLB in the centre of the domain is approximately zero. In figure 4.5 the Burger number $B_u = 0.2$ and as predicted in section 2.3.2 the ψ field adjusts to balance the ϕ field. On the other hand in figure 4.6 for $B_u = 4.0$, the opposite is observed and this time the ϕ field adjusts to balance the ψ field.

By looking at the geostrophic adjustment problem in our model we are able to verify that the model has a non-trivial balanced state even when orography is not included. We are also confident that its behaviour is as we would expect from the

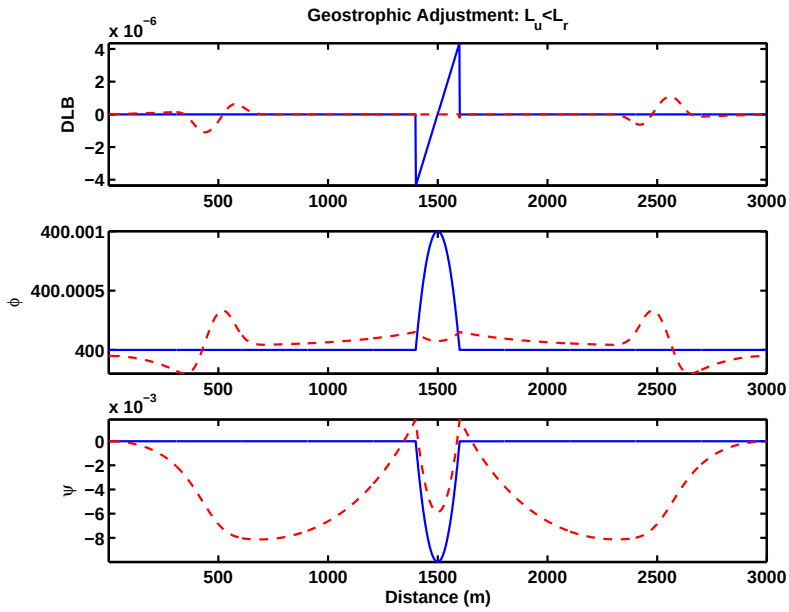

Figure 4.6: Geostrophic adjustment: $L_r > L_u$. The blue solid line indicates the initial fields, the red dashed line indicates the final fields. The top plot is the departure from balance, the middle plot is ϕ and the bottom plot is ψ.

theory described in 2.3.2.

Having now tested the non-linear model we turn our attention to developing a linear approximation to the non-linear model. This linear model is called the tangent linear model (TLM) and we also derive its adjoint. These models are required for the inner loop minimisation in the incremental 4D VAR algorithm.

4.4 Deriving The Linear Models

There are several approaches for deriving the linear models necessary to implement the incremental 4D VAR algorithm. The first we refer to as the discrete method and the second the semi-continuous method [34]. Both of the methods start from the continuous non-linear equations. The discrete method finds the discrete non-linear model equations and then linearises the non-linear code directly to obtain the tangent linear model (TLM). By considering the TLM code as a sequence of linear operations the Adjoint model can be found directly from the

TLM code by applying a few simple rules. These will be discussed in the following sections. Thus both the TLM and the adjoint model are derived. The semi-continuous method, on the other hand, first linearises the continuous non-linear model and then discretises the continuous linear equations to derive what is referred to as the perturbation forecast model. The adjoint of the perturbation forecast model can be found from the linear model code in the same way as the direct method.

The semi-continuous method was proposed in [40] and is used in the Met Office data assimilation. The choice of this method partly involved the additional freedom to design the linear model with specific goals in mind and partly due to the ambiguity of some aspects of the direct method. For instance, in NWP, there are many non-differentiable physical processes and it is not clear how to define their adjoint. Also, iterative procedures are commonly used in these models and again it is not obvious how to apply the direct method.

We choose, however, to follow the direct method since in our model none of the limitations of the direct method apply; we do not have any physical processes and the only iterative procedures lie in solving an elliptic equation that can be solved directly in the TLM, and in the departure point calculation, which only requires two iterations [60] and therefore can just be treated as two statements. We now derive the TLM and adjoint model using the direct method.

4.4.1 The Tangent Linear Model

The Tangent Linear Model is developed directly from the non-linear model source code. There are several steps which required extra analysis.

Firstly, in the departure point calculation equation (4.20) we use the Fortran intrinsic function INT, which is non-differentiable. To linearise this part of code we assume that the departure point and arrival point are in the same grid box and thus we no longer require the function. Also, since the departure point iteration is performed only a fixed number of times we are able to linearise the code directly

by assuming that each iteration is just another Fortran statement.
The exception is the elliptic iteration (4.27). In developing the TLM we actually linearise the original elliptic equation (4.26) which is equivalent to (4.27). To do this we put

$$\phi_i = \overline{\phi_i} + \delta\phi_i,$$

and

$$R_i = \overline{R_i} + \delta R_i,$$

and the linear form of (4.26) is

$$-c\delta\phi_{i+1}^{n+1} + (2c + \frac{1}{\overline{\phi_i}})\delta\phi_i^{n+1} - c\delta\phi_{i-1}^{n+1} = \delta R_i, \tag{4.38}$$

which can be solved directly as the tri-diagonal matrix formed by the left hand side is strictly diagonally dominant. Therefore an iterative procedure to solve this equation is not required.

Testing the Tangent Linear Model

We verify TLM is coded correctly by testing the models:

1. Correctness.

2. Validity.

The TLM correctness is verified by considering the residual and relative errors of the evolution of a perturbation in the non-linear model compared to that of the evolution of the same perturbation in the TLM. The evolution of the non-linear model from time t_0 to t_n is given by

$$x(t_n) = \mathcal{M}(t_n, t_0)[x(t_0)] = \mathcal{M}(t_n, t_0)[\overline{x}(t_0) + \delta x(t_0)], \tag{4.39}$$

where $\mathcal{M}(t_n, t_0)$ is the non-linear model evolution from t_0 to t_n and $x(t_0) = \overline{x}(t_0) + \delta x(t_0)$ where $\overline{x}$ is the linearisation state and $\delta x(t_0)$ is a small perturbation to it. The evolution of $\delta x(t_0)$ in the TLM is

$$\delta x(t_n)^{TL} = \mathbf{M}(t_n, t_0)\delta x(t_0), \tag{4.40}$$

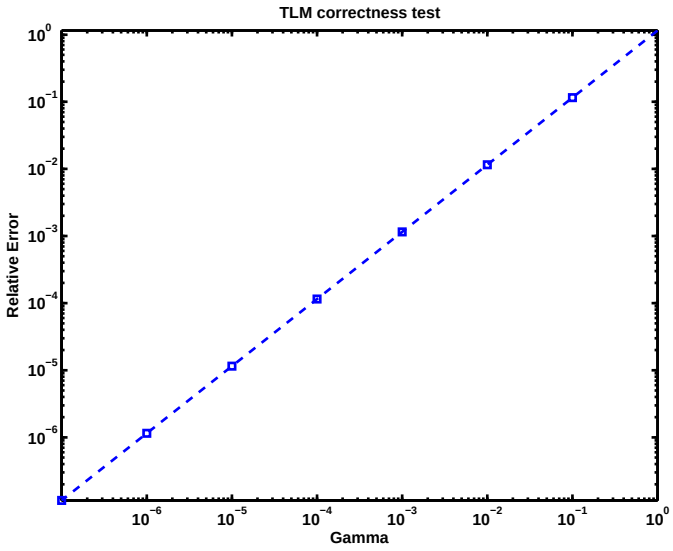

Figure 4.7: Tangent Linear Model correctness test for u:

where $\mathbf{M}(t_n, t_0)$ is the TLM evolution from t_0 to t_n and $\delta x(t_n)^{TL}$ is the perturbation evolved to time t_n by TLM. This can be compared with the total perturbation resulting from the non-linear model evolution,

$$\delta x(t_n)^{NL} = \mathcal{M}(t_n, t_0)[\overline{x}(t_0) + \delta x(t_0)] - \mathcal{M}(t_n, t_0)[\overline{x}(t_0)], \tag{4.41}$$

where $\delta x(t_n)^{NL}$ is the perturbation evolved to time t_n by non-linear model. We can now consider the relative error in the TLM evolution. This can be expressed

$$\frac{\left\|\delta x(t_n)^{NL} - \delta x(t_n)^{TL}\right\|^2}{\left\|\delta x(t_n)^{TL}\right\|^2}. \tag{4.42}$$

If we consider the perturbation $\gamma\delta x$ where γ is a scalar, then in the limit

$$\lim_{\gamma \to 0} \frac{\left\|\gamma\delta x(t_n)^{NL} - \gamma\delta x(t_n)^{TL}\right\|^2}{|\gamma|} = 0. \tag{4.43}$$

If we take $\gamma = 10^{-p}$ and plot the logarithms of relative error against γ we should see the relative error decrease linearly with γ to 0 [69]. This is what is observed in figures 4.7 for u and 4.8 for v.

The results for the correctness of the TLM are independent of the size of the perturbation. So to assess whether the TLM is a qualitatively *good* approximation to the non-linear model we need to look the model's validity. Here we assess

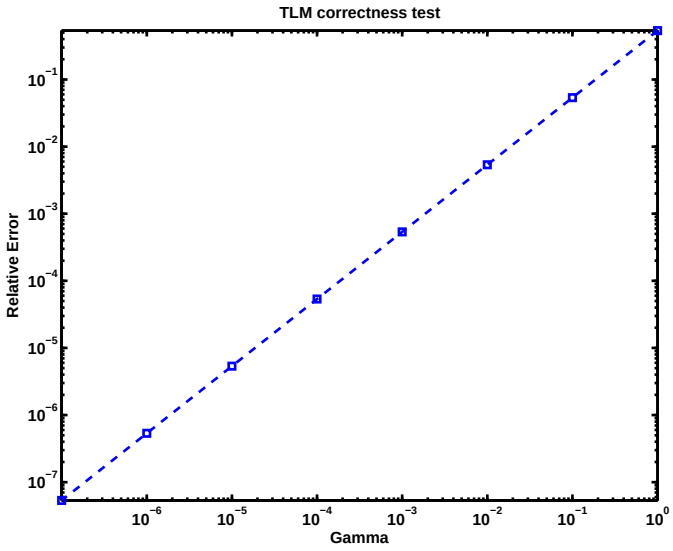

Figure 4.8: Tangent Linear Model correctness test for v

whether the TLM approximates the non-linear model well enough for a realistically sized perturbation over the length of time we require. To answer this we take 10% perturbations to all model variables. The non-linear model and TLM were integrated over a time of $10000s$, which is approximately the length of time of an operational 4D-VAR data assimilation window. We require the TLM to be a good approximation of the evolution of the perturbation over this time window. Figure 4.9 shows the non-linear perturbations, $\delta x(t_n)^{NL}$ as given by (4.41) evolved to $10,000s$ (the solid line) and the TLM evolution $\delta x(t_n)^{TL}$, given by (4.40), at the same time (the dashed line). We see that the TLM evolution stays very close to that of the non-linear evolution. The only exceptions are in a few places where there are sharp changes in the fields but this is to be expected in a linear approximation. This test was run in several regimes and similar results were obtained.

By verifying the TLM's correctness and validity we can be satisfied that the TLM does provide an accurate approximation to the non-linear model over the required time. We can now move on to deriving the adjoint of the TLM.

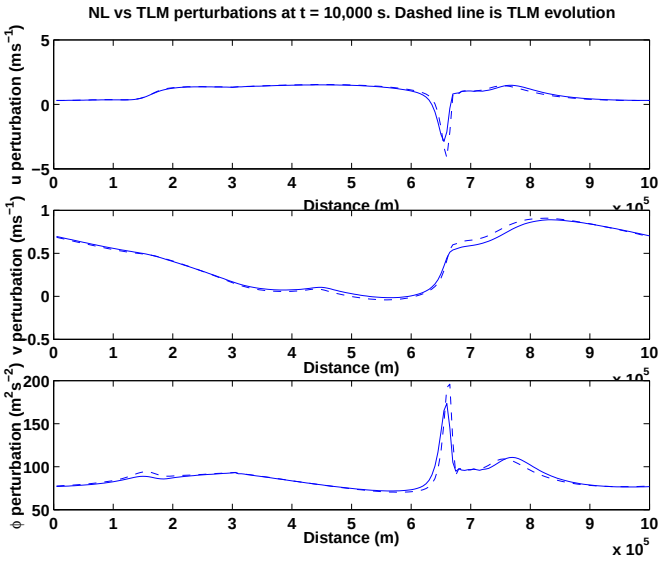

Figure 4.9: Tangent Linear Model Validity Test. Non-linear (solid line) vs Tangent Linear (dashed line) perturbations at 10,000s. The top plot shows the perturbations for u, the middle plot the perturbations for v and the bottom plot the perturbations for ϕ.

4.4.2 The Adjoint Model

Now that we have a TLM and it has been tested for correctness and validity we can derive the Adjoint directly from the TLM source code. This is achieved by viewing the code as a sequence of linear operators. This process is described in detail in [4]. Once the Adjoint model is coded the definition of an Adjoint, $\mathbf{M}^T$,

$$\langle \mathbf{M}\delta x, \mathbf{M}\delta x \rangle = \left\langle \delta x, \mathbf{M}^T\mathbf{M}\delta x \right\rangle, \tag{4.44}$$

where $\mathbf{M}$ is the TLM, $\mathbf{M}^T$ the Adjoint model, and $\langle , \rangle$ the inner product, can be applied to test the model. This test proceeds as follows,

1. Start with perturbation δx.
2. Apply the TLM to obtain $\mathbf{M}\delta x$.
3. Calculate the left hand side of equation (4.44).
4. Apply the Adjoint model to $\mathbf{M}\delta x$ i.e. calculate $\mathbf{M}^T\mathbf{M}\delta x$.

Inner Product	Value
$\langle \mathbf{M}\delta x, \mathbf{M}\delta x \rangle$	518151.6667245743
$\left\langle \delta x, \mathbf{M}^T\mathbf{M}\delta x \right\rangle$	518151.6667245743

Table 4.2: Adjoint test results

5. Calculate the right hand side of equation (4.44).

6. Verify that the equality holds to machine precision.

This test was run several times and the adjoint definition was satisfied to machine precision. Table 4.2 is a sample of the results using random numbers to generate fields. We can now be confident that the adjoint model is coded correctly.

4.5 Summary

We started this chapter by introducing the 1D SWEs, which will be used in the remainder of this work. We choose to solve the equations using a SISL scheme, a scheme used throughout NWP. The scheme is then applied to the 1D SWEs and the steps in the numerical method are outlined.
After the model was tested we then demonstrated some key properties of the model. We looked at the conservation of PV and the behaviour of the model in different Rossby regimes. We then gave two examples of geostrophic adjustment. We then applied the direct method of adjoint construction to develop both the TLM and the adjoint models from our non-linear model. The details of each model are then described and the standard tests for each presented with our test results.

Chapter 5

Control Variable Transforms for the 1D Shallow Water Equations

The final stage in the set up of the incremental 4D VAR is the development of the control variable transform. We now derive two different versions of the control variable transform; a vorticity-based transform, and a PV-based transform.
In chapter 3 the use of the control variable transform in 4D VAR was introduced as a method of modelling the background error covariances. The control variable transform has two distinct components; the first is to represent the cross-correlations between model variables referred to as the parameter transform, and the second to model the auto-correlations for each control variable, which is not considered in this work.
We now concentrate solely on the parameter transform. From the discussion in section 3.3.1 we see that most of the current implementations of the control variable transform are derived from the ideas of [47] and are vorticity-based. The only exception is the height/mass-based transform of [25] that uses very similar ideas. These transforms only approximate the balanced variable since they assume that either vorticity (or mass) is a completely balanced variable. Whilst vorticity is a good approximation to the balanced flow in a high Burger regime, it is not a valid approximation in all regimes [68], as we demonstrate in section 2.4.2. This could potentially cause the data assimilation to generate unrealistic analysis increments.

To overcome this potential limitation a PV-based transform was proposed in [11] and initial results, whilst not being conclusive, suggest potential benefit.
In this chapter we derive the vorticity-based and PV-based transforms for the 1D Shallow Water Equations (SWEs) defined by equations (4.1) to (4.3). The transforms are derived using the dynamical concepts presented in chapter 2. We recall that the linearised form of the 2D SWEs model has three normal modes. The same is true for the 1D equations. The slow, balanced mode is characterised by the linearised PV, q'. The remaining two fast modes can be related to the geostrophic departure, ${}_a\zeta'$, defined by

$$ {}_a\zeta' = f\zeta' - g\frac{\partial^2 h'}{\partial x^2}, \tag{5.1} $$

which is the one dimensional departure from equation (2.88), and the divergence $\mathcal{D}'$

$$ \mathcal{D}' = \frac{\partial u'}{\partial x}, \tag{5.2} $$

the one dimensional form of equation (2.7). We recall that, in the system defined by equations (4.1) to (4.3), we have assumed the model variables do not vary in the y direction.
In [18] the normal modes of the SWEs are used to define a mapping from physical space to $q', \mathcal{D}'$ and ${}_a\zeta'$ (${}_a\zeta'$ is referred to as the ageostrophic vorticity in [18]). A similar mapping is now used to derive the control variable transform.
The vorticity-based control variables are a convenient approximation to the normal mode decomposition where the balanced variable is approximated by the vorticity and all the rotational wind is assumed to be balanced. The PV-based variables, on the other hand, allow for an unbalanced component of the rotational wind and are similar to those used in [18]. The PV-based variables do not approximate the balanced mode.

5.1 Vorticity-Based Transform

The vorticity-based control variables are the streamfunction ψ', velocity potential χ' and 'unbalanced pressure' or, in the case of the SWEs, the residual unbalanced

height h'_{res}. The Helmholtz decomposition is used to separate velocities into rotational and divergent parts. In 1D the Helmholtz decomposition reduces to vorticity

$$\zeta' = \frac{\partial v'}{\partial x} = \frac{\partial^2 \psi'}{\partial x^2}, \tag{5.3}$$

and divergence

$$\mathcal{D}' = \frac{\partial u'}{\partial x} = \frac{\partial^2 \chi'}{\partial x^2}, \tag{5.4}$$

with velocities u' and v' given by

$$u' = \frac{\partial \chi'}{\partial x}, \tag{5.5}$$

$$v' = \frac{\partial \psi'}{\partial x}. \tag{5.6}$$

We now assume that ψ' is a totally 'balanced' variable, following [47], and use the linearised balance relationship, given by (4.5), to define a balanced height variable h'_b, i.e.

$$f\frac{\partial^2 \psi'}{\partial x^2} - g\frac{\partial^2 h'_b}{\partial x^2} = 0, \tag{5.7}$$

where the subscript b refers to the 'balanced' part of the height variable. The residual height is then given by

$$h_{res} = h - h_b.$$

5.1.1 Vorticity-Based Scheme: The T-Transform

Using the vorticity equation (5.3) and the linear balance equation (5.7) the T-transform

$$\mathbf{z}' = \mathbf{T}\mathbf{x}'$$

with

$$\mathbf{x}' = \begin{pmatrix} u' \\ v' \\ h' \end{pmatrix},$$

and

$$\mathbf{z}' = \begin{pmatrix} \psi' \\ h'_{res} \\ \chi' \end{pmatrix}$$

in the context of the 1D SWEs is given by solving a sequence of equations. We use the notation $< x >$ to denote the mean value of x. We note that when differentiated quantities are used as control variables we must store the mean values else they will be lost. The method proceeds as follows:

Step 1 Solve

$$\frac{\partial^2 \chi'}{\partial x^2} = \mathcal{D}' \tag{5.8}$$

for χ' subject to periodic boundary conditions. The solution is unique up to an additive constant.

Step 2 Solve

$$\frac{\partial^2 \psi'}{\partial x^2} = \zeta' \tag{5.9}$$

for ψ' subject to periodic boundary conditions. The solution is unique up to an additive constant.

Step 3 Calculate the residual height h'_{res} by

$$h'_{res} = h' - h'_b = h' - \frac{f}{g}\psi', \tag{5.10}$$

where

$$h' = h'_b + h'_{res}.$$

The variable h'_b in equation (5.10) is found by integrating (5.7), the linear balance equation, twice to give

$$h'_b = \frac{f}{g}\psi' + c_1 x + c_2,$$

where c_1 and c_2 are constants of integration. Then $c_1 = 0$ because both ψ' and h'_b are periodic in x, and we set $c_2 = 0$.

Step 4 Store the spatial mean values of u' and v', $< u' >$ and $< v' >$ respectively. These are extra control variables that are needed as these means cannot be otherwise recovered from ψ', χ' or h'_{res} in the U-transform (see below).

Equations (5.9) for ψ' and (5.8) for χ' are solved with periodic boundary conditions and the solutions are unique up to a constant provided the right hand side has a zero mean value. In both cases the right hand sides are derivatives of periodic functions and therefore will always have a zero mean value. The solutions are therefore unique up to a constant and we choose this constant such that the mean value $< \psi' >$ of ψ' is zero and the mean value $< \chi' >$ of χ' is also zero, where the angled brackets indicate we are taking the mean value of the variable. In solving (5.9) for ψ' and (5.8) for χ' we lose a degree of freedom by virtue of the fact that we choose ψ' and χ' to have zero mean values. This available degree of freedom is used to retain the mean values $< v >$ of v and $< u >$ of u that are lost through differentiation. The values $< u >$ and $< v >$ are also control variables. The model variables and the vorticity-based control variables now have equal degrees of freedom. The T-transform essentially solves a sequence of equations producing the control variable increments given model variable increments.

5.1.2 Vorticity-Based Scheme: The U-Transform

Using equations (5.5), (5.6) and (5.7) we are able to
derive the U-transform,

$$\mathbf{x}' = \mathbf{U}\mathbf{z}'.$$

Given the control variable increments the U-transform proceeds as follows:

Step 1 Find the velocity v' from ψ' and $< v' >$

$$v' = \frac{\partial \psi'}{\partial x} + < v' > . \tag{5.11}$$

Step 2 Find the balanced height increment h'_b from ψ' and calculate the full height increment h'

$$h' = h'_b + h'_{res} = \frac{f}{g}\psi' + h'_{res}. \tag{5.12}$$

Step 3 Find the velocity u' from χ' and $< u' >$

$$u' = \frac{\partial \chi'}{\partial x} + < u' > . \tag{5.13}$$

Again a sequence of equations is solved and we reconstruct the original model variables from the control variables.

It is worth noting at this point that the consideration of the mean values $< u' >$ and $< v' >$ is more natural in the implementation of the transforms at the Met Office. At the Met Office the transforms are solved in spectral space and are only applied to wavenumbers one and above.

5.1.3 Numerical Implementation of the Vorticity-Based Transform

We now describe the numerical implementation of the vorticity-based transform. We start with the T-transform.

T-Transform

The equations (5.9) and (5.8) for ψ' and χ' are solved in the same way. We now describe the solution method taking χ' as an example. We desire to solve

$$\frac{\partial^2 \chi'}{\partial x^2} = R(x) \tag{5.14}$$

with periodic boundary conditions. The equation has a unique solution up to an additive constant provided the right hand side has a zero mean value, this is sometimes refered to as the *compatibility condition*. In this case $R(x) = \frac{\partial u'}{\partial x}$ so the zero mean condition is satisfied since u is periodic in x. The solution χ' is therefore unique up to an additive constant and we choose this constant such that $< \chi' >= 0$.

We have now specified all the required boundary conditions and solve the equation using discrete Fourier transforms (DFT). The DFT of a field X is given by

$$\hat{X}'(k_n) = \frac{1}{\sqrt{N}} \sum_{j=1,N} X(x_j) e^{-ik_n x_j}, \tag{5.15}$$

for each wave number $k_n = \frac{2\pi n}{\Delta x N}$ with $n = 1, \ldots, N-1$, where Δx is the spatial grid length. The inverse transform is given by

$$X(x_j) = \frac{1}{\sqrt{N}} \sum_{n=1,N-1} \hat{X}(k_n) e^{ik_n x_j}. \tag{5.16}$$

We apply the Fourier transform to the discrete equation

$$\frac{\chi_{i+1} - 2\chi_i + \chi_{i-1}}{\Delta x^2} = R_i, \tag{5.17}$$

where the prime has been dropped from the χ' for convenience. This gives

$$\frac{1}{\sqrt{N}\Delta x^2} \sum_{n=1,N-1} \hat{\chi}'(k_n) \left(e^{ik_n x_{i+1}} - 2e^{ik_n x_i} + e^{ik_n x_{i-1}} \right) = \frac{1}{\sqrt{N}} \sum_{n=1,N-1} \hat{R}(k_n) e^{ik_n x_i}. \tag{5.18}$$

After further manipulation we obtain

$$\hat{\chi}'(k_n) = \frac{\Delta x^2}{(2\cos(k_n \Delta x) - 2)} \hat{R}(k_n), \tag{5.19}$$

using the identity $e^{ik_n \Delta x} = \cos(k_n \Delta x) + i \sin(k_n \Delta x)$. Thus, to solve the equation we execute the following steps:

1. Transform the right hand side:

   The right hand side is transformed to Fourier space using a discrete Fourier transform, equation (5.15), for permitted wavenumbers, k_n, with

   $$k_n = \frac{2\pi n}{\Delta x N} \quad \text{for} \quad n = -\frac{N}{2}, \ldots, \frac{N}{2} - 1$$

   where N is the number of points in domain.

2. Calculate solution in Fourier space:

   Solution is calculated in Fourier space by multiplying each mode

   $$n = -\frac{N}{2}, \ldots, \frac{N}{2} - 1$$

   by

   $$\frac{\Delta x^2}{(2\cos(k_n \Delta x) - 2)}.$$

3. Enforce boundary conditions and mean:

   We enforce the boundary condition for the mean value by setting zero mode $\hat{\chi}'(k_0) = 0$. The periodic boundary conditions are satisfied naturally in Fourier space.

4. Calculate solution in model space:

   Apply the inverse transform, equation (5.16), to obtain the solution in model space.

Once we have followed the above method we obtain the control variables χ' and ψ'. It is then simple to calculate h'_{res} from ψ',

$$h'_{resi} = h'_i - \frac{f}{g}\psi'_i, \tag{5.20}$$

where the subscript i indicates the value of the variable at point $i\Delta x$. Thus we obtain the required control variables.

U-Transform

We now describe the numerical solution of the U-transform. This transform must be efficient since it is used once every inner iteration and once every outer iteration of the incremental 4D VAR algorithm. The model variables u', v', and h' are re-constructed from the control variables as follows;

$$u'_i = \frac{\chi'_{i+1} - \chi'_i}{\Delta x} + < u' >, \tag{5.21}$$

$$v'_i = \frac{\psi'_{i+1} - \psi'_i}{\Delta x} + < v' >, \tag{5.22}$$

$$h'_i = \frac{f}{g}\psi'_i + h_{res}{}'_i, \tag{5.23}$$

for

$$i = 1, \ldots, N,$$

where N is the number of grid points and $x_{N+1} = x_1$ and $x_0 = x_N$. We note that χ' and ψ' values are located on ϕ points on the staggered grid used in the non-linear model, see figure 5.1. Thus finding u' and v' is straightforward, as is calculating h'. We now derive the PV-based control variable transform.

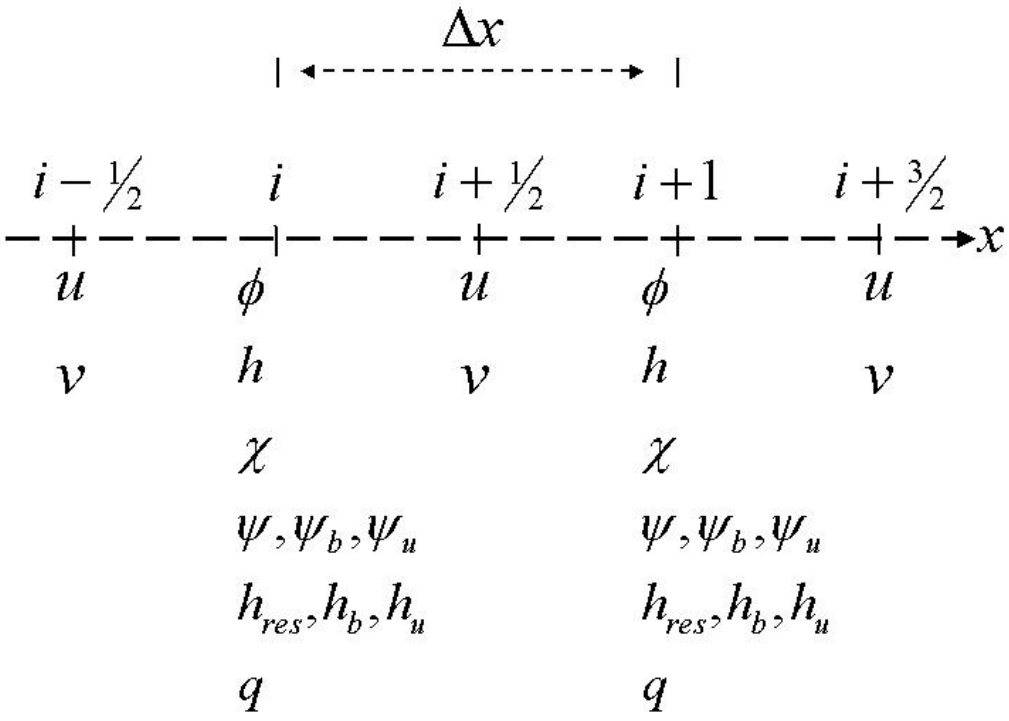

Figure 5.1: Control variables on the staggered grid.

5.2 PV-Based Transform

We start by defining our variables. Again the Helmholtz decomposition is used to separate velocities into rotational and divergent parts. In 1D the Helmholtz decomposition reduces to equations (5.3) to (5.6). Additionally we let

$$h' = h'_b + h'_u$$

and

$$v' = v'_b + v'_u \Rightarrow \psi' = \psi'_b + \psi'_u,$$

where the subscripts refer to the balanced and unbalanced parts of the variable. So we have

$$v'_b = \frac{\partial \psi'_b}{\partial x} \tag{5.24}$$

and

$$v'_u = \frac{\partial \psi'_u}{\partial x}. \tag{5.25}$$

Splitting the velocity v' in this manner allows for balanced and unbalanced components of the rotational wind. Therefore not all the rotational wind is assumed balanced, as it is with the vorticity-based variables.

We define the reference PV for our simplified model as in [67], giving

$$\bar{q} = \frac{1}{\bar{h}}\left(f + \frac{\partial \bar{v}}{\partial x}\right), \tag{5.26}$$

where the reference states $\bar{v}(x,t), \bar{h}(x,t)$ and $\bar{q}(x,t)$ are either the first guess, or background state, on the first outer loop of the incremental 4D VAR, or updates to the background on subsequent outer loops.

For the PV-based transform we define the balanced variables v_b' and h_b' such that they satisfy the linear balance equation (5.7)

$$f\frac{\partial^2 \psi_b'}{\partial x^2} - g\frac{\partial^2 h_b'}{\partial x} = 0 \tag{5.27}$$

and the linearised PV equation. To derive the linearised PV equation we follow [67] and start by linearising (5.26) around a varying reference state

$$q(x,t) = \bar{q}(x,t) + q'(x,t)$$

$$v(x,t) = \bar{v}(x,t) + v'(x,t)$$

$$h(x,t) = \bar{h}(x,t) + h'(x,t),$$

where the overbar denotes the reference state and the prime is a perturbation to it. This gives

$$q = \frac{1}{h}\left(f + \frac{\partial v}{\partial x}\right) = \frac{1}{\bar{h} + h'}\left(f + \frac{\partial \bar{v}}{\partial x} + \frac{\partial v'}{\partial x}\right).$$

Therefore, neglecting products of the perturbations, we have

$$\bar{q}\bar{h} + q'\bar{h} + \bar{q}h' = f + \frac{\partial \bar{v}}{\partial x} + \frac{\partial v'}{\partial x}$$

and using equation (5.26) gives

$$q'\bar{h} = \frac{\partial v'}{\partial x} - \bar{q}h'. \tag{5.28}$$

The balanced variables v'_b and h'_b are associated solely with the linearised PV, q', so we have

$$\frac{\partial v'_b}{\partial x} - \bar{q}h'_b = q'\bar{h}. \tag{5.29}$$

There are also corresponding equations for the unbalanced variables. We define the unbalanced variables v'_u and h'_u such that they satisfy the departure from linear balance (DLB) equation

$$fv'_u - g\frac{\partial h'_u}{\partial x} = fv' - g\frac{\partial h'}{\partial x} \tag{5.30}$$

and

$$\frac{\partial v'_u}{\partial x} - \bar{q}h'_u = 0, \tag{5.31}$$

i.e. the unbalanced variables do not contribute to the PV increment.
Re-writing these equations using the balanced and unbalanced streamfunctions ψ'_b and ψ'_u gives the following four equations

$$f\frac{\partial^2 \psi'_b}{\partial x^2} - g\frac{\partial^2 h'_b}{\partial x^2} = 0, \tag{5.32}$$

$$\frac{\partial^2 \psi'_b}{\partial x^2} - \bar{q}h'_b = q'\bar{h}, \tag{5.33}$$

$$f\frac{\partial^2 \psi'_u}{\partial x^2} - g\frac{\partial^2 h'_u}{\partial x^2} = {}_a\zeta', \tag{5.34}$$

$$\frac{\partial^2 \psi'_u}{\partial x^2} - \bar{q}h'_u = 0. \tag{5.35}$$

These equations, with appropriate boundary conditions specified later define four variables ψ'_b, ψ'_u, h'_b and h'_u, although only two are needed as control variables. We choose to use ψ'_b and h'_u (χ' is the third control variable, which is identical to the vorticity-based formulation). This gives us one balanced and two unbalanced variables to reflect the normal modes. Theoretically we could have also choosen h'_b and ψ'_u instead of ψ'_b and h'_u, but this would require division by f. This would cause problems when solving the transform on the sphere where $f \to 0$ on the equator.

5.2.1 PV-Based Scheme: The T-Transform

Using equations (5.32) to (5.35) and (5.4) the T-transform

$$\mathbf{z}' = \mathbf{T}\mathbf{x}'$$

with

$$\mathbf{x}' = \begin{pmatrix} u' \\ v' \\ h' \end{pmatrix},$$

and

$$\mathbf{z}' = \begin{pmatrix} \psi_b' \\ h_u' \\ \chi' \end{pmatrix}$$

is given by solving the following sequence of equations:

Step 1 Solve

$$\frac{\partial^2 \chi'}{\partial x^2} = \mathcal{D}' \tag{5.36}$$

for χ' subject to periodic boundary conditions. The solution is unique up to an additive constant. This is the same equation as (5.8) in the vorticity-based control variable transform.

Step 2 Solve

$$\frac{\partial^2 \psi_b'}{\partial x^2} - \frac{f\bar{q}}{g}\psi_b' = q'\bar{h} \tag{5.37}$$

for ψ_b' subject to periodic boundary conditions. The right hand side is known from the model variable increment fields. The equation has a unique solution provided $\bar{q} > 0$. This is discussed in detail in the next section.

Step 3 Solve

$$f\bar{q}h_u' - g\frac{\partial^2 h_u'}{\partial x^2} = {}_a\zeta' \tag{5.38}$$

for h_u' subject to periodic boundary conditions. As before the right hand side is known from the model variable increment fields and the equation has a unique solution provided $\bar{q} > 0$.

Step 4 Store mean values of u' and v'. These are lost through differentiation.

Equation (5.37) is found by substituting $h_b' = \frac{f}{g}\psi_b'$ from (5.32), the linear balance equation, into equation (5.33). Here we have integrated (5.32) twice with both

constants of integration defined to be zero, as was done for the vorticity-based variables. Equation (5.38) is found by substituting $\nabla^2\psi'_u$ from equation (5.35) into equation (5.34).

Equation (5.36) is solved with periodic boundary conditions for χ' and has a unique solution up to an additive constant provided the right hand side has a zero mean value. The right hand side is a derivative of a periodic function and therefore will always have a zero mean value. The solution is therefore unique up to a constant and we choose this constant such that $< \chi' >= 0$. We therefore lose a degree of freedom in χ'. We use this available degree of freedom to retain the mean value $< u >$ that is lost through differentiation. We also note that a degree of freedom in $\psi' = \psi'_b + \psi'_u$ is lost since the unbalanced streamfunction ψ'_u is found by solving (5.35) i.e.

$$\frac{\partial^2\psi'_u}{\partial x^2} = \bar{q}h'_u \tag{5.39}$$

subject to periodic boundary conditions. The right hand side is known and must have a mean value of zero for the equation to have a solution. Provided that this is true the solution ψ'_u is unique up to an additive constant, chosen to be zero. Thus, we again lose a degree of freedom. This available degree of freedom is used to store the mean value $< v >$. The degrees of freedom in both the control variables and the model variables are now equal. Again, as with the vorticity-based variables, the mean values $< u >$ and $< v >$ are also control variables.

We note from equation (5.38) that if we first apply the T-transform to our model variable increments we will always produce h'_u such that the mean of $\bar{q}h'_u$ is zero. However, the T-transform is not applied before the U-transform in the incremental 4D VAR algorithm since the inner minimisation is performed *in control space*. This issue is addressed later in this section. We first present the inverse transform, the U-transform, in the context of our simplified SWEs.

5.2.2 PV-Based Scheme: The U-Transform

Using equations (5.5), (5.24), (5.25), (5.32) and (5.35) we are able to derive the U-transform

$$\mathbf{x}' = \mathbf{U}\mathbf{z}'$$

for the simplified SWEs. Given the PV-based control variable increments the U-transform proceeds as follows:

Step 1 Find the balanced velocity increment v_b' from ψ_b'

$$v_b' = \frac{\partial \psi_b'}{\partial x}. \tag{5.40}$$

Step 2 Find the unbalanced velocity increment v_u' from ψ_u'

$$v_u' = \frac{\partial \psi_u'}{\partial x}, \tag{5.41}$$

where ψ_u' is found by solving

$$\frac{\partial^2 \psi_u'}{\partial x^2} = \bar{q} h_u'$$

subject to periodic boundary conditions. The right hand side is known and must have a mean value of zero for the equation to have a solution. Provided that this is true the solution ψ_u' is unique up to an additive constant.

Step 3 Reconstruct the full velocity increment v'

$$v' = v_b' + v_u' + < v > . \tag{5.42}$$

Step 4 Find the balanced height increment h_b' from ψ_b' and reconstruct the full height increment h'

$$h' = h_b' + h_u' = \frac{f}{g}\psi_b' + h_u'. \tag{5.43}$$

Here we have assumed that h_u' is such that equation (5.39) has a solution. We discuss this in more detail in the following sections.

Step 5 Find the velocity u' from χ' and $< u >$

$$u' = \frac{\partial \chi'}{\partial x} + < u > . \tag{5.44}$$

So the U-transform solves a series of equations that reconstructs the model variable increments from the control variable increments.

In equation (5.41) the unbalanced streamfunction ψ'_u is found by solving (5.39). The right hand side of (5.39) is known and must have a mean value of zero for the equation to have a solution. We must now consider this in the practical implementation of the U-transform in the incremental 4D VAR algorithm.

We minimise the cost function in control space and therefore the condition that the mean of $\bar{q}h'_u$ is zero may not be satisfied unless explicitly enforced. It is not straight forward how to do this since $\bar{q}$, defined by equation (5.26), is varying in x and is modified on every outer iteration. It is possible to modify h'_u on *each inner iteration* so that the mean of $\bar{q}h'_u$ is zero. This can be achieved since we are always able to subtract a constant from h'_u such that $< \bar{q}h'_u >= 0$. The mean of the full height increment is therefore split between h'_b and h'_u. The implementation is discussed in more detail in the following section.

The problem could be avoided by choosing to approximate $\bar{q}$ by a constant. An approximation of this sort was made in [11]. This would mean that we are simply able to explicitly set $< h'_u >= 0$ and the condition $< \bar{q}h'_u >= 0$ will always be satisfied. We then store the mean of the full height increment solely in h'_b. This approximation is also desirable from an operational perspective since the transform would be less computationally demanding. In the following section we consider the possible implications of making this approximation in the PV-based transform. We choose to make the approximation

$$\bar{q} = \frac{1}{\bar{h}}\left(f + \frac{\partial \bar{v}}{\partial x}\right) \approx \frac{f}{< \bar{h} >}, \tag{5.45}$$

to equation (5.26), and note that this is the PV of the linearisation state used in the linear analysis of section 2.3.1. The approximation essentially assumes that $\frac{\partial \bar{v}}{\partial x} = 0$. This is consistent with the climatological assumptions and zonally averaged quantities used in determining the background error covariance matrix, which are discussed in section 3.4; the climatological average of $\frac{\partial \bar{v}}{\partial x}$ is likely to be close to zero.

5.2.3 Numerical Implementation of the PV-Based Transform

We now describe the numerical implementation of the PV-based transform. We start with the T-transform.

T-Transform

The equation for ψ_b' is

$$\frac{\partial^2 \psi_b'}{\partial x^2} - \frac{f\bar{q}}{g}\psi_b' = R(x), \tag{5.46}$$

where $R(x) = \frac{\partial v'}{\partial x} - \bar{q}h'$. We solve the equation on a periodic domain of N points with $x_0 = x_N$ and $x_{N+1} = x_1$. Derivatives are approximated using a standard central difference so the second derivative of ψ_b' at point $x_i = i\Delta x$, where Δx is the spatial step size is given by

$$\left(\frac{\partial^2 \psi}{\partial x^2}\right)_i = \frac{\psi_{i+1} - 2\psi_i + \psi_{i-1}}{\Delta x^2}. \tag{5.47}$$

Here the subscript and the prime has been dropped from ψ_b' for convenience. Equation (5.46) therefore approximates to

$$\frac{\psi_{i+1} - 2\psi_i + \psi_{i-1}}{\Delta x^2} - \frac{f}{g}\bar{q}_i\psi_i = R_i, \tag{5.48}$$

where $\bar{q}_i = \bar{q}(x_i)$ and $R_i = R(x_i)$. Enforcing the periodic boundary conditions this approximation gives rise to a tri-diagonal system of N equations, given by

$$\begin{pmatrix} -2 - \Delta x^2 \frac{f}{g}\bar{q}_1 & 1 & & & & 1 \\ \ddots & \ddots & \ddots & & & \\ & 1 & -2 - \Delta x^2 \frac{f}{g}\bar{q}_i & 1 & & \\ & & \ddots & \ddots & \ddots & \\ 1 & & & 1 & 2 - \Delta x^2 \frac{f}{g}\bar{q}_N & \end{pmatrix} \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_i \\ \vdots \\ \psi_N \end{pmatrix} = \Delta x^2 \begin{pmatrix} R_1 \\ \vdots \\ R_i \\ \vdots \\ R_N \end{pmatrix}.$$

This system has a unique solution provided the matrix on the left hand side is diagonally dominant i.e.

$$|a_i| \geq |b_i| + |c_i| \tag{5.49}$$

with strict diagonal dominance,

$$|a_i| > |b_i| + |c_i| \tag{5.50}$$

on at least one row. Here a_i is the diagonal element on row i and b_i and c_i are the neighbouring off diagonal elements. This condition is met if $\bar{q}_j > 0$ at point j and $\bar{q}_i \geq 0$ for $i = 1, \ldots, N$. We will discuss this condition more in section 5.2.3. Assuming that this is true we solve the system using a periodic tri-diagonal solver from [17].

The equation for h'_u is

$$f\bar{q}h'_u - g\frac{\partial^2 h'_u}{\partial x^2} = R(x) \tag{5.51}$$

where $R(x) = f\frac{\partial^2 \psi'}{\partial x^2} - g\frac{\partial^2 h'}{\partial x^2}$. We solve the equation on a periodic domain of N points with $x_0 = x_N$ and $x_{N+1} = x_1$ as before. Derivatives are approximated in the same way as for ψ'_b above. Discretising the equation gives

$$f\bar{q}_i h_i - g\frac{h_{i+1} - 2h_i + h_{i-1}}{\Delta x^2} = R_i, \tag{5.52}$$

where the prime and the subscript have been dropped from the h'_u for convenience. This yields the following tri-diagonal system

$$\begin{pmatrix} 2g + \Delta x^2 f\bar{q}_1 & -g & & & & -g \\ \ddots & \ddots & \ddots & & & \\ & -g & 2g + \Delta x^2 f\bar{q}_i & -g & & \\ & & \ddots & \ddots & \ddots & \\ -g & & & -g & 2g + \Delta x^2 f\bar{q}_N & \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_i \\ \vdots \\ h_N \end{pmatrix} = \Delta x^2 \begin{pmatrix} R_1 \\ \vdots \\ R_i \\ \vdots \\ R_N \end{pmatrix}.$$

Again this system has a unique solution provided the matrix on the left hand side is diagonally dominant with strict diagonal dominance on at least one row. We can guarantee this is met if again $\bar{q}_j > 0$ for some point j and $\bar{q} \geq 0$ for $i = 1, \ldots, N$. Assuming this is true the system is solved in the same way as for ψ'_b.

The variable χ' is the same as for the vorticity-based variables, and the equation is therefore solved in exactly the same way, as described in section 5.1.3.

Equations (5.46) and (5.51) for the PV-based variables h'_u and ψ'_b were the first equations tackled in this work. The initial method chosen to solve the equations

was standard finite differences, as described above. We find that the periodic tri-diagonal system obtained once equations (5.46) and (5.51) are discretised is s.d.d. (given a minor constraint on the sign of the linearised PV, which is satisfied by choice of initial conditions). These systems are easily solved with a standard tri-diagonal solver. Therefore, this approach is straightforward to implement. Later we choose a spectral method to solve the equations for χ' and ψ' as this is a natural way to enforce the boundary conditions. In retrospect, a spectral solution method could be applied to all the equations and this might be more efficient. However, for our purposes, calculation times are not a primary consideration since the domain is relatively small.

U-Transform

We now describe the numerical solution of the U-transform. The model variables u', v', and h' are re-constructed from the control variables as follows,

$$u'_i = \frac{\chi'_{i+1} - \chi'_i}{\Delta x} + < u >, \tag{5.53}$$

$$v'_i = \frac{\psi'_{bi+1} - \psi'_{bi}}{\Delta x} + \frac{\psi'_{ui+1} - \psi'_{ui}}{\Delta x} + < v >, \tag{5.54}$$

$$h'_i = \frac{f}{g}\psi_{b_i}{}' + h_{u_i}{}' + c, \tag{5.55}$$

for

$$i = 1, \ldots, N$$

where N is the number of grid points and $x_{N+1} = x_1$ and $x_0 = x_N$. The constant c is determined as a result of adjusting h'_u so that the equation (5.39) for ψ_u has a solution. The calculation of c is given in the next section.

As before all derivatives are approximated with a standard central difference scheme, as shown in figure 5.1. Thus finding u' is straightforward as is calculating v' once we have found ψ'_u. However, there are two complications to the U-transform: Firstly, we must find ψ'_u. Secondly, choosing the constant c needed to give h'. These are now discussed.

U-Transform: Equation for ψ'_u

In the transform we must solve

$$\frac{\partial^2 \psi'_u}{\partial x^2} = \bar{q} h'_u \tag{5.56}$$

for ψ'_u. As we have seen earlier, for an equation of this form with periodic boundary conditions to have a solution we require the right hand side to have a zero mean. This may not be the case since h'_u is generated by the minimisation routine and is not constrained to meet this condition. We therefore define $\widetilde{h'_u}$ as the unbalanced height increment that is generated by the minimisation. Since there is no guarantee that $\widetilde{h'_u}$ will be such that the mean of the $\bar{q}\widetilde{h'_u}$ is zero we need to adjust the value in such a way to respect this condition, i.e. adjust $\widetilde{h'_u}$ so that

$$< \bar{q}(\widetilde{h'_u} - c) >=< \bar{q} h'_u >= 0.$$

Assuming that this is done we may solve the equation in exactly the same way as described in the T-transform when solving for χ', see section 5.1.3.
To adjust $\widetilde{h'_u}$ we define

$$\widetilde{h'_u} = h'_u + c \quad \text{where} \tag{5.57}$$

$$c = \frac{\int \bar{q}\widetilde{h'_u}\ \mathrm{dx}}{\int \bar{q}\ \mathrm{dx}}. \tag{5.58}$$

The constant c is then subtracted from $\widetilde{h'_u}$ and added back onto h.
In the next chapter we also test an approximation to the PV-based transform where we approximate $\bar{q} \approx \frac{f}{<h>}$. Since $\bar{q}$ is then a constant to solve equation (5.56) we only need to constrain the mean value of h'_u to be zero in the minimisation, which can be done by explicitly setting it to zero. Therefore there is no need to adjust h'_u, since

$$< h'_u >= 0 \Rightarrow < \bar{q} h'_u >= 0,$$

when $\bar{q}$ is a constant. When we do not have a constant $\bar{q}$ the adjustment must be calculated on every inner iteration of the minimisation, as we outlined above.

Other Considerations: $\bar{q}$

As we have mentioned in the previous section for ψ'_b and h'_u to have unique solutions in the T-transform we require that $\bar{q} > 0$, to respect condition (5.50). This condition is also present in the continuous problem as we require the solutions to be real. In the minimisation it is possible to have a situation where at some point x_i, $\bar{q}_i < 0$. We therefore need to handle this situation explicitly. We do this by smoothing $\bar{q}$ so that it never falls below a small positive number close to zero, i.e. if $\bar{q}_i < \epsilon$, with $\epsilon << 1$, say, then set explicitly $\bar{q}_i = \epsilon$.

5.3 Testing the Transforms

The transforms are tested by looking at the errors

$$e_1 = \|\mathbf{x} - \mathbf{U}(\mathbf{T}\mathbf{x})\| \tag{5.59}$$

$$e_2 = \|\mathbf{z} - \mathbf{T}(\mathbf{U}\mathbf{z})\| . \tag{5.60}$$

Here we are testing if the correct inverse is coded. The data used to generate the initial fields $\mathbf{x}$ and $\mathbf{z}$ in these tests were taken from two possible sources:

1. Actual model data.

2. Random fields.

When the tests are run for the vorticity and the PV-based transforms we find that the errors e_1 and e_2 are order 10^{-12} or less (i.e. machine precision). The discrete Fourier transform is tested in a similar way and in addition we also compare numerical solutions to exact solutions in specific cases where the analytical solutions are known. Again all errors the order of machine accuracy.
The adjoint transforms for both the vorticity and PV-based transforms are also required in the incremental 4D VAR algorithm. These are derived and tested following the methods described in section 4.4.2.

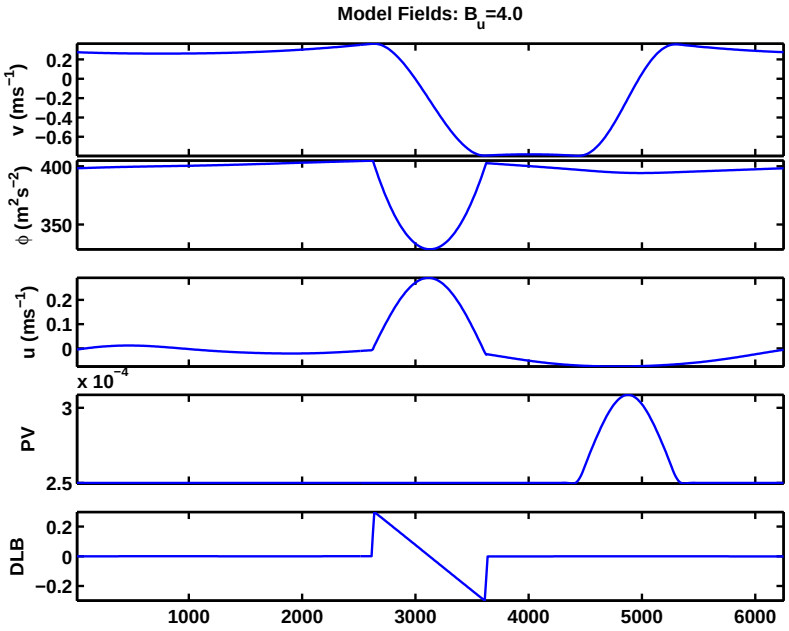

Figure 5.2: Model fields, $B_u = 4.0$: From top to bottom the first plot is v, then ϕ, u, q and DLB.

5.4 Qualitative Comparison of the Control Variables

We now qualitatively compare the control variables. We do this by transforming selected model fields into vorticity and PV-based variables (here the PV-based variables have not been approximated). The model fields are taken from model runs in a high Burger regime ($B_u = 4.0$) and a low Burger regime ($B_u = 0.2$). We recall the results of [67]: the vorticity-based variables are a good approximation to balance in high Burger regimes and therefore the sets of control variables should appear to be similar. In a low Burger regime the PV is approximated by height and so the vorticity-based variables should not accurately represent the balance variable, and the vorticity and PV-based variables should be different.
The results of the experiment with $B_u = 4.0$ are shown in figures 5.3, for the vorticity-based variables, and 5.4, for the PV-based variables. The original model fields are shown in figure 5.2 where in descending order v, ϕ, u are plotted with the

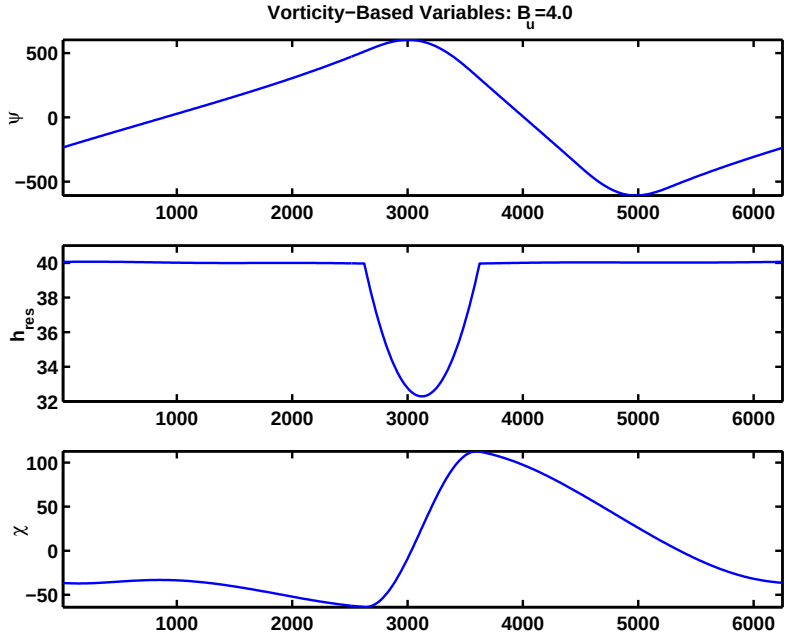

Figure 5.3: Vorticity-based variable, $B_u = 4.0$: From top to bottom the first plot is ψ, then h_{res} and χ.

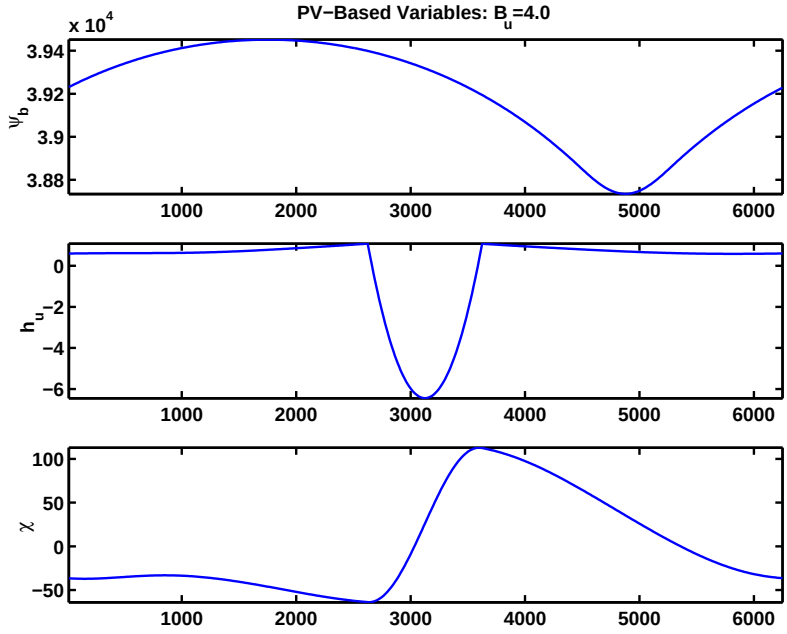

Figure 5.4: PV-based variable, $B_u = 4.0$: From top to bottom the first plot is ψ_b, then h_u and χ.

PV, q, and the departure from linear balance (DLB). Here we see that the fields are in balance, i.e.

$$DLB = fv - g\frac{\partial h}{\partial x} \approx 0,$$

everywhere except the area over the central orography defined as in equation (4.32). The two sets of control variables appear very similar (χ is the same variable in both cases). There are some differences in the form of ψ and ψ_b and also h_{res} and h_u but these are slight. We see that the unbalanced height variable in both cases has the disturbance correctly located over the central orography. The balanced variables are related to the PV, with minimum located at the point of the PV maximum, though the vorticity-based balanced variable ψ has its minimum slightly to the right of the PV minimum. We notice that the balanced variables have horizontal scales larger than the unbalanced variables.

The results of the experiment with $B_u = 0.2$ are shown in figures 5.6, for the vorticity-based variables, and 5.7, for the PV-based variables. The original model fields are shown in figure 5.5 where in descending order v, ϕ, u are plotted with the q and the departure from linear balance (DLB) as before. Again we see that the domain is approximately in balance ($DLB \approx 0$) everywhere apart from the area over the central orography. This time, however, the two sets of control variables appear very different (except χ which is the same variable in both cases). We see that the balanced PV-based variable ψ_b is very similar to the original ϕ field and the reciprocal of the PV field, which agrees with the theoretical result that the PV is dominated by height in this regime. The unbalanced component of ϕ is not present in ψ_b. The assumed 'balanced' vorticity-based variable ψ has a spurious signal over the central orography, which does not correspond to the location of the PV and, in fact, is in the unbalanced region. The unbalanced PV-based variable, h_u, is in agreement with the DLB; it has a signal over the central orography and is approximately zero elsewhere. The vorticity-based variable h_{res} has a signal across the whole domain, which does not agree with the imbalance present in the original model fields.

Finally we note the mean values of the control variables. For the vorticity-based

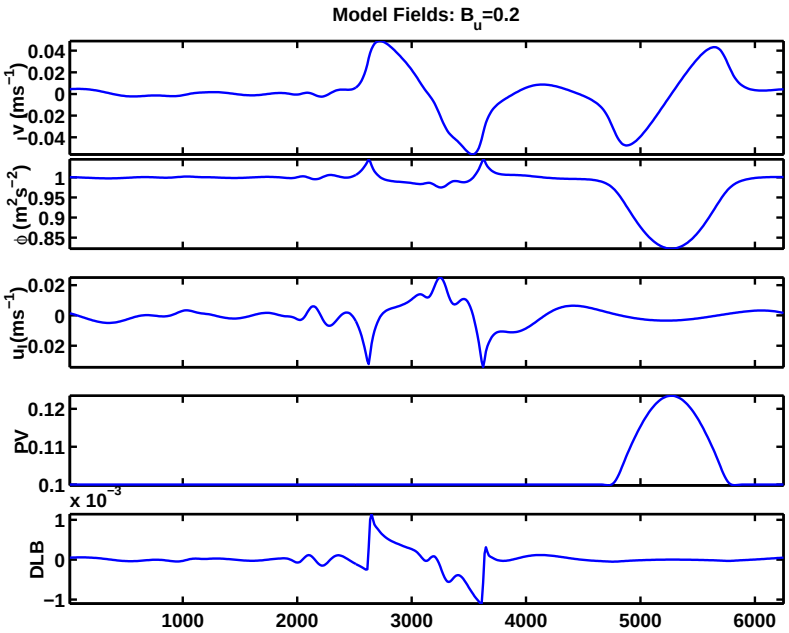

Figure 5.5: Model fields, $B_u = 0.2$: From top to bottom the first plot is v, then ϕ, u, q and DLB.

variables we saw in section 5.1.1 that the mean of ϕ, $< \phi >$, is retained totally in the mean of h_{res}, and ψ has a zero mean. In section 5.2.1 we see that the PV-based variables have $< \phi >$ split between ψ_b and h_u. The unbalanced height retains only the fraction of the mean of ϕ that is required such that the mean of $\bar{q}h_u$ is zero. The majority of $< \phi >$ is retained within ψ_b and multiplied by f^{-1}, thus it is 100 times larger.

The results of both experiments are in agreement with the theory presented in [67]. In this section the transforms were applied to full model fields, which enabled us to perform an initial comparison of the control variables. However, in the incremental 4D VAR the transforms will only be applied tp increments. In the next chapter we attempt to test the transforms quantitatively by looking at the correlation of balanced and unbalanced variables.

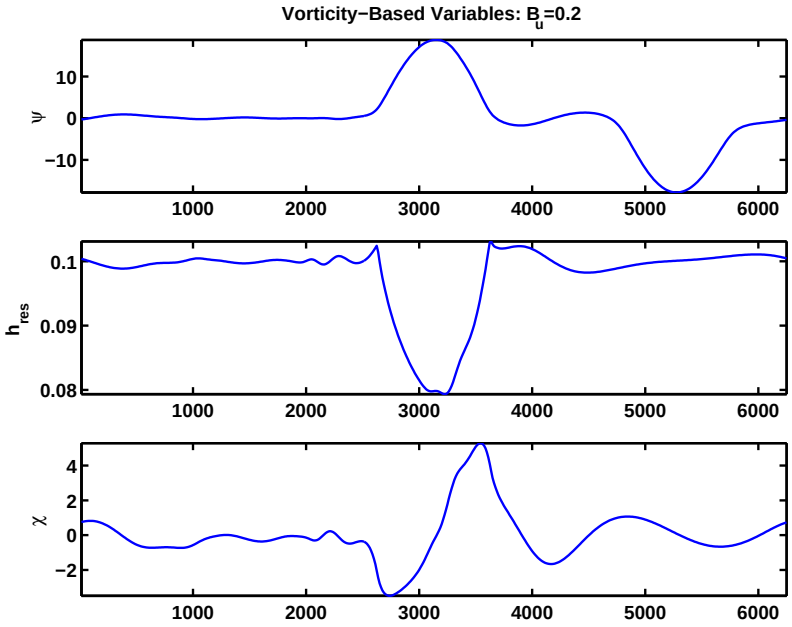

Figure 5.6: Vorticity-based variable, $B_u = 0.2$: From top to bottom the first plot is ψ, then h_{res} and χ.

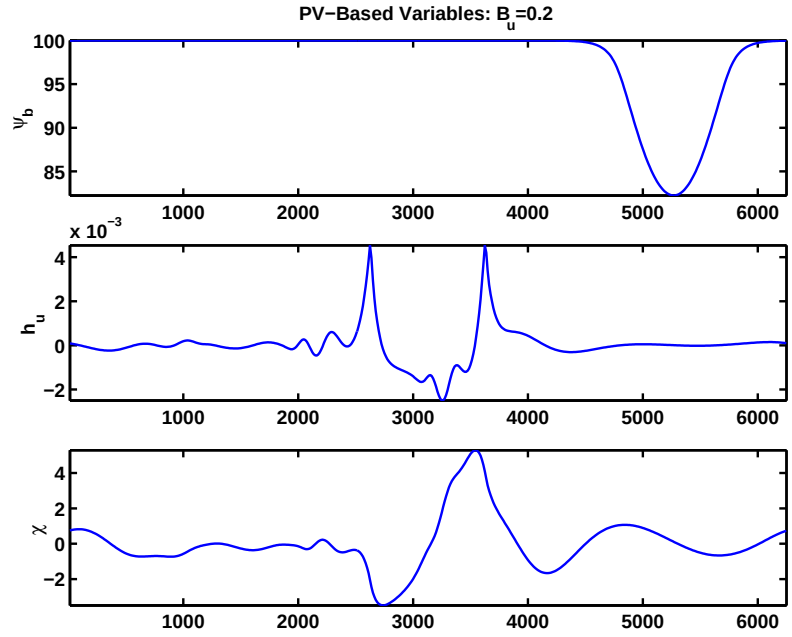

Figure 5.7: PV-based variable, $B_u = 0.2$: From top to bottom the first plot is ψ_b, then h_u and χ.

5.5 Summary

In this chapter we derived the vorticity and PV-based transforms for the 1D SWEs. We then discussed their numerical implementation. Some conditions are found that need to be handled carefully in the implementation, the most important of which is solving the equation for ψ_u in the PV-based transform, equation (5.56). The solution requires that h_u is such that the mean of $\bar{q}h_u$ is zero. We propose a solution to this problem but we also note that a convenient approximation to the PV-based transform would involve approximating the linearised PV by

$$\bar{q} \approx \frac{f}{<\bar{h}>}. \tag{5.61}$$

We can then constrain the mean of h_u to be zero. In the next chapter we consider the effect of this type of approximation.

After describing the numerical tests that were carried out to verify the transforms, we make a qualitative comparison of the control variables in different Burger regimes. The results of this experiment are in agreement with the theory described in previous chapters: We find that the vorticity-based and PV-based control variables are similar in a high Burger regime, but are different in a low Burger regime, and the PV-based variables appear to capture the balance correctly in both regimes. The vorticity-based transform on the other hand does not correctly represent the balanced flow.

In the next chapter we try to quantify the difference in the transforms by looking at the correlation of the variables in different Burger regimes.

Chapter 6

Statistical Experiments

In the previous chapter we demonstrated the qualitative behaviour of the vorticity and PV-based variables as we change Burger regime. This is in agreement with the theory. In this chapter we present new results that quantify the success of each transform in removing cross-correlations. Specifically we address the main aim of this research: How accurate is the fundamental assumption that the control variables are uncorrelated? And further what affect does approximating the PV-based transform have on the correlations between the balanced and unbalanced PV-based variables?

To do this we examine the correlation of the control variables in both high and low Burger regimes. The method we choose to test the correlation of the control variables is that used in [51], which we will call the 'quick covs method' (QCM) and was introduced in section 3.4.1. Whilst this is a relatively simple method it is found to be effective for initial determination of forecast error statistics if used carefully.

Our new results show that the whilst the PV-based variables remain uncorrelated in all regimes tested, the vorticity-based variables often show significant correlation. We are therefore able to conclude that the vorticity-based variables are not a valid choice of control variable at low Burger number.

The QCM is also used to generate auto-correlations for each control variable. The auto-correlations provide, amongst other things, a spatial length scale for each

control variable. This length scale tells us how far background information should be spread for each variable. We can then interpret this length scale as further evidence that the PV-based variables are valid across regimes. Finally, the auto-correlations are then modelled and implemented into the block diagonal components of the cost function.

We start by deriving the background error statistics implied by each transform. This highlights a key difference between the vorticity and the PV-based transforms; that the PV-based transform introduces state-dependence into the implied background statistics. We also note the importance of the auto-correlations.

6.1 Implied Background Statistics

Having derived the vorticity and PV-based transforms for the simplified SWEs we are now able to find the background error statistics implied by each control variable transform. We do this by applying only the parameter transform (defined in section 3.3.2) and then assume that the control variables are now uncorrelated with each other. So equation (3.8) becomes

$$\mathbf{U}^T\mathbf{B}^{-1}\mathbf{U} = \mathbf{\Lambda}^{-1}, \tag{6.1}$$

where $\mathbf{\Lambda}$ is a block diagonal matrix of auto-correlations. Rearranging the above equation gives

$$\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T. \tag{6.2}$$

We can write the background term of the cost function, defined by equation (3.2), for the simplified SWEs as

$$\begin{aligned} 2\widetilde{J}_b^{(k)}[\mathbf{x}_0'^{(k)}] &= (\mathbf{x}_0'^{(k)} - \mathbf{x}'^b)^T\mathbf{B}^{-1}(\mathbf{x}_0'^{(k)} - \mathbf{x}'^b) \\ &= (\delta\mathbf{u}, \delta\mathbf{v}, \delta\mathbf{h})(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T)^{-1} \begin{pmatrix} \delta\mathbf{u} \\ \delta\mathbf{v} \\ \delta\mathbf{h} \end{pmatrix} \end{aligned}$$

$$= (\delta\mathbf{u}, \delta\mathbf{v}, \delta\mathbf{h}) \begin{pmatrix} \mathbf{A}_{(u,u)} & \mathbf{C}_{(u,v)} & \mathbf{C}_{(u,h)} \\ \mathbf{C}_{(v,u)} & \mathbf{A}_{(v,v)} & \mathbf{C}_{(v,h)} \\ \mathbf{C}_{(h,u)} & \mathbf{C}_{(h,v)} & \mathbf{A}_{(h,h)} \end{pmatrix}^{-1} \begin{pmatrix} \delta\mathbf{u} \\ \delta\mathbf{v} \\ \delta\mathbf{h} \end{pmatrix} \tag{6.3}$$

where $\delta u = u_0'^{(k)} - u'^b$ etc., the block matrices $\mathbf{A}$ specify the auto-covariances and the block matrices $\mathbf{C}$ specify the implied covariances.

In the following analysis we find the background error covariance statistics that are implied by each control variable transform. For ease of notation we do not consider the mean values of the increments u' and v' although they are accounted for in the calculations.

6.1.1 Implied Background Error Covariance Statistics

For the vorticity-based control variables $\mathbf{U}_V$ is defined by equations (5.11) to (5.13). The transform can be written in matrix notation as

$$\begin{pmatrix} u' \\ v' \\ h' \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial x} & 0 \\ 0 & \frac{f}{g} & 1 \end{pmatrix} \begin{pmatrix} \chi' \\ \psi' \\ h'_{res} \end{pmatrix}.$$

Therefore the background error covariance matrix for model variables u', v' and h' implied by the transform is

$$\mathbf{B}_V = \mathbf{U}_V \mathbf{\Lambda}_V \mathbf{U}_V^T = \begin{pmatrix} \left(\frac{\partial}{\partial x}\right) \mathbf{\Lambda}_\chi \left(\frac{\partial}{\partial x}\right)^T & 0 & 0 \\ 0 & \left(\frac{\partial}{\partial x}\right) \mathbf{\Lambda}_\psi \left(\frac{\partial}{\partial x}\right)^T & \left(\frac{\partial}{\partial x}\right) \mathbf{\Lambda}_\psi \frac{f}{g} \\ 0 & \left(\left(\frac{\partial}{\partial x}\right) \mathbf{\Lambda}_\psi \frac{f}{g}\right)^T & \left(\frac{f}{g}\right)^2 \mathbf{\Lambda}_\psi + \mathbf{\Lambda}_{h_r} \end{pmatrix},$$

where $\mathbf{\Lambda}_\chi, \mathbf{\Lambda}_\psi$ and $\mathbf{\Lambda}_{h_r}$ are the auto-covariance matrices for χ', ψ' and h'_{res} respectively.

For PV-based control variables $\mathbf{U}_{PV}$ is given by equations (5.40) to (5.42). Writing the transform in matrix notation implies

$$\begin{pmatrix} u' \\ v' \\ h' \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial x} & \mathbf{Q} \\ 0 & \frac{f}{g} & 1 \end{pmatrix} \begin{pmatrix} \chi' \\ \psi'_b \\ h'_u \end{pmatrix}.$$

So the PV-based control variables imply a background error covariance matrix

$$\mathbf{B}_{PV} = \begin{pmatrix} \left(\frac{\partial}{\partial x}\right)\mathbf{\Lambda}_{\chi}\left(\frac{\partial}{\partial x}\right)^{T} & 0 & 0 \\ 0 & \left(\frac{\partial}{\partial x}\right)\mathbf{\Lambda}_{\psi_b}\left(\frac{\partial}{\partial x}\right)^{T} + \mathbf{Q}\mathbf{\Lambda}_{h_u}\mathbf{Q}^{T} & \left(\frac{\partial}{\partial x}\right)\mathbf{\Lambda}_{\psi_b}\frac{f}{g} + \mathbf{Q}\mathbf{\Lambda}_{h_u} \\ 0 & \left(\left(\frac{\partial}{\partial x}\right)\mathbf{\Lambda}_{\psi_b}\frac{f}{g} + \mathbf{Q}\mathbf{\Lambda}_{h_u}\right)^{T} & \left(\frac{f}{g}\right)^{2}\mathbf{\Lambda}_{\psi_b} + \mathbf{\Lambda}_{h_u} \end{pmatrix},$$

where the operator $\mathbf{Q}$ is given by

$$\mathbf{Q} = \left(\frac{\partial}{\partial x}\left(\nabla^{-2}\bar{q}\ \cdot\right)\right)$$

with $\nabla^2 \equiv \frac{\partial^2}{\partial x^2}$, and $\mathbf{\Lambda}_{\chi}, \mathbf{\Lambda}_{\psi_b}$ and $\mathbf{\Lambda}_{h_u}$ are the auto-covariance matrices for χ', ψ_b' and h_u' respectively.

The first thing we notice in $\mathbf{B}_V$ and $\mathbf{B}_{PV}$ is that there are no implied covariances between u' and other model variables. This de-coupling of u is a result of assuming that there is no variation in the y direction in the 1D SWEs model and hence the divergence depends solely on u. Implied covariance with u could be introduced by splitting the velocity in the same way as is done with v. We would then need a relationship defining a balanced component of u. This is not considered at present.

A comparison of $\mathbf{B}_V$ and $\mathbf{B}_{PV}$ shows that the differences between the implied background error statistics are the covariance $\mathbf{C}_{(v,h)}$, $\mathbf{C}_{(h,v)}$ and $\mathbf{A}_{(v,v)}$. The statistical model is more complicated for the PV-based variables and whilst $\mathbf{B}_V$ is static, the PV-based implied background error statistics include the reference state PV, $\bar{q}$. Thus the PV-based transforms have introduced state-dependence into the implied background error statistics by the inclusion of $\bar{q}$. The state-dependence introduces a mechanism to change the implied background statistics each outer iteration of the incremental 4D VAR algorithm through the linearisation state. This is something that is not possible if we use the vorticity-based variables. We note that some state dependence is also included in the new ECMWF transform [20]. This is achieved through the use of a non-linear balance operator, which is linearised for use in their incremental 4D VAR system.

6.1.2 Summary

By visualising the implied background error covariance matrices B_V and B_{PV} in this way we are able to see the different implications that the vorticity and PV-based transforms have on the data assimilation. Most importantly, that through the reference state PV we are able to introduce state dependence in the implied background error statistics of the PV-based control variables.
We also highlighted the importance of the auto-covariance matrix $\mathbf{\Lambda}$. In the final section of this chapter we generate these auto-covariances for each of our control variables. We observe that these auto-covariance structures also change with regime.
At the end of section 5.2 an approximation

$$\bar{q} = f / < \bar{h} >$$

to the PV-based transform was proposed. This approximation would greatly simplify the implementation of the U-transform. In the following section we try to identify the consequences of this approximation. We now introduce the statistical method we choose to investigate the correlation of the balanced and unbalanced vorticity and PV-based control variables.

6.2 The Correlation of Control Variables

It is assumed in the data assimilation that the control variables are uncorrelated. We now investigate the validity of this assumption. There are several questions we aim to address:

1. Are the balanced and unbalanced components of the flow uncorrelated?

   This is true in the linear system but how true is this assumption for the non-linear model?

2. Is the PV-based transform more successful at removing cross-correlations than the vorticity-based transform?

We have demonstrated some theoretical arguments in the previous chapters but this hypothesis is yet to be verified experimentally.

3. What are the consequences of using an approximate $\bar{q} = f/ < \bar{h} >$ in the PV-based transform?

   We can attempt to assess this by comparing the correlation between the approximated PV-based variables and the correlation between the variables when the full linearised PV is used.

We look at correlations of the vorticity and PV-based control variables where we fix the Burger number to be either high or low whilst varying the Rossby number. Before these questions can be addressed we describe the 'quick covs method' (QCM) used to generate the correlations.

6.2.1 Statistical Method

The background field in operational centres is a previous short forecast. Therefore a natural assumption is that background errors will be similar to forecast errors. There are many ways to obtain forecast error statistical data as discussed in section 3.4.1. A popular example is the NMC method, as described in [47]. This method requires an operational data assimilation system. Therefore for initial analysis we choose the much simpler approach proposed in [51] and introduced in 3.4.1. This method is sometimes referred to as the 'quick covs method' (QCM). Essentially, the QCM assumes that background error statistics are similar to forecast error statistics. This assumption is also made by the AEM and NMC methods discussed in section 3.4.1. However, whereas the NMC method assumes forecast error statistics are similar to the differences between 24 and 48-hour forecasts valid at the same time, the QCM assumes that forecast error statistics are similar to the statistics of differences between forecasts at different times. Hence, errors are assumed to have statistics similar to those of the time derivatives of each variable. Therefore, in the absence of an operational data assimilation

system, forecast differences can be used as an initial source for forecast error data, and so a proxy for background error statistics [51].

Here we generate a single long forecast using our model. We then take forecast differences at regular intervals of time apart. This is done until we have a data set of multiple time differences. Using these differences we may look at the correlations of the difference fields of each variable. We can examine the correlation between control variables by transforming the forecast time-differences using the relevant T transform.

We start by considering a time difference $\boldsymbol{\psi}'^k$, say, of field ψ at time t_k with

$$\boldsymbol{\psi}'^k = \boldsymbol{\psi}^{t_k+\tau} - \boldsymbol{\psi}^{t_k} = \left(\psi_1^{t_k+\tau} - \psi_1^{t_k}, \ldots, \psi_N^{t_k+\tau} - \psi_N^{t_k}\right)^T = \left(\psi'^k_1, \ldots, \psi'^k_N\right)^T,$$

where t_k is some time in a model forecast, τ is one time-differencing interval later, N is the number of points in the domain and the subscript indicates the grid point. The interval τ remains to be chosen. We now consider two time difference values ψ'^k_i, for psi, and h'^k_i, for h, at $x = i\Delta x$ and time $t = t_k$. We treat these values as one realisation of two random variables $\mathbf{\Psi}$ and $\mathbf{h}$. At time t_k we therefore have N realisations $\mathbf{\Psi}$ and $\mathbf{h}$ with

$$\Psi^k = \left\{\psi'^k_1, \ldots, \psi'^k_N\right\}$$

and

$$\mathbf{h}^k = \left\{h'^k_1, \ldots, h'^k_N\right\}.$$

If we now assume that we have M sets of time differences, i.e. $k = 1, \ldots, M$, of ψ and h for times $t_1, \ldots, t_M$ we then have

$$\mathbf{\Psi} = \left\{\Psi^1, \ldots, \Psi^M\right\},$$

and

$$\mathbf{h} = \left\{\mathbf{h}^1, \ldots, \mathbf{h}^M\right\}.$$

This gives a total of $N \times M$ realisations of $\mathbf{\Psi}$ and $\mathbf{h}$. We can now calculate their covariance. Writing $\mathbf{\Psi}$ and $\mathbf{h}$ in full gives

$$\mathbf{\Psi} = \left\{\psi'^1_1, \ldots, \psi'^1_N, \psi'^2_1, \ldots, \psi'^2_N, \ldots\ldots, \psi'^M_1, \ldots, \psi'^M_N\right\}, \tag{6.4}$$

and

$$\mathbf{h} = \left\{ {h'}^1_1, \ldots, {h'}^1_N, {h'}^2_1, \ldots, {h'}^2_N, \ldots\ldots, {h'}^M_1, \ldots, {h'}^M_N \right\}, \tag{6.5}$$

where the subscript indicates the grid point and the superscript the time-difference field. We now drop the primes for convenience. Their covariance is then given by

$$COV(\mathbf{\Psi}, \mathbf{h}) = < (\mathbf{\Psi} - < \mathbf{\Psi} >)(\mathbf{h} - < \mathbf{h} >) > = < \mathbf{\Psi h} > - < \mathbf{\Psi} >< \mathbf{h} >, \tag{6.6}$$

where

$$< \mathbf{\Psi h} > = \frac{(\psi^1_1 h^1_1 + \ldots + \psi^M_N h^M_N)}{N \times M}, \tag{6.7}$$

and $< \cdot >$ represents the mean value. Therefore $< \mathbf{h} >$ is given by

$$< \mathbf{h} > = \frac{h^1_1 + \ldots + h^M_N}{N \times M}, \tag{6.8}$$

and

$$< \mathbf{\Psi} > = \frac{\psi^1_1 + \ldots + \psi^M_N}{N \times M}. \tag{6.9}$$

We then calculate the correlation coefficient

$$\rho = \frac{COV(\mathbf{\Psi}, \mathbf{h})}{\sigma_\psi \sigma_h}, \tag{6.10}$$

where the σ_ψ and σ_h represent the standard deviations of $\mathbf{\Psi}$ and $\mathbf{h}$, and

$$\sigma_h = \sqrt{(< (\mathbf{h} - < \mathbf{h} >)^2 >)} = \sqrt{(< \mathbf{h}^2 > - < \mathbf{h} >^2)}, \tag{6.11}$$

with

$$< \mathbf{h}^2 > = \frac{\mathbf{h}^2_1 + \ldots + \mathbf{h}^2_{N \times M}}{N \times M}. \tag{6.12}$$

The standard deviation of $\mathbf{\Psi}$, σ_ψ, is calculated in the same way.

The correlation coefficient varies such that

$$-1 \leq \rho \leq 1,$$

with ρ close to 1 or -1 indicating strong positive or negative correlation. A value of ρ close to 0 indicates variables are uncorrelated.

Experiment Details

Correlation coefficients are calculated in the same way for (ψ', h'_{res}), (ψ'_b, h'_u), $(\psi'_b, h'_u)_{approx}$ and $(\psi, h + \widetilde{H})$, where $(\psi'_b, h'_u)_{approx}$ are the PV-based variables when the approximation $\bar{q} = f/ < \bar{h} >$ is made and ψ and $h + \widetilde{H}$ are full model fields. In the following correlation experiments the model is set up as follows. The orography is given by

$$\widetilde{H}(x) = H_c \left(1 - \frac{x^2}{a^2}\right), \tag{6.13}$$

with a and H_c specified in Table 6.1 below, along with other model parameters. The spatial and temporal grid spacing is chosen so that the sufficient condition for convergence of the displacement iteration, equation (4.22), in the numerical scheme should always be satisfied. For the grid spacing given in table 6.1 condition (4.22) becomes

$$|u_x| \, \Delta t = |u_{i+1} - u_i| \, \frac{\Delta t}{\Delta x} = |u_{i+1} - u_i| \, \frac{2.5}{12.5} < 1, \tag{6.14}$$

or

$$|u_{i+1} - u_i| < 5. \tag{6.15}$$

Thus, for the initial conditions and flow regimes in the following experiments, using this choice of grid spacing should mean that the sufficient condition is always met. We must now choose the time-differencing interval τ. The QCM can be very sensitive to the length of the time-difference interval used. In the NMC method the interval is one day to avoid contamination by the diurnal signal, which will otherwise dominate the error statistics. In our model the dominant signal will be different and we need to make sure that we have identified the cause and choose the time interval appropriately. This is analysed in the following section, here we set $U_c = 0.5ms^{-1}$ in the high Burger regime and $U_c = 2.5ms^{-1}$ the low Burger regime.

Identifying the Dominant Signal: High Burger Regime.

The QCM is very sensitive to a dominant signal in the forecast differences. We therefore need to identify if there is a dominant signal in our data and remove it if

Parameter		Value
Number of Grid Points	N	500
Grid Spacing	Δx	$12.5m$
Time Step	Δt	$2.5m$
Coriolis Parameter	f	$0.01s^{-1}$
Width of Orography	a	$40\Delta x$
Maximum Height of Orography (Low Burger)	H_c	$0.019m$
Maximum Height of Orography (High Burger)	H_c	$7.6m$
Height of Surface at Rest (Low Burger)	D	$0.1m$
Height of Surface at Rest (High Burger)	D	$40.0m$

Table 6.1: Experiment Set-Up.

it is not of relevance. In [51] the interval is 6 hours since the background state is a 6 hour forecast. However, the time-differences are then adjusted to account for diurnal changes, which would otherwise corrupt the data.

To identify the dominant oscillation in the high Burger case we generate a forecast and plot values of each variable at fixed points in space against time. Plots for u are shown in figure 6.1 for values of u_i at point x_i with $i = 125, 250, 375$ and 500, where $x_i = i\Delta x$. The plots for v and ϕ are for the same points in space and shown in figures 6.2 and 6.3. For u we see a fast oscillation with a period of approximately $300s$. We see a corresponding fast oscillation in ϕ with the same period. This fast oscillation in ϕ is combined with a slow motion that is coupled with v.

This would suggest that there is a gravity wave associated with the fast motion and the slow motion is the balanced flow associated with the advection of the PV. From figures 6.2 and 6.3 we are able to locate the PV maximum by the minimum of ϕ, and the mid-point between the minimum and maximum of v. This is advected across the entire domain in a approximately $12,500s$. Therefore its speed is approximately $N\Delta x/12500 = 0.5ms^{-1}$; the forcing flow in this experiment is $U_c = 0.5ms^{-1}$.

The gravity wave speed, c_g, defined by equations (2.19) in this experiment is

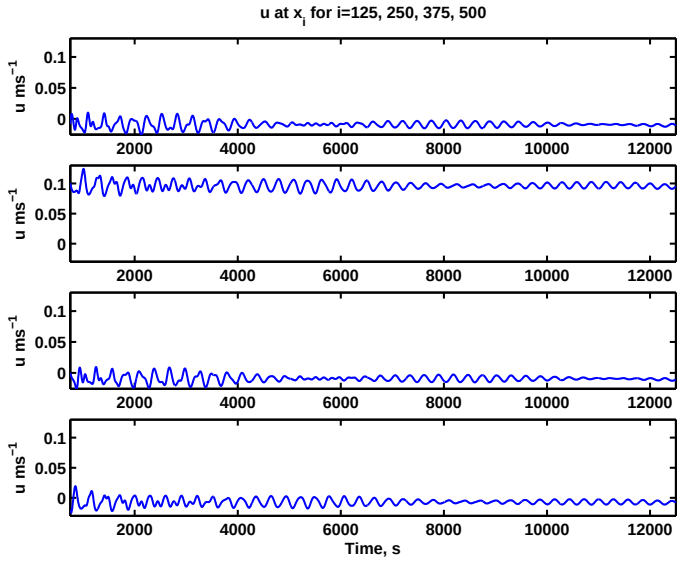

Figure 6.1: Values of u at fixed points in space plotted against time for $B_u = 4.0$. Top plot is value of u at point $x_{125} = 125\Delta x$, then following plots, in descending order, are values of u at points x_{250}, x_{375} and x_{500}.

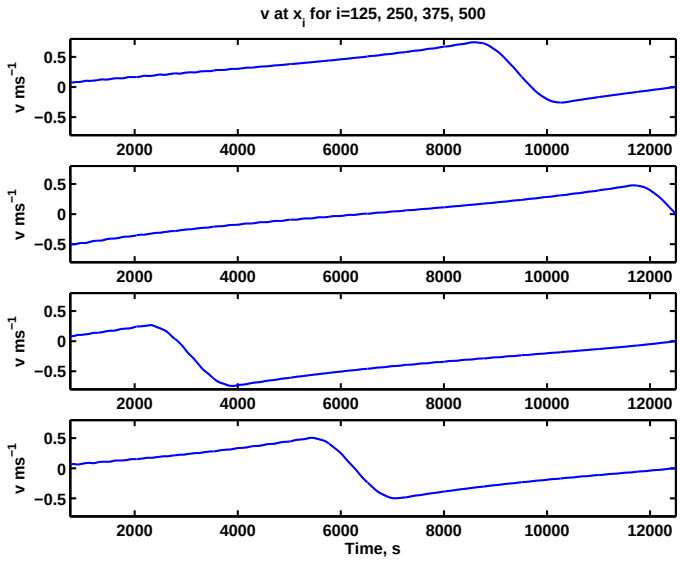

Figure 6.2: Values of v at fixed points in space plotted against time for $B_u = 4.0$.. Top plot is value of v at point $x_{125} = 125\Delta x$, then following plots, in descending order, are values of v at points x_{250}, x_{375} and x_{500}.

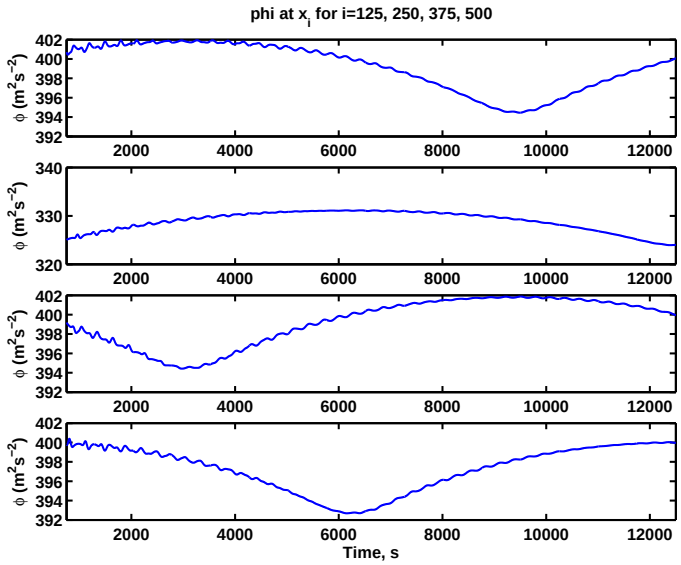

Figure 6.3: Values of ϕ at fixed points in space plotted against time for $B_u = 4.0$. Top plot is value of ϕ at point $x_{125} = 125\Delta x$, then following plots, in descending order, are values of ϕ at points x_{250}, x_{375} and x_{500}.

$c_g = \sqrt{gD} \approx 20ms^{-1}$, where D is the mean depth. Gravity waves in the positive x direction therefore cover the length of the domain in time $\frac{N\Delta x}{(\sqrt{gD}+U_c)} \approx 300s$ (accounting for the forcing flow U_c), the approximate frequency of the fast signal observed in u, v and ϕ. This gravity wave periodicity is the likely cause of this signal. The oscillation is a product of the periodicity of this particular problem. We therefore choose to filter the signal from our data as we are interested in the correlation of the control variables. We are able to filter this signal by choosing a time interval τ of approximately $300s$.

We can provide additional evidence that this choice of time-differencing interval is appropriate by calculating correlations of each variable in time. These results are given in appendix A.1. We expect to see the correlations decaying with time but this may not be the case if there is an oscillation that is dominating the data.

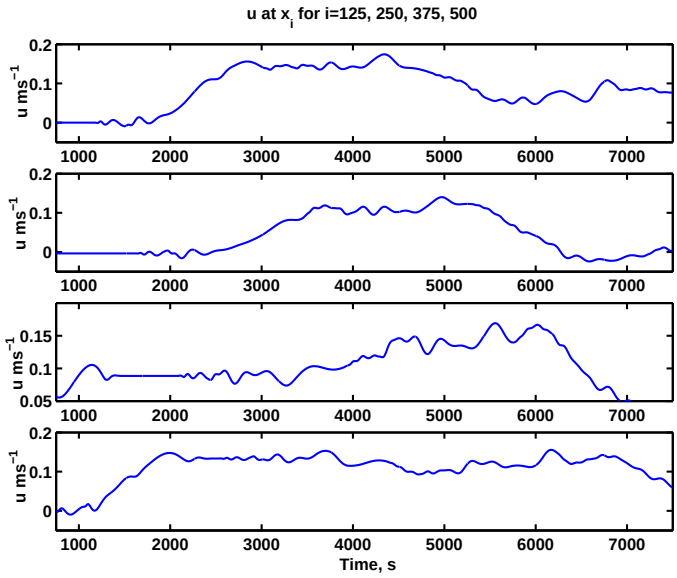

Figure 6.4: Values of u at fixed points in space plotted against time for $B_u = 0.2$. Top plot is value of u at point $x_{125} = 125\Delta x$, then following plots, in descending order, are values of u at points x_{250}, x_{375} and x_{500}.

Identifying the Dominant Signal: Low Burger Regime.

In the low Burger regime the gravity wave speed is much slower. The gravity wave speed $c_g = \sqrt{Dg} \approx \sqrt{0.1g} \approx 1ms^{-1}$, where D is the mean depth. Gravity waves in the positive x direction therefore cover the length of the domain in time $\frac{N\Delta x}{(\sqrt{gD}+U_c)}s \approx 1790s$, taking $U_c = 2.5ms^{-1}$. This time is much longer than in the high Burger regime and therefore our correlation results should not be affected by this signal. To verify this we again plot point values of the model variables as was done in the high Burger case, see figures 6.4, 6.5 and 6.6 for plots of u, v and ϕ respectively. In the plots for v and ϕ we again see a slow motion associated with the advection of the PV with $U_c = 2.5ms^{-1}$.

There is not an obvious oscillation in any of the plots as was the case with the high Burger experiments. There is no periodic signal from the gravity wave since the speed is much slower. Therefore we are free to choose the time differencing interval in the low Burger regime. We choose the interval to be the same as in the high Burger experiments. Using this interval we can check if it is appropriate by looking

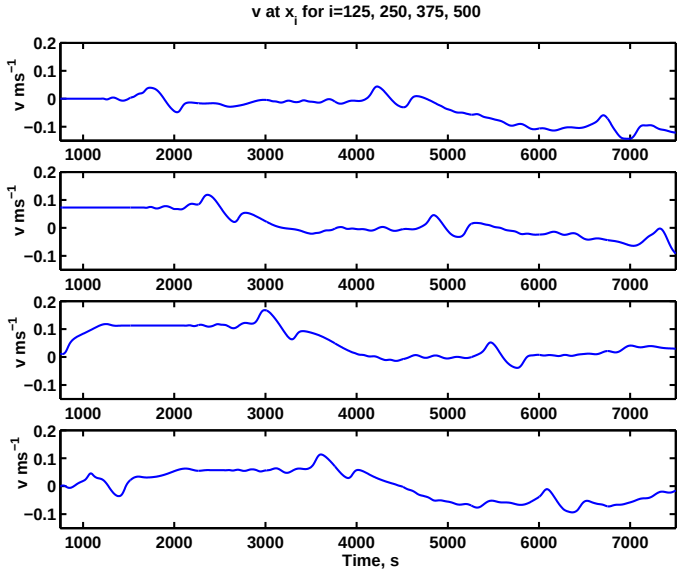

Figure 6.5: Values of v at fixed points in space plotted against time for $B_u = 0.2$. Top plot is value of v at point $x_{125} = 125\Delta x$, then following plots, in descending order, are values of v at points x_{250}, x_{375} and x_{500}.

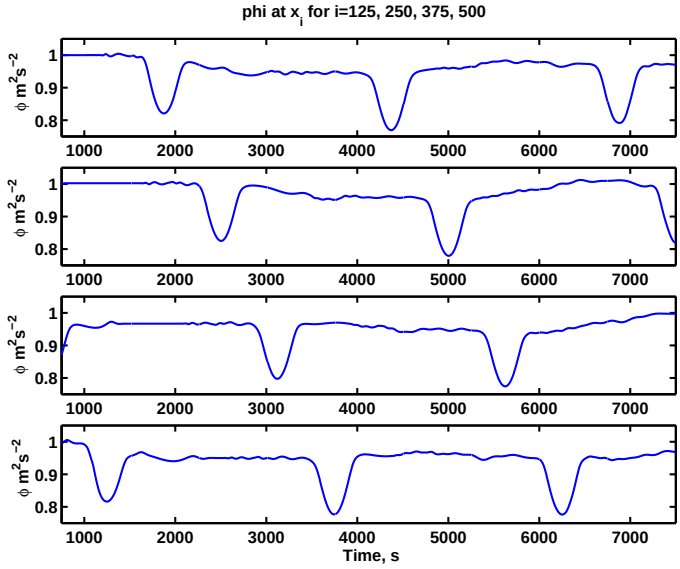

Figure 6.6: Values of ϕ at fixed points in space plotted against time for $B_u = 0.2$. Top plot is value of ϕ at point $x_{125} = 125\Delta x$, then following plots, in descending order, are values of ϕ at points x_{250}, x_{375} and x_{500}.

again at the time-correlations, see appendix A.2. We see that the time-correlations decay as we would expect and do not appear corrupted by the gravity wave signal.

6.2.2 Correlation Results

We now present the main result of our research. The experiments are run using the method introduced at the start of 6.2.1. We use a time differencing interval appropriate to both the high and low Burger regimes. We first look at correlations in a high Burger regime where the Rossby number is varied between 0.02 and 1.0.

Correlation Experiment Results: High Burger Regime

Plotted in figure 6.7 are the correlation coefficients, calculated from equation (6.10), between full model fields $(\psi, h + \widetilde{H})$, model field time-differences (ψ', h'), vorticity-based variables (ψ', h'_{res}) and PV-based variables (ψ'_b, h'_u) where the full linearised PV is used and the approximation $q = f/ < \bar{h} >$ is used.
In this experiment the flow U_c has been varied between $0.5ms^{-1}$ and $5.0ms^{-1}$. We therefore choose a time-differencing interval of $\frac{N\Delta x}{(\sqrt{gH}+2.5)}s = 277.5s$ taking an approximate average mean flow $U_c = 2.5ms^{-1}$. By varying the mean flow U_c we change the Rossby number, $R_0 = U/fL$, and the Froude number, $F_r = \sqrt{gD}f$, where $U \approx U_c$, and L and D are characteristic length and fluid depth scales. The scale D and other parameters are given in table 6.1. We take the characteristic length scale $L = a$, the width of the orography. Therefore, with

$$0.1ms^{-1} < U < 5.0ms^{-1}$$

we have

$$0.02 < R_0 < 1.0$$

and

$$0.005 < F_r < 0.25.$$

In figure 6.7 we observe the following:

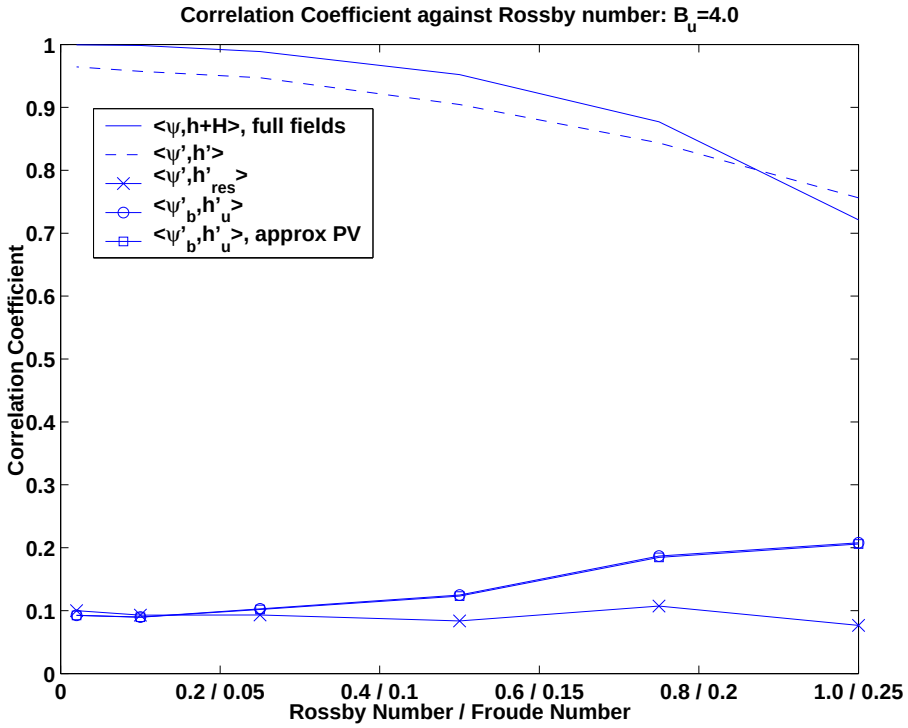

Figure 6.7: Time-differencing interval 277.5s: Plot of correlation coefficient against Rossby number for $B_u = 4.0$. The solid line is the correlation for full model field ψ and $h + \widetilde{H}$, the dashed line for model field time differences ψ' and h'. Vorticity-based control variable correlations ψ' with h'_{res} are indicated with the crosses and PV-based variables ψ'_b with h'_u using the full $\bar{q}$ are circles and the approximate $\bar{q} = f/ < \bar{h} >$, squares.

1. A strong correlation between model variables in the full fields $(\psi, h + \widetilde{H})$ and the difference fields (ψ', h'), which decreases as the Rossby number increases.

   This indicates a high degree of balance. The correlation increases as the Rossby number decreases and the flow becomes increasingly balanced. The balance relationship is less relevant as R_0 increases to 1.0, as we would expect from the analysis in section 2.2.1. This is because the balance approximation (2.89) is valid in the asymptotic limit of small Rossby number.

2. The correlation between the vorticity-based variables (ψ', h'_{res}) and between the PV-based variables (ψ'_b, h'_u) is much less than the model variables.

   Both control variable transforms have been successful in removing the strong correlation between model variables (ψ', h').

3. Correlation between vorticity-based variables (ψ', h'_{res}) and between the PV-based variables (ψ'_b, h'_u) is very similar.

   This is what we expect to see in a high Burger regime. The balanced variable is approximated well by the vorticity and therefore we should see similar correlations between the vorticity and PV-based variables.

4. The correlation between PV-based variables increases slightly as the Rossby number approaches 1.

   The correlation between vorticity-based variables stays relatively constant with changing R_0. The increase in the correlation between PV-variables could be attributed to the fact that as R_o increases the flow becomes increasingly non-linear. This type of behaviour is discussed in section 2.2.1 where, for Rossby numbers close to 1, the non-linear advective terms are no longer negligible. Thus the linear approximations used in the PV-based transform are becoming less accurate and this is reflected in the correlation of the PV-based variables. However, this increase is slight and the PV-based variables are still significantly less correlated than the model variables.

5. The correlation between PV-based variables when we use an approximate $\bar{q} = f/ < h >$ is almost exactly the same as when the full PV is used.

   The approximation therefore seems to be have little effect on the correlation between the PV-based variables.

Point 1, to some extent, gives validity to our choice of statistical method. The behaviour of the correlation between model full fields and increment fields is exactly as we would expect. As the Rossby number decreases we see an increase in correlation between these variables.

In previous sections we gave justification for the choice of the time-differencing interval $\tau = 277.5s$. To demonstrate that our choice of τ is appropriate to the high Burger regime, and that the QCM is sensitive to any dominate signal in the data, we can compare the results using different time intervals. We compare the time differencing interval $\tau = 277.5s$, figure 6.7, to whose when we use an interval that is not appropriate.

Figure 6.8 shows the correlation coefficients using $\tau = 100s$, which does not filter the gravity wave signal. The results in figure 6.8 do not correspond with the balances we understand to be present. There is very little balanced correlation in the time-differenced model fields ψ' and h'. However, the full model fields ψ and $h + \widetilde{H}$ are highly correlated. As the Rossby number is decreased the balance approximation becomes more accurate and therefore ψ and $h + \widetilde{H}$ become increasingly correlated. This behaviour is observed. The same should also be apparent in the time-difference fields. However, this is not observed in the correlations. This is because the gravity wave signal is dominating the correlations. An implication of the lack of correlation between ψ' and h' is that an unduly strong negative correlation in the vorticity-based variables, ψ' and h'_{res}, is produced. This is because h'_{res}, from equation (5.10), is given by

$$h'_{res} = h' - \frac{f}{g}\psi'$$

and since no correlation between ψ' and h' is found in the data this results in a negative correlation of ψ' and h'_{res} when the data is transformed using the

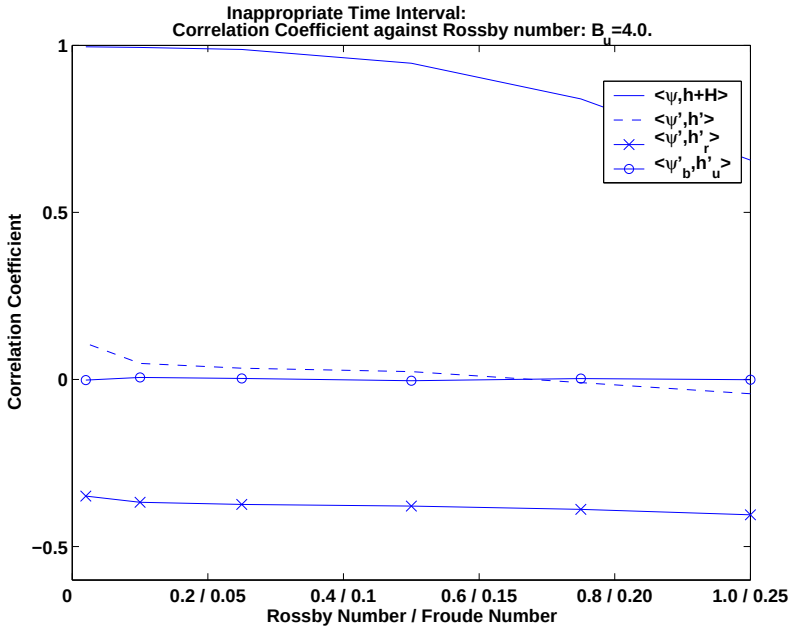

Figure 6.8: Time-differencing interval 100s: Plot of correlation coefficient against Rossby number for $B_u = 4.0$. The solid line is the correlation for full model field ψ and $h + \widetilde{H}$, the dashed line for model field time differences ψ' and h'. Vorticity-based control variable correlations ψ' with h'_{res} are indicated with the crosses and PV-based variables ψ'_b with h'_u using the full $\bar{q}$.

vorticity-based T transform. If we look at the results when the correct time filter is used, figure 6.7, we see correlations that are in line with our expectations. Correlations between full model fields and between time-differenced model increments agree almost exactly. We also observe similar correlations for the vorticity and PV-based variables. We can therefore be confident that the periodic motion of gravity waves around the domain does not corrupt these results.

Correlation Experiment Results: Low Burger Regime

In a low Burger regime we see the largest difference between the correlations of the vorticity and PV-based control variables. The correlation coefficients of $(\psi, h + \widetilde{H})$, (ψ', h'), (ψ', h'_{res}), (ψ'_b, h'_u) and $(\psi'_b, h'_u)_{approx}$ are shown in figure 6.9 for the low Burger number experiment. We observe the following:

1. Strong balanced correlation between model variables for small Rossby number.

   As the Rossby number decreases the correlations between the full fields increases as in the high Burger experiment. For Rossby number close to 1.0 we are not in a low Rossby number regime and we see that the flow is unbalanced. This is observed in the small correlation of ψ and $h + \widetilde{H}$. The correlation in the time-difference fields (ψ', h') remains relatively unaffected by changes in the Rossby number. This is attributed to a stationary wave in the forecast fields that is tied to the orography. This can be seen in figure 6.10. The stationary wave is not seen in the time-difference fields but does have a signal in the full field correlations, and further the behaviour of this standing wave is affected by the changes in Rossby number. This behaviour is not seen in the high Burger experiment as the flow is rotation dominated. Here the Froude number is larger than the Rossby number. The jump in the correlation of ψ and $h + \widetilde{H}$ corresponds to Rossby number becoming small and the Froude number decreasing below 1.

2. The PV-based variables are uncorrelated.

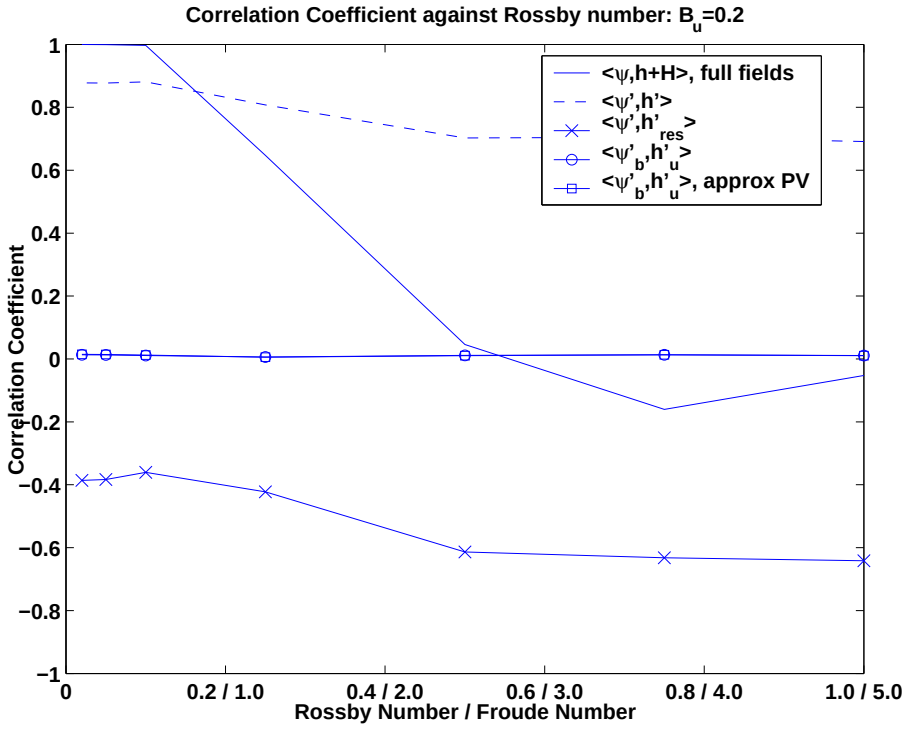

Figure 6.9: Plot of correlation coefficient against Rossby number for $B_u = 0.2$, time-differencing interval is 277.5s. The solid line is the correlation for full model field ψ and $h + \widetilde{H}$, the dashed line for model field time differences ψ' and h'. Vorticity-based control variable correlations ψ' with h'_{res} are indicated with the crosses and PV-based variables ψ'_b with h'_u using the full $\bar{q}$ are circles and the approximate $\bar{q} = f/ < \bar{h} >$, squares.

The transforms are successful in removing the correlation of the model variables and are therefore a valid choice of control variable in these regimes.

3. The vorticity-based variables show a strong negative correlation.

   The vorticity-based variables are therefore not a valid choice of control variable in a low Burger regime.

4. The correlation between the approximated PV-based variables are almost identical to when the full PV is used.

   The same behaviour is observed in the high Burger experiment.

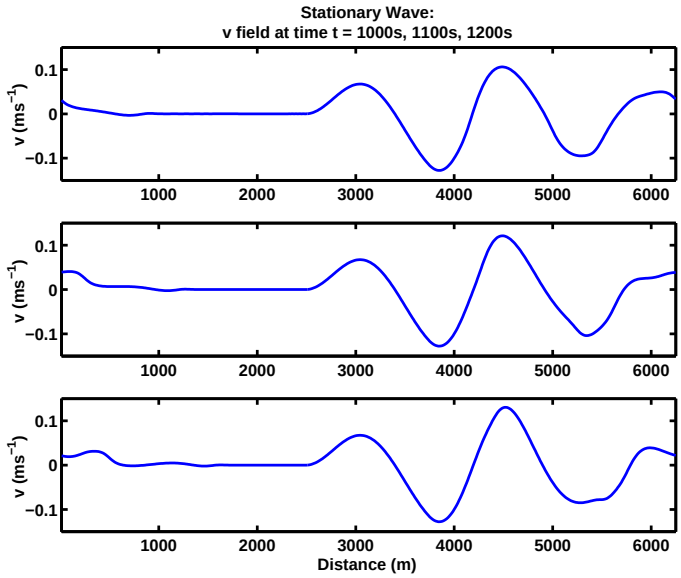

Figure 6.10: Plots show stationary wave in v for times one inertial period apart. The wave is tied to the central orography. This is not seen by the time-difference statistics but is seen in the statistics of the full model fields ψ and $h + \widetilde{H}$

The low Burger regime was found not to have a dominant oscillation. We therefore concluded that we are free to choose the time-differencing interval. To demonstrate that this is true and that these results are not sensitive to the interval we have plotted results using an interval of $100s$. These results are shown in figure 6.11. A comparison of figure 6.9 and figure 6.11 shows qualitatively very little difference in the correlations and our conclusions are the same.

6.2.3 Summary

These results demonstrate that the PV-based variables provide a much better choice of control variables than the current vorticity-based variables in a low Burger regime.

We have been able to validate our hypothesis that the PV-based variables are valid in both high and low Burger regimes, whilst the vorticity-based transforms fail in the low regime.

We also considered the practical implementation of the PV-based transform. We

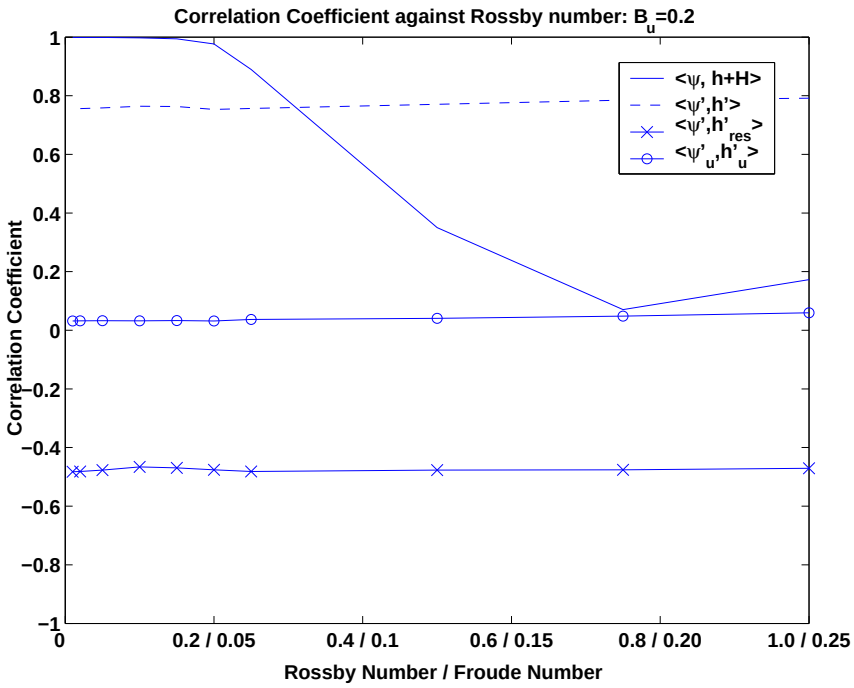

Figure 6.11: Plot of correlation coefficient against Rossby number for $B_u = 0.2$, time-differencing interval is 100s. The solid line is the correlation for full model field ψ and $h + \widetilde{H}$, the dashed line for model field time differences ψ' and h'. Vorticity-based control variable correlations ψ' with h'_{res} are indicated with the crosses and PV-based variables ψ'_b with h'_u using the full $\bar{q}$ are circles.

considered a possible approximation

$$\bar{q} = f / < \bar{h} >$$

in the PV-based transform. This approximation simplifies the transform. We demonstrated that this approximation makes little difference to the correlation between the variables in both the high and the low Burger regimes.

Before any correlation experiments were run we identified the dominant signal in our model. In the high Burger case this dominant signal was a product of the gravity wave interacting with the periodic boundary conditions. By choosing an appropriate time interval we are able to filter it producing good results. The effect of choosing the wrong time interval was demonstrated and produced clearly incorrect results. Thus demonstrating that this signal is undesirable in this case. We can therefore have confidence in our results.

In the low Burger experiments we were not able to identify a dominant signal. Therefore the choice of time-difference interval should not affect the correlation results. This we demonstrated by comparing the correlations calculated when very different time intervals were used to generate the data.

The experiments have not considered any spatial correlations of control variables, though the results are sufficient for our conclusions. The correlations we have generated actually correspond to the diagonal elements of the off-diagonal block matrices in the matrix B as shown in equation (6.3) (i.e. the diagonal elements of the block matrices C). We have also not considered the correlation between the unbalanced control variables, for example (χ, h_u). We would expect there to be some correlation here since the divergence and the geostrophic departure are linear combinations of the normal modes, see section 2.3.3. Investigation of the degree of this correlation is left as further work; the key result is demonstrating that the balanced and unbalanced variables are actually uncorrelated.

6.3 Auto-Correlations

In section 6.1.1 the importance of the auto-correlations was demonstrated. We introduced the QCM and validated the time-differencing interval in both a high and low Burger regime. We then presented results that clearly show the benefits of the PV-based transform. To verify our findings in actual assimilation experiments we must first generate auto-correlations for our control variables. The auto-correlations provide, among other things, a spatial length scale for each control variable. This length scale tells us how far background information should be spread. The auto-correlations are then modelled and implemented into the block diagonal components of the cost function.

The auto-correlations provide information regarding the inherent length scale of each variable. This length scale can be compared to the Rossby radius of deformation, discussed in section 2.3.2, and used to assess the degree to which each variable is representing balanced and unbalanced dynamics of the problem.

To generate the auto-correlations we use the QCM and time intervals as in the correlation experiments. We now describe the calculation of the auto-correlations for each control variable.

6.3.1 Generating Auto-Correlations

We assume that we have a set of forecast time-differences u', v', ϕ', ψ', ψ'_b, χ', h'_{res}, and h'_u generated using the QCM. We start by considering a time difference $\mathbf{x}'$ of a field $\mathbf{x}$ with

$$\mathbf{x'}^k = \mathbf{x}^{t_k+\tau} - \mathbf{x}^{t_k} = \left(x_1{}^{t_k+\tau} - x_1^{t_k}, \ldots, x_N^{t_k+\tau} - x_N^{t_k}\right)^T = \left(x'_1{}^k, \ldots, x'_N{}^k\right)^T,$$

where t_k is some time in a model run, τ is the time-differencing interval, and N is the number of points in the domain. The interval τ is chosen in line with the previous results in this chapter to be $\tau = N\Delta x/(\sqrt{gD} + U_c)$. We now consider two elements x'^k_i and x'^k_{i+j} of a time difference $\mathbf{x}'^k$ at time t_k. We treat these as

realizations of random variables X^k and X^k_{+j}. So we have N realisations

$$X^k = \left\{ {x'}^k_1, \ldots, {x'}^k_N \right\},$$

and N realisations

$$X^K_{+j} = \left\{ {x'}^k_{j+1}, \ldots, {x'}^k_N, {x'}^k_1, \ldots, {x'}^k_j \right\},$$

where the subscript indicates the spatial index and the domain is periodic with $x_{N+1} = x_1$. If we now assume that we have M sets of time differences for $\mathbf{x}$ we then have

$$X = \left\{ X^1, \ldots, X^M \right\},$$

and

$$X_{+j} = \left\{ X^1_{+j}, \ldots, X^M_{+j} \right\},$$

where the superscript indicates the temporal index. This gives $N \times M$ realisations of X and X_{+j} and we can now calculate their correlations.

The covariance is given by

$$C_{+j} = COV(X, X_{+j}) =< (X - < X >)(X_{+j} - < X_{+j} >) >=< XX_{+j} > - < X >^2,$$

using the fact that the expectation

$$< X >= \frac{{x'}^1_1 + \ldots + {x'}^1_N + \ldots + {x'}^M_1 + \ldots + {x'}^M_N}{N \times M} =< X_{+j} >,$$

and where

$$< XX_{+j} >= \frac{{x'}^1_1 {x'}^1_{j+1} + \ldots + {x'}^1_N {x'}^1_j + \ldots + {x'}^M_1 {x'}^M_{j+1} + \ldots + {x'}^M_N {x'}^M_j}{N \times M}.$$

Next the correlation coefficients are calculated

$$\rho_{+j} = \frac{C_{+j}}{\sigma_X^2},$$

where the σ_X represents the standard deviations of X, and we note

$$\sigma_X = \sqrt{< X^2 > - < X >^2} = \sqrt{< X^2_{+j} > - < X_{+j} >^2} = \sigma_{X_{+j}}$$

and $\rho_{+0} = 1$.

Finally we let $j = -\frac{N}{2}, \ldots, \frac{N}{2} - 1$ and plot $\rho = \left(\rho_{-N/2}, \ldots, \rho_{N/2-1} \right)^T$ against j. This generates correlation structures which are taken to be homogeneous when modelled in the data assimilation problem described in 3.4.

Experiment Details

The time differencing starts after 200 time steps ($500s$), of the forecast and each difference is taken over period τ. We generate a set of 20 time differences. The experiment is performed in a high and a low Burger regime and is set up as in Table 6.1. The mean flow U_c is either $U_c = 1.25ms^{-1}$ in the high Burger experiment or $U_c = 0.75ms^{-1}$ in the low Burger experiment.
For the high Burger experiments we choose a regime with $R_o = 0.25$. In this regime we demonstrated that there are strong correlations between the model variable time-differences $< \psi', h' >$ and also full fields $< \psi, h + \widetilde{H} >$ indicating a high degree of balance in the flow, though not totally balanced. Also the vorticity and PV-based variables have very similar correlations between balanced and unbalanced variables, see figure 6.7.
In the low Burger experiments we choose a regime with $R_o = 0.15$ and $F_r = 0.75$. This ensures that the gravity wave speed is always greater than the advective velocity (since $F_r < 1.0$). In this regime both the correlation between the time-differences $< \psi', h' >$ and the correlation between the full fields $< \psi, h + \widetilde{H} >$ are large indicating a high degree of balance in the flow, see figure 6.9.

6.3.2 Results

We now present the auto-correlations in high and low Burger regimes for model variables and each set of control variables.

Results: High Burger Number

The auto-correlations are generated as discussed in section 6.3.1 and plots of ρ against j for the high Burger regime are shown in figures 6.12, 6.13 and 6.14 for model variables, vorticity-based variables and PV-based variables respectively.
In this regime we expect the correlation structures to be very similar for the vorticity and PV-based variables since the PV is approximated well by rotation. This is what is observed in figures 6.13 and 6.14, here the correlation structures

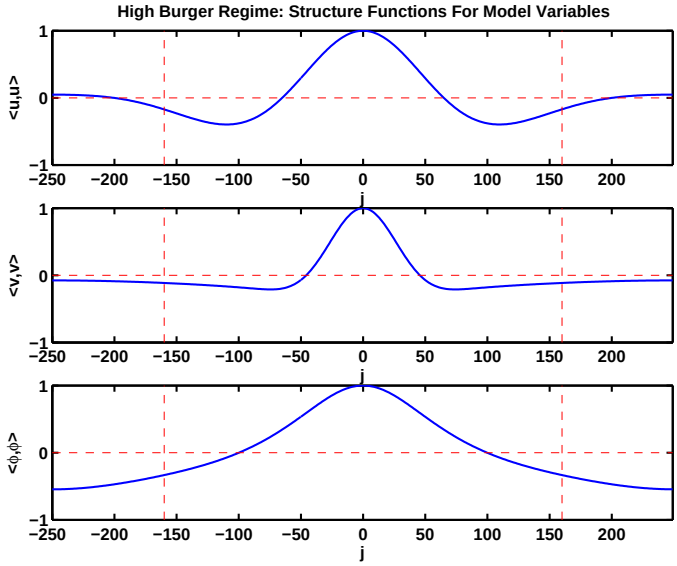

Figure 6.12: Model Variables: Auto-correlations for $B_u = 4.0$, $\tau = 300s$, $U_c = 1.25ms^{-1}$. Plot of the correlation coefficient ρ_{+j} against j, where ρ_{+j} is the correlation coefficient of grid points a distance $j\Delta x$ apart. The Rossby radius and the zero line are marked with a red dashed line.

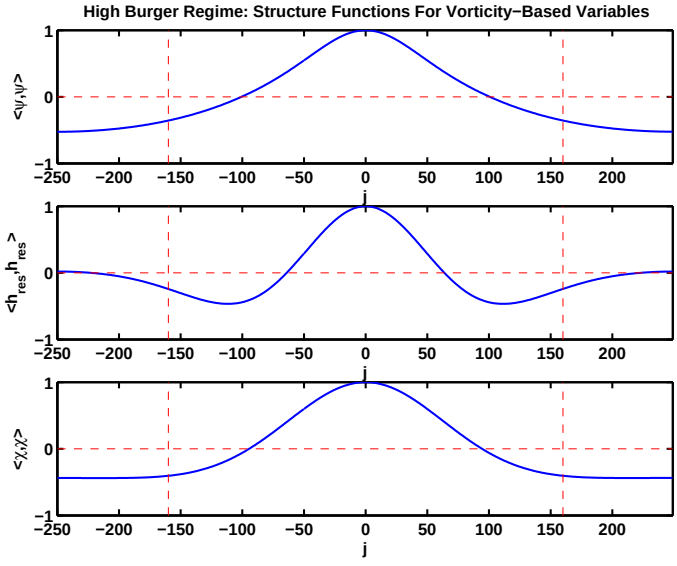

Figure 6.13: Vorticity-Based Variables: Auto-correlations for $B_u = 4.0$, $T = 300s$, $U_c = 1.25ms^{-1}$. Plot of the correlation coefficient ρ_{+j} against j, where ρ_{+j} is the correlation coefficient of grid points a distance $j\Delta x$ apart. The Rossby radius and the zero line are marked with a red dashed line.

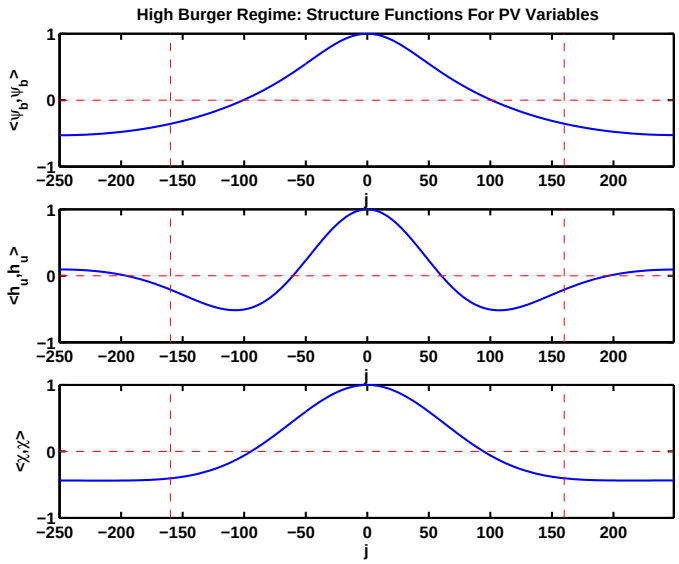

Figure 6.14: PV-Based Variables: Auto-correlations for $B_u = 4.0$, $T = 300s$, $U_c = 1.25ms^{-1}$. Plot of the correlation coefficient ρ_{+j} against j, where ρ_{+j} is the correlation coefficient of grid points a distance $j\Delta x$ apart. The Rossby radius and the zero line are marked with a red dashed line.

are almost identical for both sets of variables.

Balanced variable correlations should be on a scale $L \approx L_r$ where L_r is the Rossby deformation radius defined by equation (2.20). For length scales of approximately L_r geostrophic effects become important. The Rossby radius is also the distance a gravity wave travels in one inertial period. In this experiment

$$L_r = \frac{\sqrt{gH}}{f} \approx 2000m,$$

or 160 grid points. Unbalanced variables should have a correlation length scale $L \leq L_r$. This behaviour is explained in more detail in section 2.3.2.

We see that the correlation scale is largest for the balanced variables ψ' and ψ'_b and of order approximately L_r. The correlation scale for the unbalanced variables h'_{res} and h'_u are less than the Rossby deformation radius. These results indicate that both the vorticity-based and PV variables are capturing the balance and unbalanced flows well in this regime.

The linear balance equation (2.89) is a relationship that holds approximately

between ϕ' and ψ'. A strong correlation between the height and velocity variables would demonstrate the accuracy of this approximation. We can assess such a correlation and hence the degree of balance by comparing the auto-correlations of ϕ' with those of ψ'. We observe that the auto-correlations are very similar indicating that there is a significant amount of balance in the flow. This was demonstrated in the correlation experiment in section 6.2.2, figure 6.7. The QCM is again producing correlation structures that are in agreement with our understanding of the balance in this regime.

Results: Low Burger Number

The auto-correlations are generated as discussed in section 6.3.1 and plots of ρ against j for the low Burger regime are shown in figures 6.15, 6.16 and 6.17 for model variables, vorticity-based variables and PV-based variables respectively. In the low Burger regime the correlation structures for h'_{res} and h'_u are very different. Here the Rossby deformation radius

$$L_r \approx 100m,$$

or 8 grid points, so we would expect that correlations for the unbalanced height variable to be on a scale $L < L_r$ if it accurately represents the unbalanced motion. This is the case for the PV-based variable h'_u but not the vorticity-based variable h'_{res}. In this regime the PV is approximated by height, or mass, and the vorticity-based variables do not account for this and hence we see correlation structures for h'_{res} on too large a scale. We can therefore conclude that the PV-based variables are accurately capturing this behaviour whilst the vorticity-based variables are not.

We also might expect to see a correlation length of ψ'_b greater than ψ' since large-scale balanced correlations are captured in the balanced variable. However, this is not observed and correlation lengths of ψ'_b and ψ' are very similar. This is because in this regime large-scales are scales greater than

$$L_r \approx 100m < L_{orography} = 500m,$$

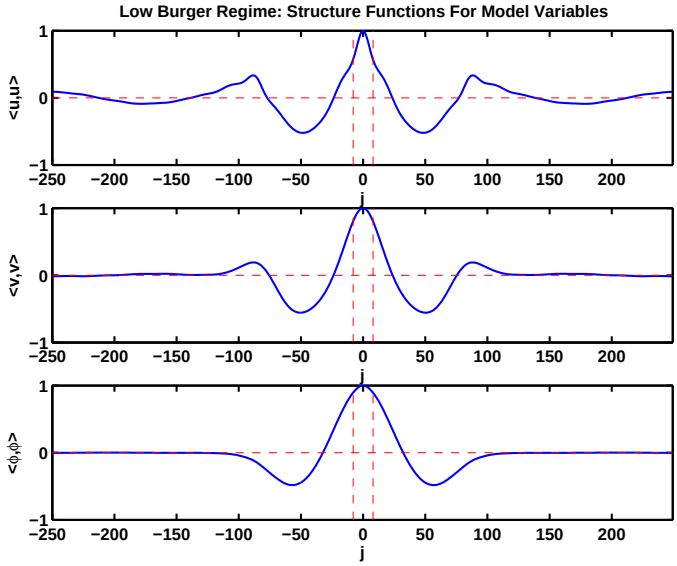

Figure 6.15: Model Variables: Auto-correlations for $B_u = 0.2$, $T = 300s$, $U_c = 0.75ms^{-1}$. Plot of the correlation coefficient ρ_{+j} against j, where ρ_{+j} is the correlation coefficient of grid points a distance $j\Delta x$ apart. The Rossby radius and the zero line are marked with a red dashed line.

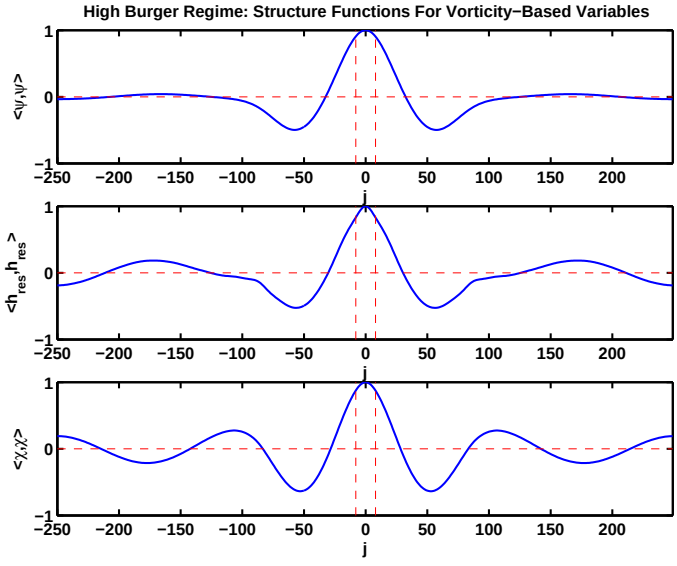

Figure 6.16: Vorticity-Based Variables: Auto-correlations for $B_u = 0.2$, $T = 300s$, $U_c = 0.75ms^{-1}$. Plot of the correlation coefficient ρ_{+j} against j, where ρ_{+j} is the correlation coefficient of grid points a distance $j\Delta x$ apart. The Rossby radius and the zero line are marked with a red dashed line.

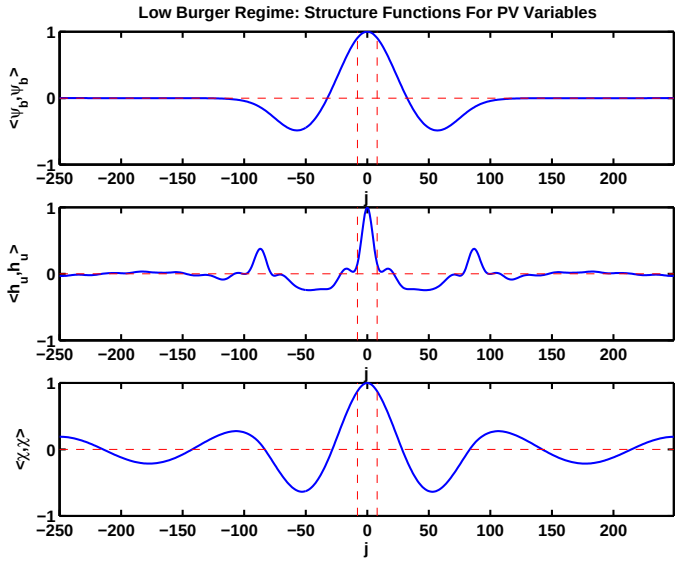

Figure 6.17: PV-Based Variables: Auto-correlations for $B_u = 0.2$, $T = 300s$, $U_c = 0.75ms^{-1}$. Plot of the correlation coefficient ρ_{+j} against j, where ρ_{+j} is the correlation coefficient of grid points a distance $j\Delta x$ apart. The Rossby radius and the zero line are marked with a red dashed line.

where the length scale $L_{orography}$ is the radius of the orography. It is the length scale $L_{orography}$ that dominates the correlation length scales of the control variable in figures 6.16 and 6.17. The radius of the orography is 40 grid points and is equal to the correlation length scales in ψ' and ψ'_b. Therefore any difference in the correlation lengths of ψ'_b and ψ' is swamped by this greater scale.
There is a strong balance signal in the low Burger regime since the auto-correlation structure of ψ' is very similar to that of ϕ'. Also the correlation scale of h'_u is of order $L < L_r$, which is the correct length scale for the unbalanced variable in this regime. We can therefore conclude that the PV-based variables are representing the balanced and unbalanced dynamics well. This cannot be said for the vorticity-based variables.

6.3.3 Summary

The auto-correlations of each control variable are needed before assimilation experiments can be run. These structures provide a spatial length scale for each control variable which is used to spread observational information in the assimilation. The length scales found from the auto-correlations are in agreement with the theory.

We have demonstrated the sensitivity of the auto-correlations to the dynamical regime, characterised by the Burger number. It is therefore necessary to use the appropriate auto-correlation statistics in each Burger regime. Usually in operational centres the auto-correlations are static and do not change with regime. These results demonstrate a high degree of sensitivity to regime for all variables and therefore this should be taken into account in the data assimilation. We also see that the balanced and unbalanced the length scale is intimately linked to the Rossby radius.

The results have also provided further evidence that the PV-based variables are able to capture the balanced and unbalanced motions in both high and low Burger regimes. The vorticity-based variables fail to do so in the low Burger regime. This can be seen by comparing auto-correlations of h'_u and h'_{res} in the low Burger case. Here the length scale of the unbalanced variable should be on a scale $L < L_r$, the Rossby radius of deformation. This is not the case for h'_{res}.

We note that the auto-correlation structures for each variable, in each regime, show some negative correlations as well as secondary peaks. In [12] negative correlations are observed in non-divergent and irrotational flows. In the data assimilation the initial objective of the auto-correlation is to primarily provide a length to spread observational information. Our model is a relatively simple test problem and we have a very simple central orography. This orography seems to have a rather large affect on the correlation scales and could also contribute the negative tails and secondary peaks in the correlations. In the initial assimilation experiments in the following chapter we choose initially to model only the positive

parts of these structures.

The correlation length scales can now be implemented into the background term of the cost function (3.9). These will be modelled using the Laplace-based correlation term as defined in section 3.4.2.

6.4 Summary and Conclusions

The experiments presented in this chapter evaluate the accuracy of the fundamental assumption that the control variables are uncorrelated. Our new results show that the assumption is accurate for both the vorticity and PV-based variables in a high Burger regime. However, when we move to a low Burger regime the assumption only holds for the PV-based variables. Thus the vorticity-based variables are not a valid choice of control variable in this regime. The correlation results demonstrate what is predicted by the theory of previous chapters, and the initial results in [11], [67]. In addition we examined the affect that approximating the PV-based transform has on the correlations between the balanced and unbalanced variables. We are able to conclude that the approximation makes very little difference to the correlation between these variables.

We started this chapter by presenting the background error covariance implied by each control variable transform. This demonstrated that the PV-based transform implies background error statistics that depend on $\bar{q}$ and are therefore state dependent. The background error statistics implied by the vorticity-based transform are static.

We then tested the hypothesis that the PV-based transform is valid across regimes by calculating the correlations between balanced and unbalanced control variables. We are able to validate the theoretical results we presented previously. We demonstrated that for the high Burger regime the vorticity and the PV-based transforms are very similar. Both transforms produce uncorrelated variables and so both are a good choice of control variable. In the low Burger case we found that the vorticity-based variables are correlated and are therefore not a valid choice of

control variable in this regime. The PV-based variables on the other hand remain uncorrelated.

We also demonstrated that the approximation

$$\bar{q} = f/ < \bar{h} >$$

in the PV-based transform makes little difference to the correlation of the variables. This is perhaps not too surprising since the PV-based transform is derived based on a linearisation about a resting state, where the linearisation PV is given by $\bar{q} = f/ < \bar{h} >$.

The implied background error statistics also illustrated the importance of the auto-correlation matrices. The auto-correlations for each variable were found for specific high and low Burger regimes selected from the results of the correlation experiments. The auto-correlations are shown to be highly dependent on the regime and this should be accounted for in the data assimilation. Additionally the auto-correlations provide further evidence that the PV-based transform is valid across regimes.

To obtain good results it was essential to remove the dominant signal in the data. In this chapter we carefully demonstrated that the source of this signal is a gravity wave propagating around the periodic domain. Given this, we note that an alternative method to remove this signal would be to filter the data before taking time-differences. For example, a low-pass filter could then be applied to remove the high frequency gravity wave signal. A huge body of work exists in this field. The low-pass filter and others are described in detail in [45]. Though this is not done in these experiments, a filter could be applied in the following way:

1. Generate the forecast fields.
2. Filter the forecast fields to remove the high frequency gravity wave signal.
3. Take time differences of the filtered model fields.
4. Transform the filtered, time-differenced fields to control variables.

5. Compute the correlations.

In the next chapter we run some initial simple assimilation experiments using each control variable transform. We intend to perform an initial evaluation of how the assimilation behaves using the control variable transforms.

Chapter 7

Assimilation Experiments

In the previous chapter we presented the main results of our research. We performed experiments to test the correlations between the control variables in different regimes. We found that the vorticity and the PV-based variables are both valid choices of control variables in a high Burger regime. However, it is found that there is significant correlation between vorticity-based variables in a low Burger regime. The PV-based variables on the other hand remain uncorrelated.
In this chapter we consider some initial simple assimilation experiments using the vorticity and PV-based variables. The aims are as follows:

1. To analyse theoretically the expected influence of each control variable transform on the analysis. This is done using the best linear unbiased estimator or BLUE.

2. To highlight potential numerical problems that may arise when using the control variable transforms.

3. To demonstrate that the transforms give similar increments in high Burger regime and different increments in a low Burger regime.

We now have a complete data assimilation system, before we run any assimilation experiments we first validate that it is working correctly. We do this using some standard tests.

In our simple experiments we consider a single observation in one variable at the initial time $t = 0$. In this simple setting the analysis of results is relatively simple and we find that much can be learnt about the effect of each of the control variable transforms on the analysis produced by the assimilation.

To aid in the analysis of the results the best linear unbiased estimate (BLUE) is adapted to include the control variable transformation. This analysis helps us to interpret the results of the assimilation experiments. The BLUE solution of the assimilation problem [33] is obtained for a special case of incremental 4D VAR that is essentially a 3D VAR setting since the observation is at the initial time $t = 0$. However, in the following experiments we run the assimilation with multiple outer iterations. We may then use the BLUE as a guide to understanding the analysis increments we might expect to obtain from the assimilation. The experimental set up is chosen this way so that the analysis of the results is clear. This procedure is sufficient for our initial comparison of the influence of the vorticity and PV-based control variables on the data assimilation. From this analysis we are able to draw conclusions regarding the behaviour of each of the transforms in the assimilation. From the BLUE analysis we can also see when we will expect to encounter numerical difficulties in the minimisation.

Before we are able to run any experiments we must choose how to model the auto-correlations for each control variable in the data assimilation. In the previous chapter we demonstrated that these are a vital aspect of the assimilation. In the BLUE analysis later in this chapter we show that care must be taken to implement the auto-correlations successfully with the control variables.

It might seem natural to use the experimental auto-correlations we generated in the previous chapter. However, we see that given the properties of the matrices they cannot be used. So for these experiments we choose to simply to model the auto-correlations using the Laplace-based smoother introduced in section 3.4.2, which models the inverse of Gaussian functions. The results obtained give us additional insight into the behaviour of the transforms. They also demonstrate the important differences in the analysis increments generated by the data assimilation

when the vorticity and the PV-based transforms are used.
Before the results are discussed we describe briefly how the data assimilation system was tested in its entirety. This final validation of the system is made using the gradient test, which is described in [44].

7.1 Testing The Data Assimilation System

The gradient test verifies the calculation of the cost function gradient as, even after thorough testing of the tangent linear model and the adjoint, as discussed in section 4.4.1 and 4.4.2, errors may still be present in the full data assimilation system.
The gradient test is described in [44]. This test verifies the correctness of the gradient and function calculation used in the data assimilation by a Taylor expansion of the cost function (3.2)

$$J[\mathbf{x}_0 + \alpha\mathbf{b}] = J[\mathbf{x}_0] + \alpha\mathbf{b}^T\nabla J[\mathbf{x}_0] + O(\alpha^2),$$

where α is a small scalar and $\mathbf{b}$ is a vector of unit length (for example $\mathbf{b} = \nabla J \left\|\nabla J\right\|^{-1}$). Rearranging the above, a function of α is found,

$$E(\alpha) = \frac{J[\mathbf{x}_0 + \alpha\mathbf{b}] - J[\mathbf{x}_0]}{\alpha\mathbf{b}^T\nabla J[\mathbf{x}_0]} = 1 + O(\alpha).$$

For small alpha we should have $E(\alpha)$ close to 1 [44].
The gradient test is run using each control variable transform and is successful in both cases. Examples of the gradient test results for each of the control variables can be found in appendix B.
Having tested the data assimilation in its entirety we now present the single observation experiments.

7.2 Single Observation Experiments: BLUE Analysis

For initial evaluation of the influence of the vorticity and PV-based transforms' on the data assimilation we consider single observation experiments. Here the single observation is taken at time $t = 0$.

In order to gain insight into how the control variable transforms influence the assimilation we start by finding the best linear unbiased estimate, or BLUE [33], for the special case of the transformed incremental 4D VAR cost function when we have only one observation at the initial time $t = 0$. We do this to analyse the affect of each control variable transform on the analysis produced by the assimilation in these single observation experiments. In addition, it aids in our interpretation of the following results. From the BLUE we are also able infer potential numerical problems that might be encountered. We bear in mind that we are only considering the parameter transform (i.e. a control variable transform that is assumed to remove only the cross-correlations of variables) and auto-correlations are specified in the block diagonal components of the transformed background error covariance matrix. For this special case, we note that the cost function is linear due to the fact the two possible sources of non-linearity, the observation operator and the non-linear model, are not present. This is because the observation is at the initial time and hence the non-linear model is not required, and the observations are of model quantities at grid points, and therefore the observation operator $\mathbf{H}$ merely picks out the location of the observation. The problem is essentially a 3D VAR minimisation. The only exception is when using the full PV-based variables over multiple outer loops. In this case, there can be a non-linear effect due to the influence of the linearisation state PV. This occurs because the linearisation state is updated after each outer loop. Thus, we can consider the BLUE to be the solution of the minimisation when model variables, vorticity-based variables and the approximate PV variables are used as control variables. However, when the full PV-based variables are used with multiple outer

loops, the BLUE is the solution of a given outer iteration of the assimilation.
To find the BLUE we consider the gradient of the transformed incremental cost function defined in equation (3.9) (where the observation is at time $t = 0$) with respect to the control variables. The gradient is given by equation (3.13), but we may simplify since our observation is at the initial time $t = 0$, to obtain

$$\nabla \widetilde{J}[\mathbf{z}'_0] = \mathbf{\Lambda}_z^{-1}(\mathbf{z}'_0 - \mathbf{z}'^b) - \mathbf{U}^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{U}\mathbf{z}'_0). \tag{7.1}$$

Here $\mathbf{\Lambda}_z$ is the background error covariance matrix for the control variables $\mathbf{z}'$. The matrix is block diagonal as the control variables are assumed to be uncorrelated with each other; thus only the auto-correlation of each control variable needs to be specified. $\mathbf{U}^T$ is the adjoint of the $\mathbf{U}$ transform. We discuss the modelling of the auto-correlations in the next section.
At the minimum we have

$$\nabla \widetilde{J}[\mathbf{z}'^a] = \mathbf{\Lambda}_z^{-1}(\mathbf{z}'^a - \mathbf{z}'^b) - \mathbf{U}^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{U}\mathbf{z}'^a_0) = 0, \tag{7.2}$$

where $\mathbf{z}'^a$ is the analysis at the minimum. Rearranging this equation and adding and subtracting $\mathbf{U}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{U}\mathbf{z}'^b$ gives

$$\mathbf{z}'^a - \mathbf{z}'^b = \left(\mathbf{\Lambda}_z^{-1} - \mathbf{U}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{U}\mathbf{z}'^b). \tag{7.3}$$

We note that here we consider $\mathbf{H}$ to be linear. After further manipulation we can write the analysis increment $\mathbf{z}'^a - \mathbf{z}'^b$ as

$$\mathbf{z}'^a - \mathbf{z}'^b = \mathbf{\Lambda}_z\mathbf{U}^T\mathbf{H}^T\left(\mathbf{H}\mathbf{U}\mathbf{\Lambda}_z\mathbf{U}^T\mathbf{H}^T + \mathbf{R}\right)^{-1}(\mathbf{d} - \mathbf{H}\mathbf{U}\mathbf{z}'^b), \tag{7.4}$$

which is the BLUE using the control variable transform for this problem.
We can simplify further since we consider an experiment where we have only one observation at point j. The matrix $\mathbf{R}$ is now simply a scalar and, since the observation is of a model quantity, the operator $\mathbf{H} = \mathbf{e}_j^T$ is the jth unit vector (i.e. $\mathbf{H}$ is $1 \times n$ matrix, where n is the dimension of $\mathbf{z}$, with zeros in every entry except at the jth point where we have a 1). We also note that

$$\left(\mathbf{H}\mathbf{U}\mathbf{\Lambda}_z\mathbf{U}^T\mathbf{H}^T + \mathbf{R}\right)^{-1}(\mathbf{d} - \mathbf{H}\mathbf{U}\mathbf{z}'^b)$$

is also a scalar and therefore

$$\mathbf{z}'^a - \mathbf{z}'^b \propto \mathbf{\Lambda}_z \mathbf{U}^T \mathbf{e}_j. \tag{7.5}$$

The $\mathbf{e}_j$ on the right hand side will pick out the jth column of $\mathbf{\Lambda}_z \mathbf{U}^T$. Therefore we derive a proportional relationship for the analysis increment; the analysis increment in control space, $\mathbf{z}'^a - \mathbf{z}'^b$, is proportional to a column of the transpose of the U-transform acted on by the auto-correlation matrix $\mathbf{\Lambda}_z$.

The BLUE clearly demonstrates the importance of the control variable transform and the prescribed background auto-correlations to the analysis increment. In fact the influence from the observation comes in the form of a multiplicative constant.

From equation (7.5) we can also observe some potential problems. If there is a null space in the matrix $\mathbf{\Lambda}_z$ the problem is ill-posed and does not have a unique solution. The null space of a matrix A is the set of all vectors x for which $Ax = 0$. Any multiple of a vector in the null space can be added without changing the solution of equation (7.4). If there is a nearly null space, i.e. if $\mathbf{\Lambda}_z$ has some very small eigenvalues, then we can expect the problem to be ill-conditioned and there will be numerical difficulties when solving the minimisation; the minimisation will require a large number of iterations to fully converge to the BLUE solution.

In our single observation test if we use only model variables that are assumed to be uncorrelated we obtain

$$\mathbf{x}'^a - \mathbf{x}'^b \propto \mathbf{\Lambda}_x I \mathbf{e}_j = \mathbf{\Lambda}_x \mathbf{e}_j,$$

since we replace $\mathbf{U}$ with the identity matrix, i.e. the analysis increment is proportional to a column of $\mathbf{\Lambda}_x$, as mentioned in section 3.3. However, when we consider the control variable transforms the analysis increments in control space for the vorticity and PV-based variables will differ due to the influence of $\mathbf{U}_V^T$ and $\mathbf{U}_{PV}^T$, where the subscript indicates the vorticity or PV-based transform respectively.

Applying this analysis to our transforms, for an arbitrary outer loop, we see that

when vorticity-based control variables are used we have

$$\begin{pmatrix} \boldsymbol{\chi}'^a - \boldsymbol{\chi}^b \\ \boldsymbol{\psi}'^a - \boldsymbol{\psi}^b \\ \mathbf{h}'^{\,a}_{res} - \mathbf{h}^b_{res} \end{pmatrix} \propto \begin{pmatrix} \sigma^2_\chi \hat{\mathbf{\Lambda}}_\chi & 0 & 0 \\ 0 & \sigma^2_\psi \hat{\mathbf{\Lambda}}_\psi & 0 \\ 0 & 0 & \sigma^2_{h_{res}} \hat{\mathbf{\Lambda}}_{h_{res}} \end{pmatrix} \begin{pmatrix} \left(\frac{\partial}{\partial x}\right)^T & 0 & 0 \\ 0 & \left(\frac{\partial}{\partial x}\right)^T & \frac{f}{g} \\ 0 & 0 & 1 \end{pmatrix} \mathbf{e}_j,$$

writing the matrices in block form as in section 6.1.1. For PV-based variables we have

$$\begin{pmatrix} \boldsymbol{\chi}'^a - \boldsymbol{\chi}^b \\ \boldsymbol{\psi}'^a_{\ b} - \boldsymbol{\psi}^b_b \\ \mathbf{h}'^a_u - \mathbf{h}^b_u \end{pmatrix} \propto \begin{pmatrix} \sigma^2_\chi \hat{\mathbf{\Lambda}}_\chi & 0 & 0 \\ 0 & \sigma^2_{\psi_b} \hat{\mathbf{\Lambda}}_{\psi_b} & 0 \\ 0 & 0 & \sigma^2_{h_u} \hat{\mathbf{\Lambda}}_{h_u} \end{pmatrix} \begin{pmatrix} \left(\frac{\partial}{\partial x}\right)^T & 0 & 0 \\ 0 & \left(\frac{\partial}{\partial x}\right)^T & \frac{f}{g} \\ 0 & \frac{\partial}{\partial x}\left(\nabla^{-2}\bar{q}\,\cdot\right)^T & 1 \end{pmatrix} \mathbf{e}_j.$$

Here the outer loop index k has been dropped for convenience. The block diagonal components $\mathbf{\Lambda}_z$ represent the auto-covariance of each control variable, as indicated by the subscript. The auto-covariance matrices are written as the product of the appropriate auto-correlation matrix and variance.

So if we have a high Burger regime experiment with a single u or h observation from the BLUE we would expect the analysis increments *in control space* for the vorticity and PV-based variables to be very similar to each other. This is because these columns in the $\mathbf{U}_V^T$ and $\mathbf{U}_{PV}^T$ are equal and the auto-covariances are very similar. However, these control variable analysis increments will then be transformed to model variable increments using the inverse transforms. So the actual model variable increments that result might be different.

We can also see that the increments when we have a single v observation will be different; in the vorticity case only ψ' will be incremented, whereas for the PV-based variables we expect to see increments to both balanced and unbalanced variables ψ'_b and h'_u respectively. This can be seen above since a single v observation will have $\mathbf{H}^T = \mathbf{e}_j$, with j a point such that we pick out a column in the central block column of $\mathbf{U}_V^T$ and $\mathbf{U}_{PV}$. In $\mathbf{U}_V^T$ only the central block, which corresponds to ψ', is non-zero. The implications of this can be fully understood by examining the vorticity-based transform. In section 5.1.2, equation (5.11) is the only means available to increment v. Further, since the h'_{res} increment for a single

v observation will be zero, equation (5.12) becomes $h' = \frac{f}{g}\psi'$. So the increment to $\phi' = gh'$ satisfies the linear balance equation (2.89). Therefore, we can conclude that for a single v observation the vorticity-based variables will only produce a *balanced* increment to the model variables. However, this behaviour is only present in the 1D problem we are studying.

Despite these differences in control variable increments, the theory we presented in this work suggests that in a high Burger regime we would expect that the model variable increments produced by both will be similar, since the vorticity-based variables should approximate the PV-based variables. The aim of the following experiments is to test whether this is in fact the case.

We also note the state dependence in $\mathbf{U}_{PV}^T$ by virtue of $\bar{q}$. So in our iterative assimilation $\bar{q}$ will constantly be updated on each outer loop. However, $\mathbf{U}_V^T$ is static and will be the same on each outer iteration of the assimilation.

Finally, we see that there will be problems solving the minimisation if the variances of the control variables are very different sizes. Since the cross-correlation of the control variables is assumed to be removed the background term is essentially split into three separate terms for each control variable. For the minimisation to be appropriately weighted the background error variances should be of similar magnitude. As well as this the observation term must be weighted to balance the background terms. We also see that $\mathbf{U}_V^T$ is block-upper triangular. Therefore, if the variance of h'_{res} is very small relative to the other variances, the matrix $\mathbf{U}_V^T$ will have a large nearly null space.

We now describe how we choose to model the auto-correlations in the following experiments.

7.3 Modelling The Auto-Correlations

In section 3.4 we discussed how auto-correlations are generated and used in operational centres. A proxy quantity for background error statistics is used to generate statistical data. Then a method is applied to model the auto-correlations

within the data assimilation. Due to the size of the problem this is not done exactly and, since the auto-correlations are global averages, this may not be appropriate in any case. So, before we are able to run our simple assimilation experiments, we must choose how we will model the auto-correlations. We now examine two possibilities.

7.3.1 Experimental Auto-Correlations

In this section we analyse whether it is possible to use the auto-correlations generated from the experiments in section 6.3 directly in the data assimilation. To do this we create isotropic and homogeneous correlation matrices using the experimental auto-correlations and investigate their properties.

A valid correlation matrix must be symmetric semi-definite with a unit diagonal [26]. The matrix A is said to be positive definite if, for all vectors $x \in \mathbf{R}^n$, $x^T Ax > 0$, and A is semi-definite if $x^T Ax \geq 0$. A matrix is positive definite if and only if all the eigenvalues of the matrix are positive. Any positive semi-definite matrix A has a Cholesky decomposition $A = X^T X$.

The experimental auto-correlation matrices are symmetric and have a unit diagonal by construction. However, it is found that the matrices are not symmetric positive definite. Some relatively small negative eigenvalues exist. As well as this there are many very small positive eigenvalues suggesting that there is a large nearly null space. For example, the auto-correlation matrix for ψ has 500 eigenvalues of which 471 are in the range -5×10^{-5} and $+5 \times 10^{-5}$. As shown in section 7.2, a nearly null space will cause the minimisation to be ill-conditioned and so these matrices cannot be used.

If the auto-correlation matrix is not symmetric positive definite then from a statistical perspective this is not correct. Also, from a least-squares perspective, this means that the normal equations of the problem are not defined since the auto-correlation matrix $\mathbf{\Lambda}_z$ cannot be written as $A^T A$ for some matrix A.

The existence of these small negative (and positive) eigenvalues can be attributed

to numerical noise. Examples of this problem are found in many disciplines. In finance this is a common problem [26]. There are methods for finding close approximate matrices that are valid correlation matrices, for example [26]. However, we leave this as further work and choose to model the auto-correlations using Gaussian functions since matrices formed in this way will be valid correlation matrices.

7.3.2 Gaussian Auto-Correlations

For each control variable we use the Laplace-based smoother introduced in section 3.4.2 to model the inverse of Gaussian functions. The Gaussian functions have the correlation length scales of the auto-correlation structures found in section 6.3.1. Plots of the auto-correlations and a column of the modelled auto-correlation matrices $\mathbf{\Lambda}$ are found in appendix C in figures C.1 to C.3 for the high Burger regime and figures C.4 to C.6 for the low Burger regime. Here the number of grid points separation, j, is plotted against the correlation coefficient.

Figure 7.1 is an example of how the auto-correlations are modelled. We see that we do not attempt to model exactly the auto-correlations found in practice, as derived in Chapter 6. The Laplace-based smoother captures well the central, positive correlations. This is by no means the only way to represent the auto-correlations, some methods used in operational centres are described in [30] and [25]. The method we select is used in [32] and is chosen for ease of implementation. In these experiments we use the auto-correlations as a guide to providing an appropriate length scale with which to spread background information.

We now present some simple experimental results using each of the control variable transforms. These results demonstrate the conclusions we made in section 7.2 regarding the influence of the transforms on the data assimilation.

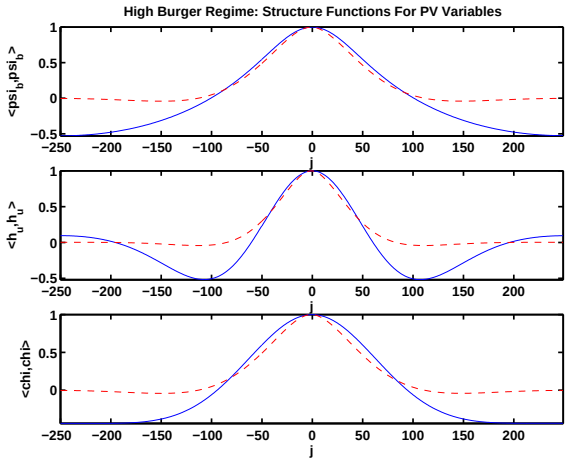

Figure 7.1: High Burger regime, PV-based variables:
Auto-correlations for ψ_b (top), h_u (middle) and χ (bottom) plotted with Laplace-smoother approximation (red dashed line).

7.4 Assimilation Experiments

We now present the single observation assimilation experiment results. These results demonstrate the conclusions made from the analysis in section 7.2 of the BLUE. We will show that the vorticity and PV-based control variable transforms produce similar analysis increments in high Burger regime, but different increments in a low Burger regime. Before we present the results we first give specific details of the experiments.

7.4.1 Experiment Details

The following experiments are set up as in section 6.2.1 and 6.3.1. The truth and the background are fields from the same forecast run, with a separation in time equal to the time-differencing interval used in the experiments in section 6.2.1. The observation point is taken to be at $i = 423$, or $423\Delta x = 5287.5m$. At this point the error in the background guess is due to the PV being incorrectly located. Since the background field is taken to be a length of time later in the same forecast the PV has been advected further than in the true field.

The correlation scales used in the Laplace-based smoother and the background error variances for each control variable are given in Table 7.1. The background error variances for each control variable are calculated from the experiments in section 6.2.1. We see that they are very different in magnitude and thus it is difficult to obtain good results. However, for the observation error variances given in Tables 7.2 and 7.3 we were able to converge the minimisation, though many iterations were required. In each experiment the minimisation was run for 20 outer loops, each with 5000 inner iterations. Such a large number of iterations was required since the difference in size of the variances for the height variables and the ψ variables is vast and the background terms in the cost function have very different weights. To verify this we tested the data assimilation with equal variances. In these tests the convergence was much faster and each control variable was incremented evenly. A result of using different observation error variances in these experiments is that we cannot expect the scales of the increments from the vorticity and PV-based variables to be exactly the same. However, we are able to compare the shape of the increments.

	High Burger Regime		Low Burger Regime	
Control Variable	Variance	Length Scale	Variance	Length Scale
χ	6.7×10^{-4}	40	1.6×10^{-5}	15
ψ	1.1×10^{-1}	40	1.1×10^{-4}	15
h_{res}	7.6×10^{-9}	30	2.4×10^{-11}	15
ψ_b	1.1×10^{-1}	40	9.5×10^{-5}	15
h_u	6.3×10^{-9}	30	1.6×10^{-12}	5

Table 7.1: Experiment Set-Up.

In the high Burger regime the Rossby radius is

$$L_r = \frac{\sqrt{gH}}{f} \approx 2000m,$$

or 160 grid points and is the same as in section 6.3.2. In the low Burger regime the

Rossby radius is

$$L_r \approx 100m,$$

or 8 grid points as in section 6.3.2. We now present the results of the simple single observation experiments.

	High Burger Regime		
Transform	u	v	ϕ
PV	7×10^{-9}	1×10^{-6}	6×10^{-11}
Vorticity	7×10^{-9}	1×10^{-4}	6×10^{-13}

Table 7.2: Observation Error Variances for high Burger regime.

	Low Burger Regime		
Transform	u	v	ϕ
PV	1×10^{-10}	1×10^{-14}	1×10^{-13}
Vorticity	1×10^{-10}	1×10^{-14}	1×10^{-12}

Table 7.3: Observation Error Variances for low Burger regime.

7.4.2 Assimilation Results: High Burger Regime

The analysis increment from the experiment for a single u observation are shown in figure 7.2. The control variable and model variable increments are the same when using vorticity and PV-based variables since both sets of variables use the same χ. The increment to u is just the derivative of the χ increment. More interesting results are obtained when we have a single v or h observation.
The analysis increments from the experiment with a single h observation are shown in figure 7.3 for the vorticity variables and figure 7.4 for the PV-based variables. We see that all the increments have a very similar shape for both variables. This demonstrates the theory presented in section 2.4 that in the high Burger regime we expect the vorticity and PV-based variables to be similar. We also demonstrate

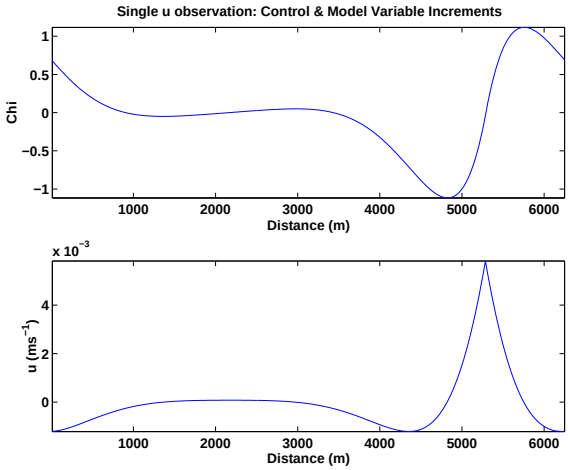

Figure 7.2: High Burger Regime, Vorticity and PV-Based Control Variables: Single u observation. Plot of analysis increment to χ (top) and resulting increment to u (bottom).

one of the conclusions from the BLUE analysis in section 7.2 that the increments to the control variables, when we have a single h observation, are very similar. Our analysis of the BLUE in section 7.2 predicts that we will see a difference in control variable increments when we have a single v observation. In this case we can see from the BLUE that, for the PV variables, we obtain an increment to both the balanced streamfunction and the unbalanced height, this is demonstrated in figure 7.6.

For the vorticity variables we concluded from the BLUE analysis that we only obtain an increment to the full streamfunction ψ, we are able to demonstrate this in figure 7.5 where the increments to χ and h_{res} are zero. Therefore, when the control variable increment is transformed to model variable increments, we are only able to obtain a *geostropically balanced* increment to ϕ. This can be seen since from equation (5.12); since $h_{res} = 0$ the increment to $h = \frac{f}{g}\psi$, which satisfies the linear balance equation (2.89). This behaviour is only present in the 1D problem we are studying. However, for the PV-based variables the increments to v and h are not totally balanced as they are a combination of unbalanced and balanced increments. Despite this the resulting increments to the model fields are very

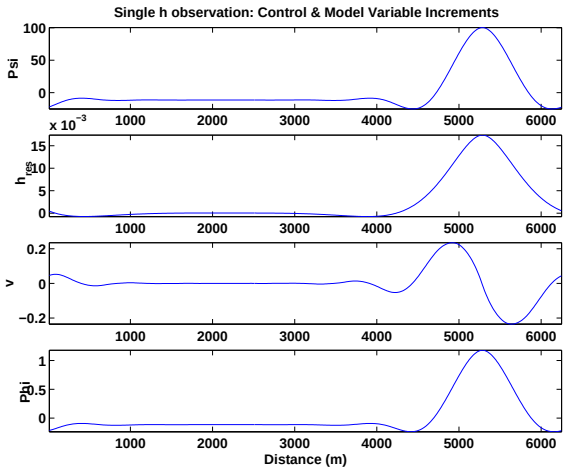

Figure 7.3: High Burger Regime, Vorticity Based Control Variables: Single h observation. Plots, from the top down, are the analysis increment to control variables ψ and h_{res}. Next are the resulting model variable increments to v and ϕ.

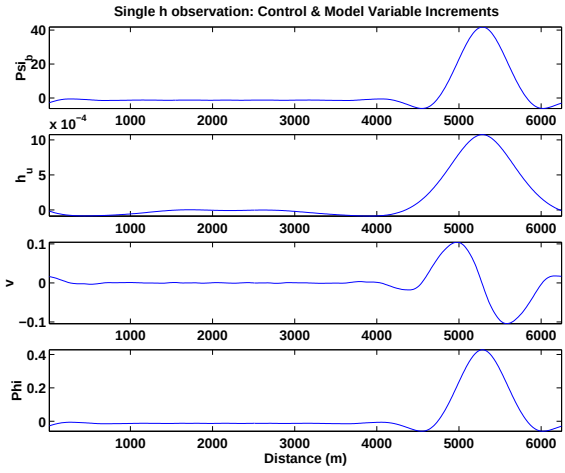

Figure 7.4: High Burger Regime, PV-Based Control Variables: Single h observation. Plots, from the top down, are the analysis increment to control variables ψ_b and h_u. Next are the resulting model variable increments to v and ϕ.

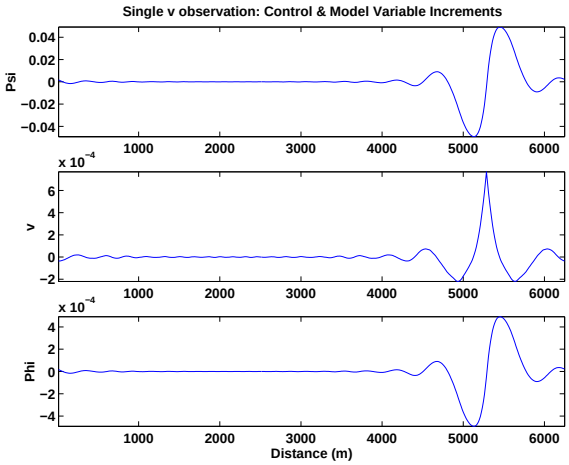

Figure 7.5: High Burger Regime, Vorticity Based Control Variables: Single v observation. Plots, from the top down, are the analysis increment to control variables ψ, followed by the resulting model variable increments to v and ϕ.

similar. Again we have been able to demonstrate experimentally the behaviour that we expected to see in the high Burger regime.

With these simple single observation experiments we have been able to demonstrate that in the high Burger regime the increments from both the vorticity and PV-based variables are very similar. We see that when we have an observation of v the vorticity variables can only provide a balanced increment to ϕ that is in balance with ψ. However, this behaviour is only present in the 1D problem we are studying.

We now move onto the low Burger regime experiments. Here we expect to see some differences in the transforms. These come from the behaviour of each transform as we change regime, which also produces differences in length scales of the auto-correlations (the length scale in the Gaussian correlations were chosen to fit the actual auto-correlations).

7.4.3 Assimilation Results: Low Burger Regime

The PV-based transform has been designed with the ability to represent the balanced flow in all regimes. The vorticity-based variables on the other hand will

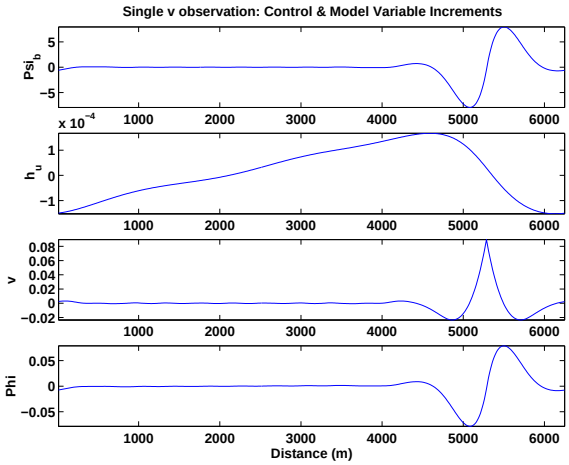

Figure 7.6: High Burger Regime, PV-Based Control Variables: Single v observation. Plots, from the top down, are the analysis increment to control variables ψ_b and h_u. Next are the resulting model variable increments to v and ϕ.

not be able to represent the balanced flow in a low Burger regime. Therefore, in a low Burger regime we would expect the data assimilation to produce different increments when using the vorticity and PV-based control variables. The results in this section aim to demonstrate this experimentally.

We have thus far shown that the PV and vorticity based variables are similar in a high Burger regime. We therefore aim to show that the resulting model variable increments in a low Burger regime are different when using the vorticity based variables, which do not represent the balanced flow well, to those obtained when the PV-based transform is used. Indeed, this is what we find. We do not present the results for χ here since this is the same variable in both cases in our experiments. We are interested in experiments with a single v and h observation. We start with the results for h. These are shown in figure 7.7 for the vorticity based variables and figure 7.8 for the PV-based variables. We see that the length scales of the increments are different. The Rossby deformation radius in this experiment is

$$L_r \approx 100m,$$

compared to approximately $2000m$ in the high Burger regime. The PV-based variables are on a smaller scale than they were in the high Burger experiments.

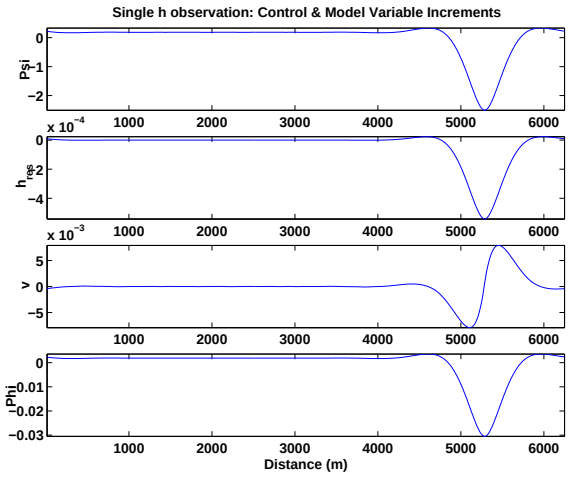

Figure 7.7: Low Burger Regime, Vorticity-Based Control Variables: Single h observation. Plots, from the top down, are the analysis increment to control variables ψ and h_{res}. Next are the resulting model variable increments to v and ϕ.

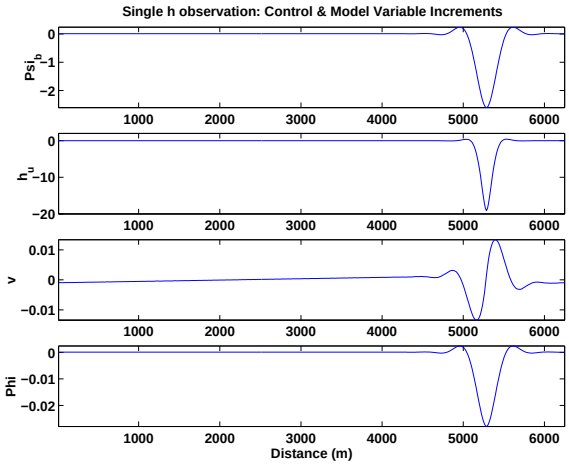

Figure 7.8: Low Burger Regime, PV-Based Control Variables: Single h observation. Plots, from the top down, are the analysis increment to control variables ψ_b and h_u. Next are the resulting model variable increments to v and ϕ.

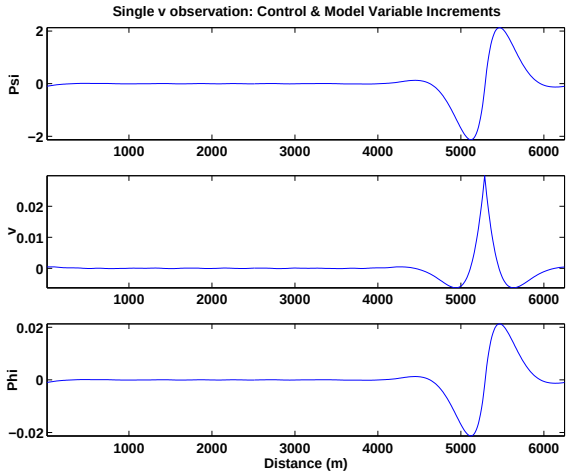

Figure 7.9: Low Burger Regime, Vorticity Based Control Variables: Single v observation. Plots, from the top down, are the analysis increment to control variable ψ, followed by the resulting model variable increments to v and ϕ.

This reflects the change in the intrinsic length scale, the Rossby radius, which is much smaller in a low Burger regime than a high regime. In the vorticity case the increments are on a larger scale than the PV-based case and, in fact, have length scales that are not very different to that of the increments in the high Burger regime. This does not reflect the change in the Rossby radius from the high Burger regime to the low Burger regime.

The increments from the assimilation experiments with a single v observation are shown in figure 7.9 for the vorticity based variables and figure 7.10 for the PV-based variables. Here we see the greatest difference. Again we are able to demonstrate that we only produce a balanced increment through ψ in the vorticity case as we concluded from the BLUE analysis previously. However, the PV-based variables have a large-scale increment to the unbalanced height resulting in an increment to v that is different from that produced by the vorticity based variables. The vorticity-based control variables are not able to increment the unbalanced flow at all in this case.

In the vorticity case, increments to v are produced only via ψ as shown by the BLUE, which is assumed to be totally balanced, that is to say, ψ is assumed to

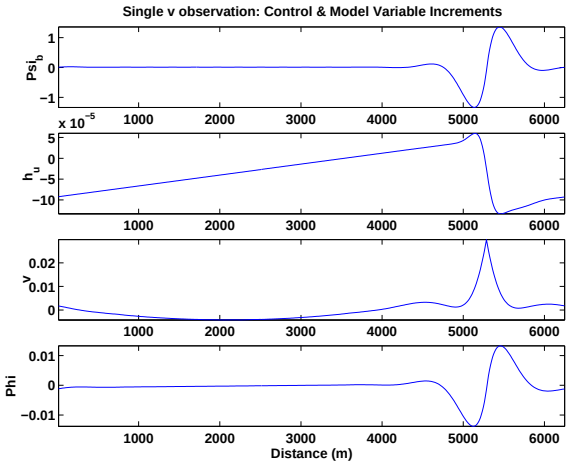

Figure 7.10: Low Burger Regime, PV-Based Control Variables: Single v observation. Plots, from the top down, are the analysis increment to control variables ψ_b and h_u. Next are the resulting model variable increments to v and ϕ.

satisfy the linear balance equation exactly. However, for the PV-based variables the velocity v is split and increments to v can be produced from either its balanced or unbalanced component, or both.

7.4.4 Summary

In these simple assimilation experiments we were able to demonstrate the main differences in the transforms, which we concluded from the analysis of the BLUE and from the theory in section 2.4. We demonstrated that the increments generated by the two control variable transforms are indeed very similar in a high Burger regime, as expected from the theory.

In the low Burger regime we succeeded in showing that the transforms produce different increments. In this regime, for a single h observation, the increments produced by the PV-based variables appear more representative of the intrinsic length scale. The increments are on smaller scales than those of the vorticity-based variables and represent the change in intrinsic length scale from high to low Burger regimes.

Due to the large differences in the sizes of the background error variances the

weighting of the background terms was far from equal. Therefore, it was hard to obtain good results; many iterations were required to obtain solutions that we expect from the BLUE analysis. This is clearly not feasible in a operational assimilation system but was possible in our experiments. Before further assimilation experiments are run with this system a preconditioning step should be implemented. This is left as further work.

7.5 Conclusions

In this chapter we considered some simple assimilation experiments. Before presenting the assimilation results we first analysed theoretically the expected influence of each control variable transform on the analysis using the best linear unbiased estimator or BLUE. From the BLUE we were able to see a major difference between the analysis increments we can obtain when using the vorticity-based control variables and the PV-based control variables. The vorticity-based variables will only produce a balanced analysis increment when we have a single v observation. This is because the only variable that is able to increment v is ψ, which is related to ϕ by the linear balance relationship. However, this behaviour is only present in the 1D problem we are studying. The PV-based variables on the other hand can produce both balanced and unbalanced increments to v.

From the BLUE we are also able to determine when we will encounter numerical problems in the minimisation. We see that if the auto-correlation matrix has a nearly null space then the problem will be ill-conditioned. This is in fact the case when we examine the properties of the experimental auto-correlation matrices and therefore they may not be used in the assimilation. Since the cross-correlation between control variables is assumed to be removed by the transform the background term is essentially three separate terms for each control variable. We see that if the variances of these control variables are not of similar magnitude then the background terms will not be evenly weighted and the minimisation will take

many iterations to fully converge. This is what is observed in these experiments. The experiments we presented demonstrate the behaviour that is expected from the theory; the vorticity and PV-based control variables are similar in high Burger regimes and both represent the balance in the model, but they are different in low Burger regimes were only the PV-based control represents the balanced flow accurately, therefore the analysis increments obtained are different.

Chapter 8

Conclusions and Further Work

Data assimilation is the process of finding the best estimate of the current state of a system. In numerical weather prediction this system is the atmosphere and oceans. It is a critical stage in the production of weather forecasts: More accurate initial conditions produce more accurate forecasts. However, the problem is huge and, to be practical to solve, sophisticated methods must be developed that simplify the problem, whilst respecting the inherent characteristics of large-scale atmospheric motions.

The 4D VAR data assimilation problem became practical to solve when an incremental formulation was introduced in [7]; here a sequence of linearised quadratic cost functions are solved. Even with these simplifications further approximations are necessary to handle the background error covariance matrix. Typically, this is achieved using the control variable transform. The use of a control variable transform in variational data assimilation was first introduced in [47]. Here the model variables are transformed into a new set of variables and the new variables are assumed to be uncorrelated with each other. The background error covariance matrix is then block diagonal. Thus, we may think of the control variable transform as having a dual purpose; it is a necessity required to simplify the problem, and also a means to introduce important dynamical relationships into the data assimilation.

The quality of the analysis produced by the assimilation is highly dependent on

the choice of control variables. In most operational centres the control variables are based on a balanced / unbalanced partitioning of the model variables first used in [47]. Here the assumption is that the balanced flow is uncorrelated with the unbalanced flow. However, all of the control variables used operationally at best only approximate this balanced / unbalanced partitioning. These control variables are essentially vorticity-based, that is to say that vorticity is assumed to be totally balanced. One exception are height/mass-based variables where it is assumed that the height/mass is totally balanced [25]. This method has similar deficiencies since it too approximates the balanced variable, in addition there are complications around the equator. In low Burger regimes the vorticity-based variables are no longer representative of the balanced and unbalanced dynamics and our assumption that these variables are uncorrelated breaks down. This is because the PV, or the balanced flow, is dominated by height in this regime, something that the vorticity-based variables are not capable of representing. The reverse is true for the height/mass-based variables.

In this work we consider a new set of PV-based control variables proposed in [11] that should be valid across all dynamical regimes. In Chapter 2 we introduce some key dynamical concepts using the 2D SWEs, defined by equations (2.1) - (2.3). We define balance and make the link between PV and the balanced flow by means of a normal mode analysis of the 2D SWEs. In 2.3.3 we show that of the three normal modes of the linearised system, two are unbalanced and one balanced. The balanced mode is represented by the linearised PV. We then show that the linearised PV can be approximated well by vorticity in a high Burger regime and height in a low Burger regime.

In Chapter 3 we introduce the data assimilation problem and discuss 4D VAR. We describe the incremental 4D VAR algorithm when the control variable transform is included. We give an overview of the various implementations of the transform in operational centres around the world. We conclude that there is a deficiency in the definition of the balanced variables in all the current operational control variables; all current control variables are either vorticity-based or height/mass-based. These

definitions of balance are not appropriate in all Burger regimes and therefore can potentially violate the assumption that the control variables are uncorrelated. Next, we introduce the PV-based control variable transform, which should address these problems. We discuss the current work in this area and motivate our research.

We choose to investigate the control variables in the framework of the 1D SWEs, equations (4.1) - (4.3). This is a relatively simple model that retains key dynamics similar to those of the full atmosphere. In Chapter 4 we describe the numerical model and demonstrate it properties. Two vital equations are the linear balance equation (4.5) and the PV equation (4.8). We then develop both the tangent linear and adjoint models required for the data assimilation.

In Chapter 5 we present the key relationships and approximations we need to derive the vorticity and PV-based transforms in the framework of the 1D SWEs. We derive the transforms and their numerical implementation. We then propose an approximate version of the PV-based transform where

$$\bar{q} \approx \frac{f}{<\bar{h}>}.$$

This approximation will improve the efficiency of the transform. We then test the numerical implementation of the transforms and make a qualitative comparison of their properties.

We present new results in Chapter 6 concerning the validity of the fundamental assumption that the control variables are uncorrelated. We introduce a statistical method, the 'quick-covs method', which we use to generate statistical data. We then test the correlation of the control variables in a variety of dynamical regimes. Next, we generate auto-correlations for each variable.

Finally, we develop a complete incremental 4D VAR data assimilation system for the 1D SWEs, which implements both the vorticity and PV-based control variable transforms. In Chapter 7 we consider simple assimilation experiments with the system where we have a single observation at the initial time $t = 0$.

We now discuss the conclusions we are able to draw from the new results of this

work.

8.1 Conclusions

In this work we make a new link between the theory of the vorticity and PV-based transforms behaviour in different dynamical regimes and what may be verified experimentally. Our new results extend the theory and experiments that are presented in [11], [67] and [68] by quantifying the accuracy of our assumptions regarding the balanced and unbalanced variables. In addition, we implement the control variable transforms into an assimilation system where they may be easily studied.

We make a qualitative comparison of the control variables that demonstrates the PV-based variables provide a better representation of the balanced and unbalanced dynamics than the vorticity-based variables in a low Burger regime. We show that the PV-based variables represent the balanced and unbalanced parts of the flow correctly in both and high and low Burger regime. However, in the low Burger regime the vorticity-based variables fail to capture the balanced flow, and in fact locate this flow in a region that is actually unbalanced.

We show that the PV-based control variable transforms imply background error statistics that include state dependence through the linearisation state PV. The vorticity-based variables imply background error statistic that are static.

The main new results of our research test the validity of the assumption that the control variables are uncorrelated. We demonstrate that whilst both the vorticity and PV-based control variables are uncorrelated in a high Burger regime, only the PV-based variables remain uncorrelated in a low Burger regime. The vorticity-based variables show significant negative correlations. Therefore, they are not a valid choice of control variables in this regime. The PV-based variables on the other hand are valid. We are able quantify the accuracy of the fundamental assumption that the balanced and unbalanced flows are uncorrelated by the correlations of the balanced and unbalanced control variables in both the vorticity

and PV-based variables. These are new results that are in agreement with the theory regarding the vorticity and PV-based variables in different dynamical regimes.

We also show that the approximation we propose to the PV-based transform, $\bar{q} \approx \frac{f}{<h>}$, has little affect on the correlations of the balanced and unbalanced PV-based variables in all the regimes we test. Therefore, the approximate PV-based control variables also are valid in all regimes. This will be convenient in more complex systems.

The auto-correlations we generate for each control variable show that these structures are highly dependent on the dynamical regime and it is not correct to use the same auto-correlations in all regimes, as is done operationally. They also demonstrate how successfully each set of control variables represents the balanced and unbalanced flows in each regime. We show that in a low Burger regime the unbalanced height PV-based variable has a correlation scale less than the Rossby radius as we would expect from a truly unbalanced variable. The residual height vorticity-based variable on the other hand has a length scale that is much larger, which is in fact more appropriate for the balanced variable.

We consider data assimilation experiments with a single observation at the initial time. Using the BLUE modified to include the control variable transform, we show that a fundamental difference between the vorticity and PV-based control variables is that when we have a single v observation the vorticity-based variables can only produce a balanced increment to the model variables v and ϕ. However, the PV-based variables can produce both the balanced and unbalanced increments. The superior representation of balance and unbalanced dynamics provided by the PV-based transform will improve the analysis increments produced by the data assimilation since both balanced and unbalanced errors can be corrected when a v observation is available.

Our single observation experiments reinforce the theory and also demonstrate the differences in the vorticity and PV-based variables. Both the control variables produce very similar analysis increments in a high Burger regime. However, in a

low Burger regime we demonstrate that the analysis increments produced using the PV-based variables are are different to those produced when the vorticity-based variables are used. In this regime the PV-based increments that appear to be better representations of the balance in the system.
In addition, using the assimilation system that we develop we have been able to demonstrate the potential numerical problems that we see might exist from the BLUE analysis. The auto-correlation matrices must be valid correlation matrices and also the weighting of background terms must be equal to have good convergence. Further work is needed to precondition the problem before more extensive assimilation experiments can be run.

8.2 Further Work

The main task for further research is to investigate the impact of the control variable transforms on the data assimilation in a more realistic setting. This would involve an incomplete set of observations and assimilations run using the vorticity and PV-based control variables in high and low Burger regimes. Identical twin experiments would be run where we perturb a known true state to obtain a background field. This background field would then be used as an initial guess in the assimilation experiments. There are many ways that we could choose to perturb the truth; we suggest the following:

1. Use a background field from the same true trajectory but valid at a different time.

2. Add a perturbation to the truth such that the perturbation to v and ϕ is balanced.

3. Add a perturbation to the truth such that the perturbation to v and ϕ is unbalanced.

Suitable methods to assess the quality of the analysis and the performance of the data assimilation then need to be used. There are a variety of ways we propose to

do this:

1. Compare the root mean square of the analysis error $\mathbf{x}_0^a - \mathbf{x}_0^t$, where $\mathbf{x}_0^t$ is the truth at time $t = 0$, when each control variable transform is used.

2. Compare the root mean square of the forecast error $\mathbf{x}_m^f - \mathbf{x}_m^t$, where $\mathbf{x}_m^f$ is the forecast at time $t = t_m$, generated from each analysis.

3. Assess how accurately the analysis captures the balanced mode.

   Since the PV represents the balanced mode we can consider the root mean square error between the analysis PV and the true PV. We can also look at the size of the departure from linear balance in the truth and the analysis.

4. Compare the convergence properties and computational efficiency of the assimilation using each control variable.

5. Consider the conditioning of the minimisation by generating matrix forms of each transform and computing the Hessian matrix of the incremental cost function.

   For the PV-based variables we would have to do this for each dynamical state separately by including the linearised PV. We could then also examine how the conditioning changes as the outer loops proceed.

Before these experiments can be run we must first address the numerical issues in the minimisation since a very large number of iterations is required. This would involve finding a better model for the auto-correlations. This might be achieved by developing a method to adjust the experimental auto-correlation matrices so that they can be used in the assimilation. We could also find an alternative way to represent these correlations that respects the characteristics of each of the variables auto-correlations. We would also implement further transforms to precondition the problem and diagonalise the block components of the auto-covariances. This would produce a cost function that is equally weighted and thus improve the convergence of the minimisation.

Once we have a good representation of the auto-correlations and a well conditioned cost function the assimilation system can be used to perform further statistical experiments with methods such as the NMC method. We might also consider the correlations of the two unbalanced variables as well. We know from the normal mode analysis that the two unbalanced modes are combinations of the geostrophic departure and the divergence, so the unbalanced variables might have some correlation. It would be interesting to quantify this for both the control variable transforms.

We could then consider additions to the PV-based transform. We could split the u velocity into balanced and unbalanced variables,

$$u = u_b + u_u,$$

using an appropriate equation with which to define the balanced component of u. We can then assess the impact of this on the behaviour of the PV-based transforms in the same way as done for the standard PV-based transform. These ideas are considered in more detail in [21].

Finally, we would like to examine the effect of the transforms in a full operational forecasting model or at least a more sophisticated model than the 1D SWEs. The same methods we employ here can be used to assess the performance of the variables in this framework.

Appendix A

Time Correlations

A.1 High Burger Regime

We can provide additional evidence that a time-differencing interval of approximately $300s$ is appropriate by calculating correlations of each variable in time. We expect to see the correlations decaying with time. If there is a dominant oscillation that is producing an artificial signal in the data then this will dominate the correlations. We therefore look for any unrealistic time-correlations as we vary the time interval τ.

Covariance for the time correlations are calculated in much the same way as in the correlation experiment described previously. We again assume that each component of each vector is equivalent to one realisation of a single random variable. So therefore we have a set of $N \times M$ realisations of each random variable where N is the number of points in the domain and M is the number of time differences, for example

$$\psi^{+0} = \left\{\psi_1^1, \ldots, \psi_N^1, \psi_1^2, \ldots, \psi_N^2, \ldots\ldots, \psi_1^M, \ldots, \psi_N^M\right\}, \tag{A.1}$$

and

$$\psi^{+j} = \left\{\psi_1^{1+j}, \ldots, \psi_N^{1+j}, \psi_1^{2+j}, \ldots, \psi_N^{2+j}, \ldots\ldots, \psi_1^{M+j}, \ldots, \psi_N^{M+j}\right\}, \tag{A.2}$$

where the subscript indicates the grid point and the superscript the difference field index. So here we are calculating the covariance of the same variable for

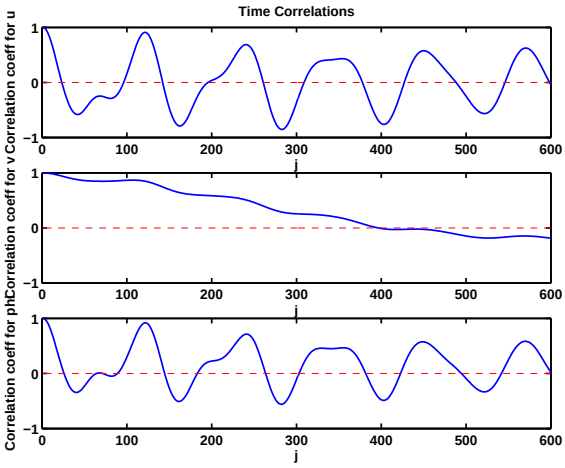

Figure A.1: Inappropriate Time Interval: Time correlations with the time-difference interval $\tau = 100s$. The correlation coefficient between two time-difference fields $j\tau$ apart is plotted against j.

time-differences $j\tau$ apart. The covariance is calculated as follows

$$COV^{+j}(\psi', \psi'^{+j}) =< (\psi' - < \psi' >)(\psi'^{+j} - < \psi'^{+j} >) > \tag{A.3}$$

where the superscript j indicates that we are calculating the covariance of time-differences $j\tau$ apart. By calculating the correlation co-efficient we can plot correlation against j, the temporal separation of the time-differences.
The time correlation experiments are run with several time intervals in line with the observations of the dominant oscillation. We have also included a plot of time-correlations against j where we have chosen a sampling time interval that is not in line with our observations.
The time difference interval in figure A.1 is one inertial period, or $100s$ (40 time steps, $\Delta t = 2.5s$). We expect the correlations of each variable to decay with time. The results give time-correlations that are swamped by another signal and do not decay with time. This demonstrates the QCM's sensitivity to this dominant signal and that this time difference interval is not appropriate.
The time correlations for time-difference intervals of $287.5s, 312.5s$ and $337.5s$ are shown in figures A.2 to A.4. These particular time-difference intervals are chosen to be around $300s$ following observations of the dominant signal and the

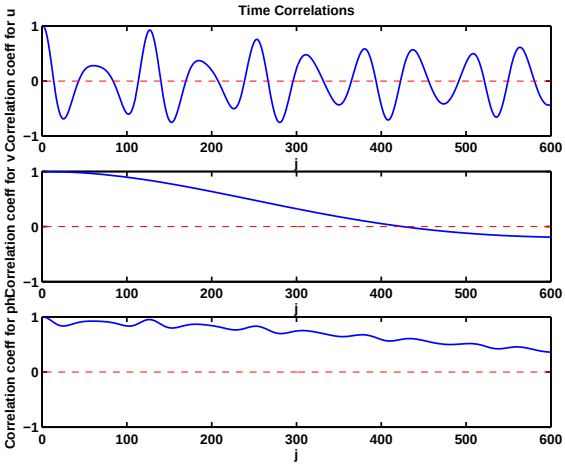

Figure A.2: Time correlations with time-difference interval $\tau = 287.5s$.

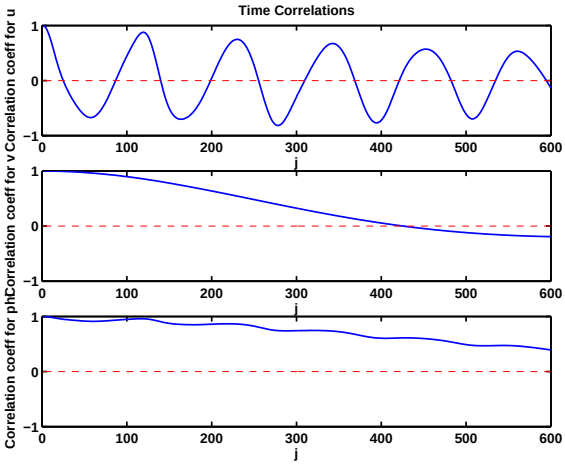

Figure A.3: Time correlations with time-difference interval $\tau = 312.5s$.

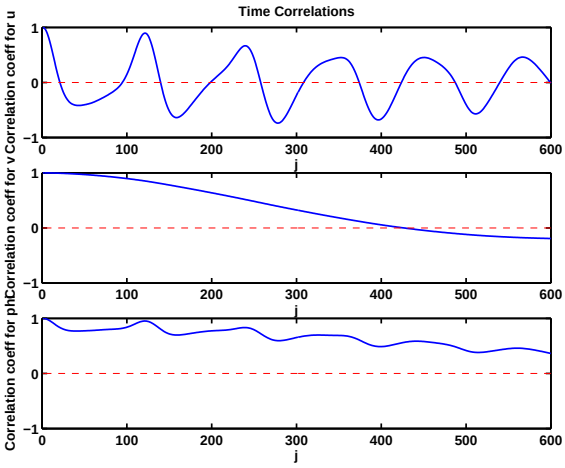

Figure A.4: Time correlations with time-difference interval $\tau = 337.5s$.

hypothesis that it is caused by gravity wave periodicity. The time interval that is most effective at filtering the wave is $312.5s$ or 125 time steps. We observe smooth correlation structures and start to lose the smooth structure for intervals either side of $312.5s$.

We notice that the correlations of ϕ decay much faster with an interval of $312.5s$ than $100s$, and are no longer dominated by the inertial-gravity signal. Also the correlations of u are smoother, more coherent and decaying slowly. Therefore choosing a time-difference interval of $312.5s$ has given the best results.

A.2 Time Correlations: Low Burger Regime

For the low Burger regime, figure A.5, we see that the time-correlations decay as we would expect and do not appear corrupted by the gravity wave signal.

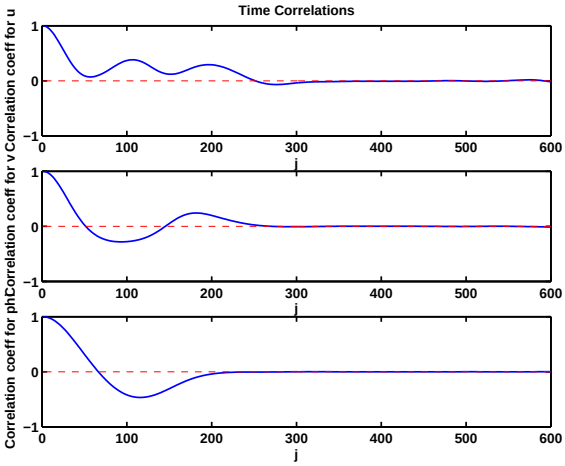

Figure A.5: Low Burger regime: Time-correlations for $T = 100s$.

Appendix B

Gradient Test Results

In figures B.1 to B.4 we see the results of the gradient test when the vorticity-based and PV-based transforms are used in the data assimilation. Figures B.1 and B.3 are plots of the quantity $E(\alpha)$, as defined in section 7.1 and [44], against perturbation size and figures B.2 and B.4 are plots of $E(\alpha) - 1$ against the perturbation size. We expect to see the value of $E(\alpha)$ approach 1 as α decreases. The results agree with those described in [44].

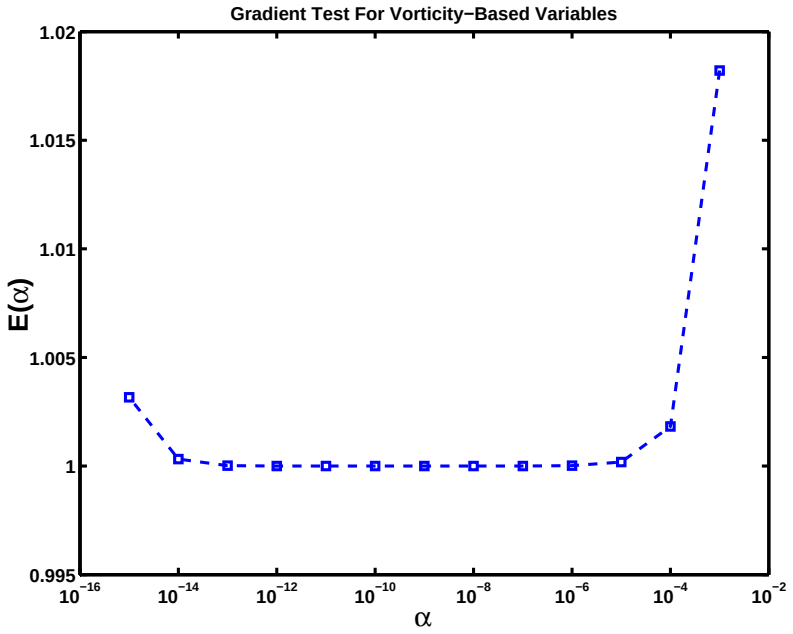

Figure B.1: Gradient Test Results: Vorticity-based variables.

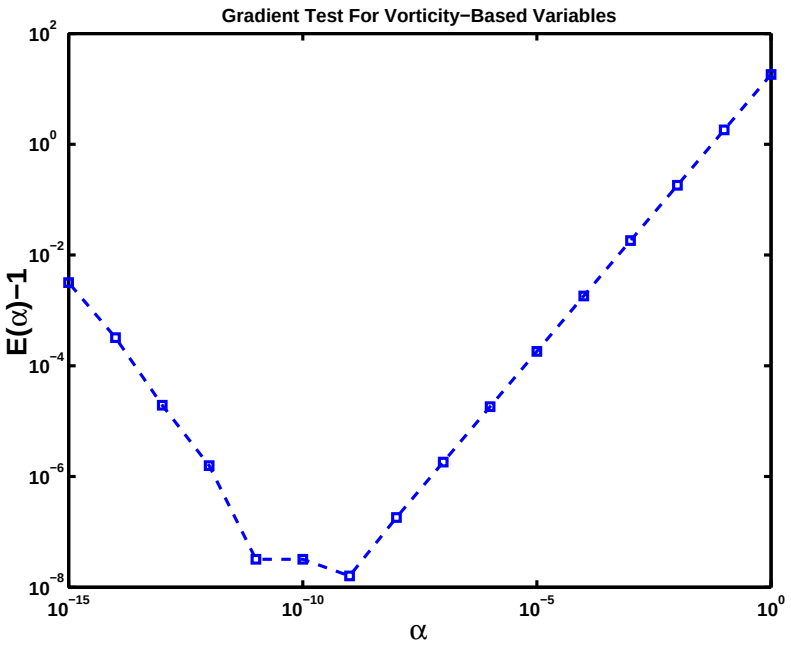

Figure B.2: Gradient Test Results: Vorticity-based variables.

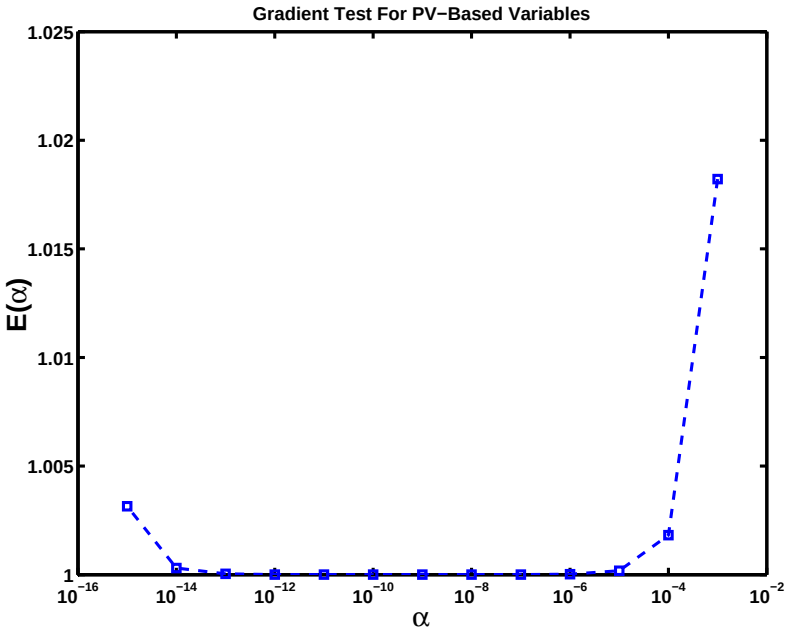

Figure B.3: Gradient Test Results: PV-based variables.

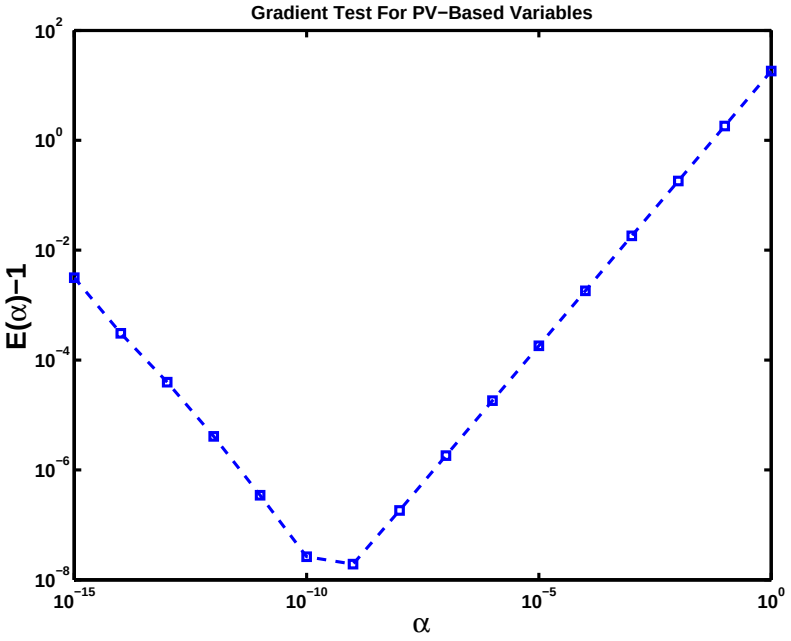

Figure B.4: Gradient Test Results: PV-based variables.

Appendix C

Modelled Auto-Correlations

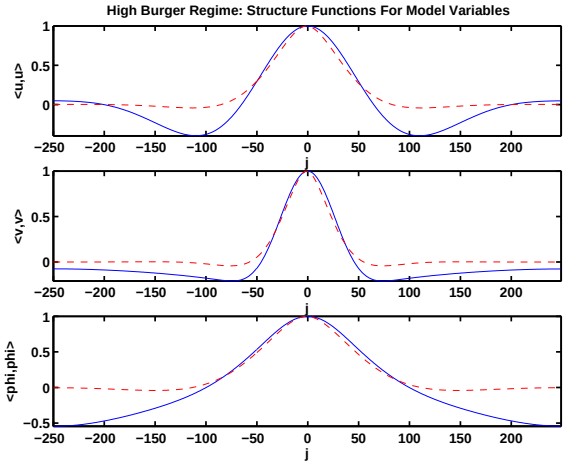

Figure C.1: High Burger regime, model variables:
Auto-correlations plotted with Laplace-smoother approximation.

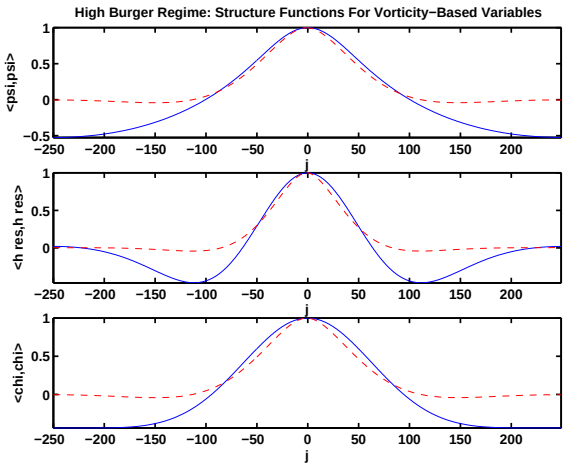

Figure C.2: High Burger regime, Vorticity-based variables:
Auto-correlations plotted with Laplace-smoother approximation.

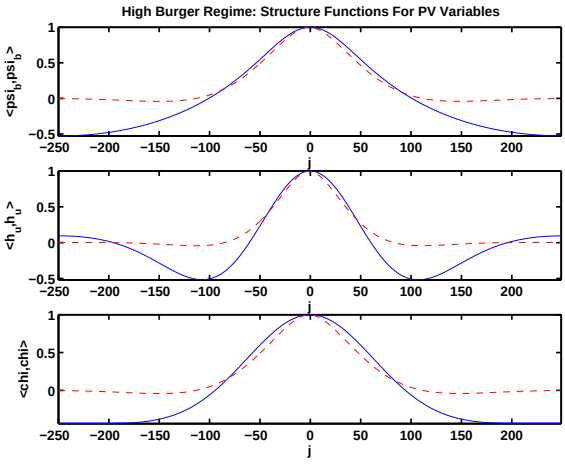

Figure C.3: High Burger regime, PV-based variables:
Auto-correlations plotted with Laplace-smoother approximation.

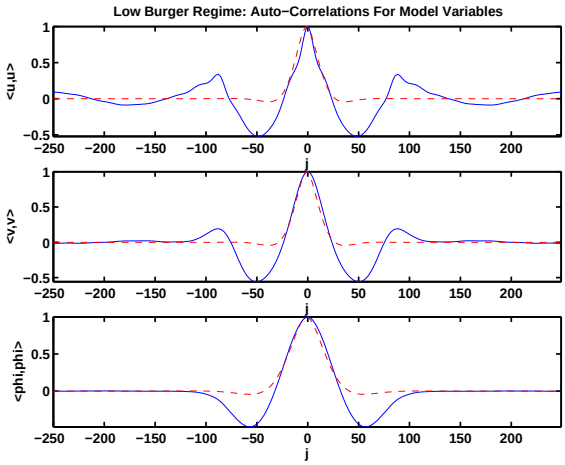

Figure C.4: Low Burger regime, model variables: Auto-correlations plotted with Laplace-smoother approximation.

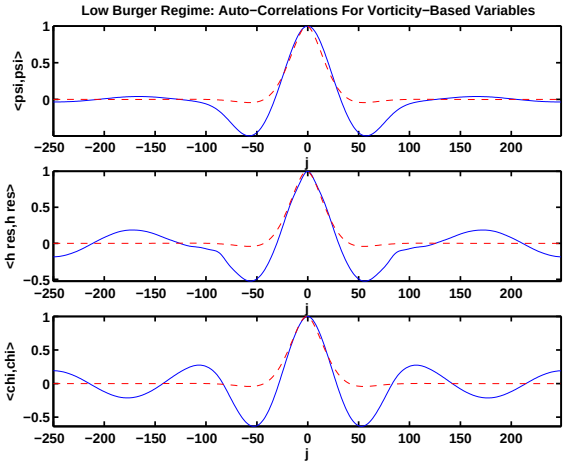

Figure C.5: Low Burger regime, Vorticity-based variables: Auto-correlations plotted with Laplace-smoother approximation.

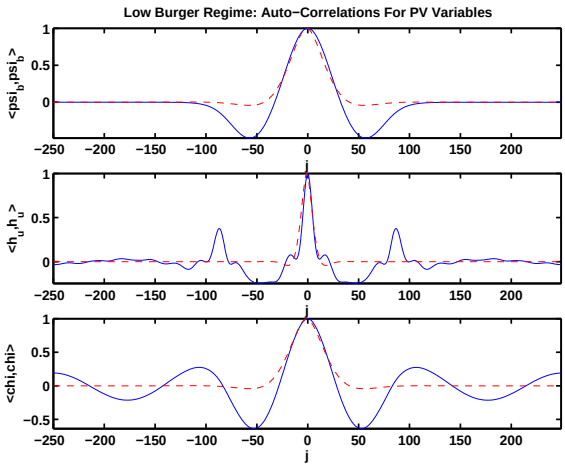

Figure C.6: Low Burger regime, PV-based variables: Auto-correlations plotted with Laplace-smoother approximation.

Glossary of Symbols and Acronyms

Continuous Model Related

h	Fluid Depth
u	Velocity in x direction
v	Velocity in y direction
ϕ	Geopotential Height
$\widetilde{H}$	Orographic Height
f	Coriolis Parameter
g	Gravitational Acceleration
U_c	Forcing mean flow
$\mathcal{D}$	Divergence
ζ	Relative Vorticity
q	Potential Vorticity
η	Departure of free surface from rest level
ψ	Streamfunction
χ	Velocity Potential
R_0	Rossby Number
B_u	Burger Number
F_r	Froude Number
L_r	Rossby Radius of Deformation
c_g	Gravity Wave Speed
H	Characteristic Height
D	Characteristic Depth
L	Characteristic Horizontal Length
U	Characteristic Horizontal Velocity
T	Characteristic Time
N_0	Characteristic Surface Height Departure

Numerical Model Related

Δx	Spatial step size
Δt	Temporal step size
x_i	Distance $i\Delta x$
t_m	Time $m\Delta t$
ϕ_i^j	Variable, ϕ in this example, at time t_j and position x_i

Data Assimilation

J	Cost function
$\widetilde{J}^{(k)}$	Incremental cost function on outer loop k
$\nabla \widetilde{J}^{(k)}$	Gradient of incremental cost function on outer loop k
$\mathbf{x}_i$	Model state at time $t = t_i$
$\mathbf{x}^b$	Background field
$\mathbf{y}_i^o$	Observation vector at $t = t_i$
$\mathbf{B}$	Background error covariance matrix
$\mathbf{R}_i$	Observation error covariance matrix at time $t = t_i$
$\mathcal{H}_i$	Observation operator at time $t = t_i$
$\mathcal{M}_i$	Non-linear model evolution to time $t = t_i$
$\mathbf{H}_i$	Linearised observation operator at time $t = t_i$
$\mathbf{M}_i$	Tangent Linear model evolution from time t_{i-1} to t_i
$\mathbf{M}_i^T$	Adjoint model at time $t = t_i$
$\mathbf{d}_i$	Innovation vector at time $t = t_i$
$\mathbf{x'}_0^{(k)}$	Model variable increment at time $t = t_0$ for outer loop k
$\mathbf{x'}^b$	Background increment

Control Variable Transforms

$\mathbf{z'}_0^{(k)}$	Control variable increment at time $t = t_0$ for outer loop k
$\mathbf{z'}^b$	Background increment in control space

$\mathbf{U}$	U-transform
$\mathbf{U}^T$	Adjoint of U-transform
$\mathbf{T}$	T-transform
$\mathbf{\Lambda}$	Block diagonal auto-covariance matrix
$\hat{\mathbf{\Lambda}}$	Block diagonal auto-correlation matrix
σ^2	Variance of control variable indicated by subscript
${}_a\zeta$	Geostrophic departure
h_{res}	Residual height vorticity-based control variable
h_b	Balanced height variable
h_u	Unbalanced height PV-based control variable
v_b	Balanced component of v
v_u	Unbalanced component of v
ψ_b	Balanced streamfunction PV-based control variable
ψ_u	Unbalanced streamfunction PV-based control variable

Acronyms

PV	Potential Vorticity
NWP	Numerical Weather Prediction
HIRLAM	High Resolution Limited Area Model
ECMWF	European Centre for Medium Range Weather Forecasting
SWEs	Shallow Water Equations
4D VAR	Four-Dimensional Variational Data Assimilation
3D VAR	Three-Dimensional Variational Data Assimilation
TLM	Tangent Linear Model
AM	Adjoint Model
SLSI	Semi-implicit, semi-Lagrangian scheme
DLB	Departure from Linear Balance

Bibliography

[1] E. Andersson, J. Haseler, P. Unden, P. Courtier, G. Kelly, D. Vasiljevic, C. Brankovic, C. Cardinali, C. Gaffard, A. Hollingsworth, C. Jakob, P. Janssen, E. Klinker, A. Lanzinger, M. Miller, F. Rabier, A. Simmons, B. Strauss, J. N. Thepaut and P. Viterbo: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III: Experimental results. *Quart. J. Royal Met. Soc.*, 124:pp. 1831-1860 (1998)

[2] A. F. Bennett: Inverse Modelling of the Atmosphere and Oceans. *Cambridge University Press* (2002)

[3] O. Bokhove: Slaving Principles, Balanced Dynamics, and the Hydrostatic Boussinesq Equations. *Journal of the Atmospheric Sciences*, 54: pp. 1662-1674 (1995)

[4] W. C. Choa, and L-P Chang: Development of a four dimensional Variational Analysis System Using the Adjoint Method at GLA. Part1: Dynamics. *Monthly Weather Review* volume 120. (1992)

[5] P. Courtier and O. Talagrand: Variational assimilation of meteorological observations with the direct and adjoint shallow water equations. *Tellus*, 42A:pp. 531-549 (1990)

[6] P. Courtier, E. Andersson, W. A. Heckley, G. Kelly, J. Pailleux, F. Rabier, J. N. Thepaut, P. Unden,

D. Vasiljevic, C. Cardinali, J. Eyre, M. Hamrud, J. Haseler, A. Hollingsworth, A. P. McNally and A. Stoffelen: Variational assimilation at the ECMWF. *ECMWF Tech. Memo.*, 194 (1993)

[7] P. Courtier, J. Thepaut and A. Hollingsworth: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Royal Met. Soc.*, 120:pp. 1367-1387 (1994)

[8] P. Courtier, E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier and M. Fisher: The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. *Quart. J. Royal Met. Soc.*, 124:pp. 1783-1807 (1998)

[9] M. J. P. Cullen: *A Mathematical Theory of Large-Scale Atmosphere / Ocean Flow*, Imperial College Press (2006)

[10] M. J. P. Cullen: Large-scale non-turbulent dynamics in the atmosphere. *Quart. J. Royal Met. Soc.*, 128:pp. 2623-2639 (2002)

[11] M. J. P. Cullen: Four-dimensional variational data assimilation: A new formulation of the background-error covariance matrix based on a potential vorticity representation. *Quart. J. Royal Met. Soc.*, 129:pp. 2777-2796 (2003)

[12] R. Daley: Atmospheric Data Analysis. *Cambridge University Press* (1993)

[13] R. Daley: Generation of Global Multivariate Error Covariances by Singular-Value Decomposition of the Linear Balance Equation. *Monthly Weather Review* volume 124:pp. 2574-2587 (1997)

[14] T.Davies, M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White and N. Wood A new dynamical core

for the Met Office's global and regional modelling of the atmosphere. *Quart. J. Royal Met. Soc.*, 131:pp 1759-1781 (2005)

[15] J. Derber: A variational continuous assimilation technique. *Monthly Weather Review* volume 117:pp. 2437-2446 (1989)

[16] J. Derber and F. Bouttier: Formulation of the background error covariances in the ECMWF global data assimilation system. *Tellus*, 51A:pp. 195-221 (1999)

[17] Dale R. Durran: Numerical Methods for Wave Equations in Geophysical Fluid Dynamics. *Springer* (1999)

[18] A. R. Mohebalhojeh and D. G. Drischel: Hierarchies of Balance Conditions for the f-plane Shallow Water Equations. To be published in *Journal of the Atmospheric Sciences*

[19] Mike Fisher: Background Error Covariance Modelling. *ECMWF Data Assimilation Training Course* (2003)

[20] Mike Fisher: Background error covariance modeling. *Proc. ECMWF Seminar on "Recent Developments in Data Assimilation for Atmosphere and Ocean"* 8-12 Sept 2003, Reading, U.K., 45-63.

[21] S. J. Fletcher: PhD Thesis: Higher Order Balance Conditions Using Hamiltonian Dynamics for Numerical Weather Prediction. *University of Reading* (2004)

[22] P. Gauthier, C. Charette, L. Fillion, P. Koclas and S. Laroche: Implementation of a 3D Variational Data Assimilation System at the Canadian Meteorological Centre. Part 1: The Global Analysis. *Atmosphere-Ocean* 37(2), 103-156 (1999)

[23] R. Giering and T. Kaminski: Recipes For Adjoint Code Construction. *Max-Planck institute for Tech.* (1998)

[24] Adrian E. Gill: Atmosphere-Ocean Dynamics. *Academic Press* (1982)

[25] N. Gustafsson, L. Berre, S. Hornquist, X.-Y. Huang, M. Lindskog, B. Navascues, K. S. Mogensen and S. Thorsteinsson:
Three-dimensional variational data assimilation for a limited area model.

Part 1: General formulation and the background error constraint. *Tellus*, 38A:pp. 111-136 (1986)

[26] N. J. Higham: Computing the nearest correlation matrix - a problem from finance. *IMA Journal of Numerical Analysis*, 22:pp. 329-343 (2002)

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

[28] M. Hortal: The development and testing of a new two-time-level semi-Lagrangin scheme (SETTLS) in the ECMWF forecast model. *Monthly Weather Review* volume 128:pp. 1671-1687 (2002)

[29] D. D. Houghton and A. Kasahara: Nonlinear shallow fluid flow over an isolated ridge. *Commun. Pure Appl. Math.*, 21:1-23 (1968)

[30] N. Bruce. Ingleby: The statistical structure of forecast errors and its representation in The Met Office Global 3-D Variational

Data Assimilation Scheme. *Quart. J. Royal Met. Soc.*, 127:pp 209-231 (2001)

[31] C. Johnson: PhD Thesis: Information Content of Observations in Variational Data Assimilation. *University of Reading* (2003)

[32] C. Johnson, Brian. J. Hoskins and N. K. NICHOLS: A Singular Vector Perspective of 4D-Var: Filtering and Interpolation. *Quart. J. Royal Met. Soc.*, 131:pp 1-19 (2005)

[33] Eugenia Kalnay: Atmospheric Modeling, Data Assimilation and Predictability. *Cambridge University Press* (2003)

[34] A. Lawless: PhD Thesis:. *University of Reading* (1999)

[35] A. S. Lawless, N. K. Nichols and S. P. Ballard: A comparison of two methods for developing the linearization of a shallow water model. *Quart. J. Royal Met. Soc.*, 129:pp 1237-1254 (2003)

[36] A. S. Lawless, S. Gratton and N. K. Nichols: An investigation of incremental 4D-Var using non-tangent linear models. *Quart. J. Royal Met. Soc.*, 131:pp 459-476 (2005)

[37] A. S. Lawless and N. K. Nichols: Inner loop stopping criteria for incremental four-dimensional variational data assimilation. To appear in *Monthly Weather Review* (Preprint - Numerical analysis report 5/05)

[38] F. X. Le Dimet and O. Talagrand: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, 38A:97110 (1986)

[39] J. Lewis and J. Derber: The use of adjoint equations to solve a variational adjustment problem with advective constraint. *Tellus*, 37A:pp. 309-322 (1985)

[40] A. C. Lorenc, S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, F. W. Saunders: The Met. Office Global 3-Dimensional Variational Data Assimilation Scheme. *Quart. J. Royal Met. Soc.*, 126:pp 2991-3012 (2000)

[41] A. C. Lorenc: Modelling of error covariances by 4D-Var data assimilation. *Quart. J. Royal Met. Soc.*, 129:pp 3167-3182 (2003)

[42] E, N. Lorenz: The Essence of Chaos. *UCL Press,* pp. 227

[43] I. M. Navon and D. M. Legler: Conjugate-Gradient Methods for Large-Scale Minimization in Meteorology. *Monthly Weather Review*, 115:pp. 1479-1502 (1987)

[44] I. M. Navon, X. Zou, J. Derber and J. Sela: Variational data assimilation with an Adiabatic Version of the NMC Spectral Model. *Monthly Weather Review*, 120:pp. 1433-14466 (1992)

[45] by A. V. Oppenheim and R. W. Schafer: *Digital Signal Processing*, Prentice-Hall (1975)

[46] C. A. Parrett and M. J. P. Cullen: Simulation of hydraulic jumps in the presence of rotation and mountains. *Quart. J. Royal Met. Soc.*, 110:pp. 147-165 (1984)

[47] D. Parrish and J. C. Derber: The National Meteorological Center's Spectral Statistical-Interpolation Analysis System. *Monthly Weather Review*, 120:pp. 1747-1763 (1992)

[48] T. Payne, Met Office: Exact Results on Control Variable Transforms using the Shallow Water Equations. *Personal Communication*

[49] J. Pedlosky: *Geophysical Fluid Dynamics*, Springer-Verlag (1979)

[50] S. Polavarapu: Divergent wind analysis in the oceanic boundary layer. *Tellus*, 47A:pp. 221-239 (1995)

[51] S. Polavarapu, S. Ren, Y. Rochon, D. Sankey, N. Ek, J. Koshyk and D. Tarasick: Data Assimilation with the Canadian Middle Atmosphere Model. *Atmosphere-Ocean* 43(1), 77-100 (2005)

[52] J. Pudykiewicz, R. Bonoit and A. Staniforth: Preliminary results from a partial LRTAP model based on an existing meteorological forecast model. *Atmosphere-Ocean* 23, 267-303 (1985)

[53] F. Rabier, H. Jarvinen, E. Klinker, J.-F. Mahfouf and A. Simmons: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Royal Met. Soc.*, 126:pp. 1143-1170 (2000)

[54] M. K. Ramamurthy and I. M. Navon: The Conjugate-Gradient Variational Analysis and Initialization Method: An Application to MONEX SOP 2 Data. *Monthly Weather Review*, 120:pp. 2360-2377 (1992)

[55] C. Rivest, A. Staniforth and A. Robert: Spurious resonant response of semi-Lagrangian discretizations to orographic forcing: diagnosis and solution. *Monthly Weather Review*, 122:pp. 366-376 (1994)

[56] A. Robert: A stable numerical integration scheme for the primitive meteorological equations. *Atmosphere-Ocean* 19, 35-46 (1981)

[57] Y. Sasaki: Some basic formalisms in numerical variational analysis. *Monthly Weather Review*, 98:875883 (1970)

[58] D. F. Shanno: On the convergence of a new conjugate gradient algorithm. *Siam Journal of Numerical Analysis* volume 15, 1247-1257 (1978)

[59] D. F. Shanno and K. H. Phua: Remark on algorithm 500 - a variable method subroutine for unconstrained nonlinear minimization. *ACM Trans. on Mathematical Software*, 6, 618-622 (1980)

[60] A. Staniforth and Jean Cote: Semi-lagrangian integration schemes for atmospheric models - A review. *Monthly Weather Review* volume 119. (1991)

[61] C. Temperton: Implicit Normal Mode Initialization. *Monthly Weather Review*, 116:pp. 1013-1031 (1988)

[62] M. Thurlow: VAR Scientific Documentation Paper 13: Control Variable Transforms. *Met Office Scientific Documentation* (1999)

[63] T. Warn, O. Bokhove, T. G. Shepherd and G. K. Vallis: Rossby number expansions, slaving principles, and balance dynamics. *Quart. J. Royal Met. Soc.*, 121:pp. 723-739 (1995)

[64] A. Weaver and P. Courtier: Correlation modelling on the sphere using a generalised diffusion equation. *Quart. J. Royal Met. Soc.*, 127:pp. 1815-1846 (2001)

[65] A. Weaver, J. Vialard and D. L. T. Anderson: Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part 1:

Formulation, Internal Diagnostics, and Consistency Checks. *Monthly Weather Review* 131:pp. 1360-1378 (2003)

[66] A. T. Weaver, C. Deltel, E. Machu, S. Ricci and N. Daget: A multivariate balance operator for variational ocean data assimilation. *ECMWF tech. memo.* 491. To be published in *Quart. J. Royal Met. Soc.*

[67] M. Wlasak: The examination of balanced and unbalanced flow using Potential Vorticity in Atmospheric Modelling. *PhD Thesis, University Of Reading* (2002)

[68] M. Wlasak, N. K. Nichols and I. Roulstone: Use of potential vorticity for incremental data assimilation. *Quart. J. Royal Met. Soc.*, 132:pp. 2867-2886 (2006)

[69] Yong Li et al: Four Dimensional Variational Data Assimilation experiments with a multilevel Semi Lagrangian Semi Implicit General Circulation Model. *Monthly Weather Review* volume 122, 966-983 (1993)

[70] M. Zupanski, D. Zupanski, T. Vukicevic, K. Els and T. Vonder Haar: CIRA/CSU Four-Dimensional Variational Data Assimilation System. *Monthly Weather Review* volume 133, 829-843 (2005)