Optimal Filtering of Orography for NWP and Climate Models

Ian C. Rutt

A thesis submitted for the degree of Doctor of Philosophy
Department of Meteorology
June 2003
Declaration

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

Ian C. Rutt
Abstract

Numerical models of the atmosphere are known to experience problems with near-grid-scale orographic forcing, particularly the formation of spurious grid-point storms. This can seriously undermine the accuracy and stability of model integrations, so possible methods for reducing them are of interest to all users of weather prediction and climate models. Previous studies indicate that smoothing the orographic field is effective in addressing these issues, and they motivate this work.

The representation of orographic forcing in a one-dimensional, linear shallow water model is studied. The most significant misrepresentation of orographic forcing is found on scales shorter than approximately six grid-lengths, which agrees with previous studies. However, this model is too simple to be of further use in studying the problem.

Two potential disadvantages of orographic smoothing are the loss of height from important barrier ridges and the adjustment of sea-points to non-zero height. To counter these effects, a new variational smoothing method is developed, which emulates a class of linear filters, but allows the imposition of other conditions on the smoothed orography. The properties of the method are explored analytically and confirmed in practice. Also discussed is the related problem of identifying those orographic ridges whose height should be maintained.

The variational method is then evaluated in a global, non-linear shallow water model, under a variety of flow regimes. The results highlight the importance of non-linearity and balance in determining the success of orographic smoothing; since NWP models are generally more non-linear and more balanced than the model used here, this evidence is taken to support the use of orographic smoothing in an NWP context. The benefits of extra smoothing constraints are weakly supported, but they need to be further evaluated in a more complete model.
Acknowledgements

This thesis could not have been written without the help and support of many people and organisations. First among these is my departmental supervisor, John Thuburn. His patient encouragement, insightful scholarship, and unflagging optimism were crucial throughout the course of my studies. I have been very fortunate to benefit from such generous and committed supervision.

I would like to thank the Natural Environment Research Council for providing the major part of my funding, and to the Met Office for generously supplementing it though a CASE award. In connection with this, I am grateful to Andrew Staniforth, my supervisor at the Met Office, for his support and guidance, and for the clarity he has brought to my often rather muddy thought processes.

Others in the Department have made helpful contributions as the work progressed: Ian James and David Stephenson, as members of my thesis committee, encouraged me to look at the work in a broader context, while Mike Blackburn’s encyclopedic knowledge of spectral models saved me from frustration several times.

I am indebted to my friends for their support and companionship, in good times and bad. James, Elspeth, Chris and Hilary were particularly good at keeping the stress levels down, while Susan, Jo, Ros and Ellie all helped me overcome pessimism and put my troubles into perspective. I’m also grateful to my numerous housemates over the years, who made Brighton Road a really pleasurable place to live, and to Giles for many entertaining and illuminating conversations.

Lastly, I wouldn’t have been able to get to this point without the love and support of my parents. I am so grateful to them for all the things they have done for me, including putting a roof over my head while I was writing up.
Dedicated to my parents, with gratitude,

and to the memory of Alan Harwood (1946–2003), friend and teacher.
The hills are shadows, and they flow
From form to form, and nothing stands.
They melt like mist, the solid lands,
Like clouds they shape themselves and go.

– Tennyson (In Memoriam)

‘I have endeavoured . . . to keep a steadier eye upon the general purpose and design. With this object in view, I have put a strong constraint upon myself from time to time, in many places; and I hope the story is better for it, now.’

– Charles Dickens (Preface to Martin Chuzzlewit)
Contents

1 Introduction 1

1.1 Problems with near-grid-scale orography 2

1.2 Orographic Smoothing 4

1.2.1 A possible solution? 4

1.2.2 Issues 5

1.2.3 Evaluating smoothed orography 5

1.2.4 Questions addressed by the thesis 6

1.3 Outline of thesis content 7

2 Experiments with a one-dimensional shallow water model 9

2.1 Introduction and aims 9

2.2 Model specification 10

2.2.1 Derivation of model equations 10

2.2.2 The model schemes 12

2.3 Steady state accuracy 15

2.3.1 Steady state for the continuous equations 15

2.3.2 Eulerian, linear model 16
2.3.3 Semi-Lagrangian, linear model .................................. 18
2.3.4 Conclusions ....................................................... 22
2.4 Transients and their accuracy ...................................... 26
  2.4.1 Introduction ................................................... 26
  2.4.2 Experimental design .......................................... 26
  2.4.3 Orography ...................................................... 28
  2.4.4 Results ........................................................ 28
2.5 Summary and conclusions .......................................... 36
3 Two-dimensional orography – a variational approach to smoothing 38
  3.1 Introduction ...................................................... 38
  3.2 Mathematical basis ............................................... 39
    3.2.1 Preliminaries ................................................ 39
    3.2.2 A simple example in one dimension ....................... 40
    3.2.3 A more general linear filter ............................... 41
    3.2.4 Discretization of the cost function ....................... 45
    3.2.5 Determining the Padé coefficients ......................... 46
    3.2.6 Minimising the cost function ............................ 48
  3.3 Minimising the cost function in two dimensions ............... 49
    3.3.1 ‘Planar’ smoothing ........................................ 49
    3.3.2 ‘Spherical’ smoothing ..................................... 51
    3.3.3 Practical minimization .................................... 65
### CONTENTS

3.4 Other constraints .................................................. 67
   3.4.1 Two different types of constraint ....................... 68
   3.4.2 Constraining barrier height ............................. 68
   3.4.3 Constraining the flatness of the sea .................... 69

3.5 Summary .......................................................... 70

4 Applying the variational method to real orography .............. 72
   4.1 Introduction .................................................... 72
   4.2 Grids in spectral models ..................................... 72
      4.2.1 Model truncation ........................................ 73
      4.2.2 Choosing a grid for a spectral model ................. 74
   4.3 Mean orography ................................................ 76
      4.3.1 Calculating the mean orography ....................... 76
      4.3.2 Orographic spectra ...................................... 76
   4.4 Methods for identifying important ridges .................. 78
      4.4.1 Local maxima ............................................ 79
      4.4.2 Perpendicular second derivatives ...................... 80
      4.4.3 Height threshold ....................................... 84
      4.4.4 Ridge searches ......................................... 85
      4.4.5 Choosing a ridge-finding method ..................... 91
   4.5 The variational smoothing method in practice ............... 94
      4.5.1 ‘Planar’ geometry ..................................... 95
      4.5.2 ‘Spherical’ geometry .................................. 102
   4.6 Summary and conclusions .................................... 112
5 Experiments with a global shallow water model 113

5.1 Introduction ................................................. 113

5.1.1 Background ............................................. 113

5.1.2 Aims and hypotheses ................................. 114

5.2 The model .................................................. 115

5.2.1 Description ............................................. 115

5.2.2 Testing ................................................. 118

5.3 Experimental method ...................................... 119

5.3.1 Basic considerations .................................. 119

5.3.2 Model configuration .................................... 122

5.3.3 Flow regimes ......................................... 124

5.3.4 The effect of initialisation on model balance ....... 128

5.3.5 Quantitative comparison between model runs ...... 132

5.3.6 Procedure for exploring the parameter space ...... 134

5.4 Results ..................................................... 135

5.4.1 The control run ....................................... 135

5.4.2 Sensitivity of the model solution to $\gamma$ ........ 138

5.4.3 Sensitivity of the model solution to $\epsilon$ ........ 152

5.4.4 Sensitivity of the model solution to sea flatness .... 159

5.4.5 Discussion ............................................. 163

5.4.6 The impact on the results of the choice of diffusion ... 166

5.5 Summary and conclusions ............................... 167
6 Discussion and conclusions 169

6.1 Thesis summary . . . . . . . . . . . . . . . . . . . . . . . . . . . 169

6.2 Applications of this work . . . . . . . . . . . . . . . . . . . . . 171

6.3 Remaining questions and future work . . . . . . . . . . . . . . 172

6.4 Final thoughts . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 174
Chapter 1

Introduction

The presence of mountains and hills on the surface of the Earth has a profound effect on the atmosphere. At the most fundamental level, the orography forces the atmospheric system by causing parcels of air to be raised and lowered or deflected as they flow across the Earth’s surface. This forcing generates waves in the atmosphere, as well as having important thermodynamic effects, triggering convection and bringing air parcels to saturation. At a local level, a striking correspondence may be seen between the locations of maxima in the mean annual UK rainfall (see, for instance, Met Office (2003)) and the locations of mountainous areas. Globally, orography plays a significant role in the dynamics of the atmosphere. A good example of this is the generally zonal flow around Antarctica (see Turner (2002)), which depends on the approximately symmetrical distribution of orography in that region. This contrasts markedly with the situation around the North Pole, where the distribution of orography is far from symmetrical, and the flow is more complex. And, in numerous places on the globe, high orography (e.g. the Andes) effectively blocks the flow, and leads to a fundamentally different climatology from that which would exist if there were no mountains. On a more subtle level, but no less importantly, the waves generated by flow over orography transport angular momentum up through the atmosphere, and constitute a very important part of the global angular momentum budget (see, for example, Peixoto and Oort (1991)). Thus, in view of the big impact that orographic forcing has on the atmosphere, its representation in numerical models of the atmosphere is of great importance.
1.1 Problems with near-grid-scale orography

The representation of orography in NWP\(^1\) and climate models is problematic for several reasons, but probably the most important of these is the nature of the orographic spectrum. As shown by Balmino (1993), the spectrum of the Earth’s orography decreases quite slowly with increasing wavenumber, and this means that, whatever the resolution of the model, there will always be a spectral component of the orography present near the grid-scale. Numerical models generally perform badly near the grid-scale, and so the presence of forcing at these scales is undesirable. In the simplest dynamics-only models\(^2\), these inaccuracies concern both freely propagating waves and the steady forced part of the motion. As demonstrated by, for instance, Durran (1999), numerical dispersion errors are generally at their most severe at the grid-scale; inaccuracies in the steady forced part of the solution are also at their worst near the grid-scale, as is discussed in chapter 2 of this thesis. Indeed, the representation of information at these scales is poor enough for Lander and Hoskins (1997) to term them ‘unbelievable’.

The forcing of model dynamics at scales which are poorly represented generates near-grid-scale noise. In itself, this is undesirable, as it contaminates the model output, and obscures our view of the flow at larger, ‘believable’ scales. However, in the more complicated models used for NWP and climate studies, other factors come into play which make the problem much more serious. As Lander and Hoskins (1997) point out, model parameterizations of such processes as convection are very sensitive to small changes in the fields that are passed to them, and so the noise associated with near-grid-scale forcing could be sufficient to erroneously trigger these parameterizations. This kind of process can lead to the ‘grid-point storms’ noted by Webster et al. (2003), which may generate vertical motions that violate the Courant-Friedrichs-Lewy (CFL) stability criterion, and thus cause the integration to fail. Even if such instability is not generated, the presence of grid-point storms in the model integration is very bad for the accuracy of the model solution. An example of this type of problem is shown in figure 1.1.

\(^1\)Numerical weather prediction.
\(^2\)Dynamics is the name given to the parts of the model which simulate the resolved flow; physics refers to sub-grid-scale parameterizations of such things as radiation, moisture, gravity-wave drag, etc.
Another problem, specific to semi-implicit semi-Lagrangian advection schemes, is known as spurious orographic resonance. This can occur when the Courant number is greater than 1. (The Courant number is defined as $u \frac{\Delta t}{\Delta x}$, where $u$ is the flow speed; $\Delta t$ and $\Delta x$ are the time-step and grid-spacing, respectively.) The problem of spurious orographic resonance was first noticed by Coiffier et al. (1987) and Kaas (1987), and is most closely associated with near-grid-scale orography. Numerical dispersion causes some waves to be stationary relative to the forcing. In this situation, resonance occurs, so that these waves can come to dominate the model solution; in extreme cases, they may breach stability criteria, and so cause the model to ‘blow up’. The most thorough analysis of the problem was undertaken by Rivest et al. (1994), who showed that it could be solved by off-centring the averaging operator in the scheme, and including a third time-level to maintain $O(\Delta t^2)$ accuracy. This approach was generalized by Côté et al. (1995), and may well provide the best solution to the problem, although work by Héreil and Laprise (1996) highlighted possible problems using it in more complicated models.
1.2 Orographic Smoothing

1.2.1 A possible solution?

The general problem with near-grid-scale noise, and the more specific spurious orographic resonance problem, both suggest that the smoothing of the orographic forcing field to remove the near-grid-scales may be beneficial to atmospheric model integrations. Although a possible solution to the orographic resonance problem has been found, it may also be alleviated to some degree by smoothing the orography, and this approach may avoid some of the possible problems with a solution based on off-centring the numerical scheme.

Most previous studies of orographic smoothing have been concerned with the removal of Gibbs ripples, present in the orography used in spectral transform models. Gibbs ripples are caused by truncating the spectrum of the orography, and then transforming it back onto the grid used in the model. They are particularly associated with rough terrain, which contains significant spectral components near the truncation limit. Gibbs ripples generate spurious waves in the model, and can cause problems with thermodynamics by elevating sea points above zero height, where they act as spurious heat-sources. Studies by Navarra et al. (1994) and Stephenson et al. (1998) indicate that orographic smoothing is of benefit in the context of spectral transform models.

Gibbs ripples present a fairly obvious case for the use of orographic smoothing, since they are clearly visible in plots of the orography (see, for instance, Bouteloup (1995)). However, as Webster et al. (2003) and others make clear, the problem of near-grid-scale orography is not confined to spectral transform models. Complementary to the work done by Lander and Hoskins (1997), on believable and unbelievable scales in spectral transform models, is the study by Davies and Brown (2001), which looked at the issue of near-grid-scale orography in a grid-point model. They studied idealized cases of flow over isolated hills and ridges, and looked at how the size of those hills relative to the grid affected the accuracy of the flow in the model. As a result of this, they came to the conclusion that it would be beneficial to remove scales shorter than six grid-lengths from the orographic field. Their suggestion of the Raymond filter
(Raymond 1988) as a suitably scale-selective candidate for this role was successfully taken up by Webster et al. (2003).

1.2.2 Issues

So, there are reasons to suppose that orographic smoothing might be worthwhile, and there have also been some studies which suggest that this is the case. However, orographic smoothing has the potential to introduce problems of its own into a numerical model. The first problem concerns the lowering of orographic barrier height by the smoothing process; a linear filter will generally reduce the height of local orographic maxima. Given the importance of barrier height to atmospheric dynamics outlined above, this could be a serious problem. In an extreme case, the smoothing could change the nature of the flow quite fundamentally, and so it would be good to find a way to avoid this problem.

The second potential problem concerns the raising of sea-points to non-zero heights. The boundary between the completely flat sea and the land is a sharp one, but, clearly, linear filtering will smudge it, with the effect that some sea points will no longer be at sea-level. Sea and land points are treated differently in NWP and climate models, and, as Navarra et al. (1994) point out, raised sea points act as spurious heat sources. Webster et al. (2003) note two possible solutions to this problem: the elevated sea points may either be set to zero manually (though this is only possible in a grid-point model), or have their surface temperature adjusted to take account of their new height. The disadvantage with the former method is clear – grid-scale cliffs are introduced into the smoothed orography – but Webster et al. (2003) found that it was no less effective than adjusting the surface temperature. However, it might be advantageous if some means could be found to avoid the problem in the first place, since both ‘fixes’ are non-ideal.

1.2.3 Evaluating smoothed orography

As Davies and Brown (2001) point out, the evaluation of smoothed orography in a full NWP model is not straightforward. Firstly, any comparison between model runs
with and without orographic smoothing is necessarily subjective. Given the impossibility of analytically calculating the correct solution, and the difficulty of generating a well-resolved, high-resolution solution, one is reduced to making a judgement about which output ‘looks more realistic’. While comparisons could be made between the model output and the actual evolution of the atmosphere recorded in synoptic analyses, these analyses rely on the same numerical model for their generation, which would clearly be unsatisfactory from an experimental design point of view. The second problem with an evaluation made in an NWP model is that the model’s sub-grid-scale parameterizations will have been tuned to work optimally with the unsmoothed orography. Thus, when the orography is smoothed, the parameterizations will not necessarily be configured correctly for the new orography, and the introduction of smoothing may thus appear to be detrimental to the accuracy of the solution.

These difficulties suggest that the evaluation of orographic smoothing would be best undertaken in the context of simplified models. Such models may not reproduce the problems seen in NWP models, such as grid-point storms, but at a more fundamental level, they suffer from near-grid-scale noise in exactly the same way. Moreover, a simple model may be run at high resolution to obtain a reference solution, and the absence of sub-grid-scale parameterizations avoids possible problems with their tuning. In addition, a simple model is easier to understand and analyse. For these reasons, the use of simple models is well-established in the testing of numerical schemes, and the study of problems with orography (see, for instance, Erbes (1993), Côté et al. (1995) or Lindberg and Alexeev (2000)).

1.2.4 Questions addressed by the thesis

From the arguments presented above, the underlying question which is the subject of this thesis was formulated, namely: what is the optimal orographic smoothing method for NWP and climate modelling applications? The term smoothing is used to emphasize that the question does not limit itself simply to linear filtering, but is concerned with a broader approach to the problem of removing the near-grid-scales. Knowledge of previous studies, and the desire to use simple models to aid understanding, led to these three specific questions:
• What is the optimal orographic filter to use with a one-dimensional, linear shallow water model?

• How can the orographic field be smoothed so that the near-grid-scale component is significantly lessened, but the flatness of the sea and the height of important barrier ridges are maintained?

• Can the lessons learned in answering the first two questions be applied to an evaluation of a new constrained smoothing method in a simple numerical model of the atmosphere?

Answering these more detailed questions should go some way towards addressing the broader question of optimal orographic smoothing, and greatly increase our knowledge of the problem. The content and layout of the thesis are thus determined by the need to answer these specific questions.

1.3 Outline of thesis content

The thesis begins with a study of orographic smoothing in a one-dimensional, linear shallow-water equation model, which is perhaps one of the simplest possible models. The results from this work agree broadly with those of Davies and Brown (2001), but nevertheless lead to the conclusion that a more complex model is needed to study the orographic smoothing problem properly.

Chapters 3 and 4 give an account of a new variational smoothing method which is designed to tackle the two potential problems of smoothing identified above. As well as being able to emulate a class of linear filters, additional constraints are described which enable the flatness of the sea and the height of important barriers to be maintained, with still further constraints possible, if needed. The properties of this method are elucidated, and practical issues concerning its application are addressed. Also discussed are possible means of identifying important ridges whose heights might be constrained by this method. Finally, the method is shown to work as expected, and its spectral properties are analysed.
The final major chapter of the thesis brings the themes of the previous chapters together in an experiment to evaluate the variational smoothing method in a global shallow water model. Smoothed orography generated by the variational method with a range of parameter combinations is tested under different flow regimes, and the model solutions are compared with high-resolution reference solutions. While the results do not show a clear benefit from smoothing in the context of a shallow water model, they point to the importance of non-linearity and balance in problems with near-grid-scale orography. The experiment gives some support to the idea that smoothed orography is beneficial to numerical model integrations, but also highlights the negative effects that such smoothing may have on the representation of some aspects of the flow.

In the concluding chapter, the implications of the work presented in the preceding four are summarized and some outstanding questions raised by the work are discussed. Although the thesis is broadly supportive of smoothing, and presents a new smoothing method which may be of use to the designers of NWP and climate models, more work needs to be done to enable the generation of optimally smoothed orography for operational use; some suggestions for achieving this are made at the end of the thesis.
Chapter 2

Experiments with a one-dimensional shallow water model

2.1 Introduction and aims

As explained in chapter 1, the problem of near-grid-scale orography is difficult to investigate in the context of a full NWP or climate model. Leading on from this is the idea that the problem might best be studied in a very simple model; this idea was pursued in a study by Davies and Brown (2001). However, although the model they used was simple in comparison with an NWP model, it was nevertheless a three-dimensional model with non-linear dynamics. The purpose of the present work is to take the rationale of using simple models a stage further. Taking one of the simplest possible models, a linear, one-dimensional shallow water model, the properties of the steady flow and transient response are studied to determine what the optimal form of orographic smoothing would be for this model. The starting point of this work was the development of model code to perform this task; however, due to the linear nature of the model, it was convenient to undertake much of the investigation by analysing the governing equations, rather than running the model.
2.2 Model specification

2.2.1 Derivation of model equations

The non-linear shallow water equations on a rotating $x$-$y$ plane may be written as follows:

\begin{align}
\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + fu + g \frac{\partial \eta}{\partial y} &= 0, \\
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - fv + g \frac{\partial \eta}{\partial x} &= 0, \\
\frac{\partial \eta}{\partial t} + u \frac{\partial \eta}{\partial x} + v \frac{\partial \eta}{\partial y} + (\eta + H - h_0) \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) &= v \frac{\partial h_0}{\partial x} + v \frac{\partial h_0}{\partial y}.
\end{align}

In these equations, $H$ is the height of the undisturbed fluid surface above a fixed datum, $h_0(x,y)$ is the height of the orographic surface above the same datum, $\eta(x,y,t)$ is the height of the perturbation to the fluid surface, $u(x,y,t)$ is the velocity in the $x$-direction and $v(x,y,t)$ is the velocity in the $y$-direction. The acceleration due to gravity is denoted $g$, and $f$ is the Coriolis parameter. The general arrangement of the domain is shown schematically in figure 2.1.

The terms on the right-hand side of equation (2.3) are the orographic forcing terms, which appear here because the equations have been formulated with surface displacement ($\eta$) as the height variable. The same formulation is used by Lindberg and Alexeev (2000) in their study of spurious orographic resonance. However, some formulations of the shallow water equations (e.g. Rivest et al. (1994)) have the forcing in the $u$ equation, which is a result of formulation the equations with the total fluid depth ($\Phi/g = H + \eta - h_0$) as the height variable. The former convention is used here, and for the same reasons as Lindberg and Alexeev give, namely that this best represents the physics of the forcing. In physical terms, the orography forces the free surface of the shallow water model by raising up parcels of fluid incident upon it. The rate of ascent is given by the product of the wind-speed and the orographic gradient, which is exactly the form of the forcing term in this case.

In order to linearize the shallow water equations, and restrict motion to one dimension, the variables are decomposed into a steady part and a part that varies only in $x$, so that

\begin{align}
\bar{u} + u'(x,t); \quad v' = v'(x,t); \quad \bar{\eta} + \eta'(x,t).
\end{align}
The fluid depth $\Phi/g$ has already been decomposed into steady and varying parts $(H - h_0$ and $\eta$) in the formulation of the equations of motion, but it is necessary to introduce a decomposition of $\eta$ to provide for a gradient in $y$ which is in geostrophic balance with $\mathbf{u}$. Note also that there is no steady flow in the $y$ direction. In order to prevent the advection of $\eta$ in the $y$-direction, a gradient in the $y$-direction of the orography is also introduced so that $h_0 = \overline{h}_0(y) + h_0'(x)$, with $\frac{\partial \eta}{\partial y} = \frac{\partial \overline{h}_0}{\partial y}$. Thus, there is no variation of the fluid depth in $y$.

In linearising the equations, it is taken that $\eta'$, $u'$, $v'$ and $h_0'$ are all small. Substituting (2.4) into (2.1–2.3), and discarding small quantities, gives the continuous, one-dimensional, linear shallow water equations. Dropping the primes for clarity, so that $\eta$, $u$, $v$ and $h_0$ are now the small perturbations, these are:

\begin{align}
\frac{\partial v}{\partial t} + \bar{u} \frac{\partial v}{\partial x} + f(u + \bar{u}) + g \frac{\partial \eta}{\partial y} &= 0, \\
\frac{\partial u}{\partial t} + \bar{u} \frac{\partial u}{\partial x} - f v + g \frac{\partial \eta}{\partial x} &= 0, \\
\frac{\partial \eta}{\partial t} + \bar{u} \frac{\partial \eta}{\partial x} + H \frac{\partial u}{\partial x} &= \bar{u} \frac{\partial h_0}{\partial x}.
\end{align}

In implementing these equations numerically, the model domain is made periodic in $x$. 

---

**Figure 2.1:** A schematic of the model domain. $L$ is the length of the domain in the $x$-direction; other symbols are as detailed in the text.
2.2.2 The model schemes

The numerical model is equipped with two integration schemes, both of which are semi-implicit. One scheme uses Eulerian advection, the other, semi-Lagrangian. Both schemes are discretized on a one-dimensional C-grid (with the $u$ data points displaced from the other variables by half a grid length; this was first proposed by Arakawa and Lamb (1977)). The C-grid is widely used in operational models because it results in less severe dispersion errors; see Durran (1999) for more details. In the case of both these schemes, the following discretization is used:

\begin{align}
  x &= j \Delta x, \\
  t &= n \Delta t,
\end{align}

so that, for instance, $u(x, t)$ in the continuous equations becomes $u(j \Delta x, n \Delta t)$ when discretized, and is notated as $u^n_j$.

2.2.2.1 Eulerian scheme

The discretized equations for this three-time-level scheme are as follows:

\begin{align}
  \frac{v_{j+1}^n - v_{j-1}^n}{2\Delta t} + \frac{u^n}{\Delta x} \bigg|_{j+\frac{1}{2}} + \frac{f}{2} (u_{j-\frac{1}{2}}^n + u_{j+\frac{1}{2}}^n) &= 0 \tag{2.10} \\
  \frac{u_{j+\frac{1}{2}}^{n+1} - u_{j-\frac{1}{2}}^{n-1}}{2\Delta t} + \frac{u^n}{\Delta x} \bigg|_{j+\frac{1}{2}} - \frac{f}{2} (v_{j+1}^n + v_{j+1}^n) + \frac{g}{2} \left[ \frac{\partial \eta}{\partial x} \bigg|_{j+\frac{1}{2}} + \frac{\partial \eta}{\partial x} \bigg|_{j+\frac{1}{2}} \right] &= 0 \tag{2.11} \\
  \frac{\eta_{j+1}^{n+1} - \eta_{j}^{n-1}}{2\Delta t} + \frac{\partial \eta}{\partial x} \bigg|_{j} + \frac{H}{2} \left[ \frac{\partial u}{\partial x} \bigg|_{j+1} + \frac{\partial u}{\partial x} \bigg|_{j} \right] &= \mathcal{H}_j \tag{2.12}
\end{align}

The spatial derivatives are calculated using a centred difference scheme, using the nearest grid-points to the point of evaluation. Examples are given in equations (2.13) and (2.14).

\begin{align}
  \frac{\partial v^n}{\partial x} \bigg|_{j+\frac{1}{2}} &= \frac{v_{j+1}^n - v_{j-1}^n}{2\Delta x} \tag{2.13} \\
  \frac{\partial u^{n+1}}{\partial x} \bigg|_{j} &= \frac{u_{j+\frac{1}{2}}^{n+1} - u_{j-\frac{1}{2}}^{n+1}}{\Delta x} \tag{2.14}
\end{align}

The term denoted $\mathcal{H}_j$ is the orographic forcing term. In this scheme, the term is given by

\begin{align}
  \mathcal{H}_j = \frac{h_{0,j+1} - h_{0,j-1}}{2\Delta x} \tag{2.15}
\end{align}
Since the treatment of the gravity wave terms is implicit, the equations are solved using standard Gaussian elimination (see Durran (1999), pp. 441–442).

### 2.2.2 Semi-Lagrangian scheme

The essence of the semi-Lagrangian method is the treatment of the advective derivative as a single term. This is in contrast with the Eulerian scheme described above, where the advective derivative is split into separate time and space derivatives. The advective derivative is the time derivative of a quantity which takes into account the motion of the fluid. Thus, to calculate it, it is necessary to know the values of the variable in question at the two ends of the trajectory rather than at specific grid points.

In the semi-Lagrangian method, the arrival point of a fluid parcel is always chosen to be a grid-point, but the departure point still needs to be calculated, and the value of the variable being differentiated needs to be determined for that point. Determining the departure point is straightforward for a linear model, and may be accomplished approximately for a non-linear one (see Durran (1999), pp. 310–313). The value of the variable at the departure point is then interpolated from the grid-point values. An example of the discretization of an advective derivative is given in equation (2.16).

\[
\frac{D\phi}{Dt}_{j} = \frac{\phi_{j}^{n+1} - \phi_{(d),j}^{n}}{\Delta t}
\]  

(2.16)

Here, and in the other equations, the subscript \((d), j\) indicates the point of departure at time-level \(n\) for a parcel which arrives at point \(j\) at time-level \(n + 1\). Thus, the discretized equations for this two time-level scheme are formulated as follows:

\[
\frac{v_{j}^{n+1} - v_{(d),j}^{n}}{\Delta t} + f u_{(m),j}^{n+\frac{1}{2}} = 0
\]  

(2.17)

\[
\frac{u_{j+\frac{1}{2}}^{n+1} - u_{(d),j+\frac{1}{2}}^{n}}{\Delta t} - f v_{(m),j+\frac{1}{2}}^{n+\frac{1}{2}} + \frac{g}{2} \left[ \frac{\partial \eta^{n+1}}{\partial x}_{j+\frac{1}{2}} + \frac{\partial \eta^{n}}{\partial x}_{(d),j+\frac{1}{2}} \right] = 0
\]  

(2.18)

\[
\frac{\eta_{j}^{n+1} - \eta_{(d),j}^{n}}{\Delta t} + \frac{H}{2} \left[ \frac{\partial u_{j}^{n+1}}{\partial x} + \frac{\partial u_{(d),j}^{n}}{\partial x} \right] = \mathcal{H}_{j}
\]  

(2.19)
Chapter 2  Experiments with a one-dimensional shallow water model

Figure 2.2: Schematic representation of the semi-Lagrangian concept. The figure shows a space-time diagram with data-points on three time-levels indicated by circles. The \( j \) and \( j + \frac{1}{2} \) grids are indicated with black and white circles, respectively. \( u \) is discretized onto the latter set of points, while \( \eta \) and \( v \) are present on the former. Two fluid parcel trajectories are shown, advecting the quantity \( \phi \). One trajectory ends at point \( j \) at time level \( n + 1 \), and one ends at point \( j + \frac{1}{2} \) at the same time. The yellow circles indicate points on the trajectories (and their respective notation) used in a 2 time-level semi-Lagrangian scheme such as the one described in the text. Note that, in a linear model, the trajectories are straight lines.

Note that the Coriolis terms are evaluated at the mid-points of the trajectories, denoted with a subscript \((m)\). To calculate these values, first the fields at time-levels \( n \) and \( n - 1 \) are extrapolated forward in time to level \( n + \frac{1}{2} \), using:

\[
\phi_{j}^{n+\frac{1}{2}} = \frac{3}{2} \phi_{j}^{n} - \frac{1}{2} \phi_{j}^{n-1}
\]  

Then, the value of fields at the mid-point of the trajectory is calculated by interpolation of the grid-point values at time-level \( n + \frac{1}{2} \). As Temperton and Staniforth (1987) point out, this extrapolation is weakly unstable; however, the damping which arises from the interpolation detailed below is probably sufficient to counter it, and, in any case, no instability was observed when the model was run.

The locations of departure points and mid-points of trajectories are illustrated in figure 2.2. The interpolation used in this scheme is cubic Lagrange, so that
\[
\phi_{(d),j}^n = -\frac{\alpha(1-\alpha^2)}{6}\phi_{j-p-2}^n + \frac{\alpha(1+\alpha)(2-\alpha)}{2}\phi_{j-p}^n - \frac{(1-\alpha^2)(2-\alpha)}{2}\phi_{j-p}^n - \frac{\alpha(1-\alpha)(2-\alpha)}{6}\phi_{j-p+1}^n, \quad (2.21)
\]

where \( p \) and \( \alpha \) are, respectively, the integer and fractional parts of \( \pi \Delta t/\Delta x \). The spatial derivatives on the grid-points are evaluated in the same way as in the Eulerian scheme above; in calculating the derivatives at the departure points of the trajectory, the fields are differentiated first, and then interpolated to the departure points.

The orographic forcing term, \( \mathcal{H}_j \) is evaluated using a spatially averaged Eulerian scheme, as described by Ritchie and Tanguay (1996), so that

\[
\mathcal{H}_j = \frac{\pi}{2} \left[ \frac{d\theta_0}{dx} \right]_j + \frac{d\theta_0}{dx} \right|_{(d),j}. \quad (2.22)
\]

### 2.3 Steady state accuracy

In a linear model with constant \( \overline{\eta} \), the flow may be decomposed into two parts: the steady wave field, and freely-propagating waves. Of these two, only the steady waves are affected by the orography. This was the obvious place to start the investigation.

#### 2.3.1 Steady state for the continuous equations

The steady state of the linear continuous equations may be found by treating each perturbation Fourier component separately, so that

\[
u = \hat{v}e^{ikx} \quad \eta = \hat{\eta}e^{ikx} \quad h_0 = \hat{h}_0e^{ikx} \quad (2.23)
\]

here \( \hat{u} \), \( \hat{v} \), \( \hat{\eta} \) and \( \hat{h}_0 \) are perturbation amplitudes, and may be complex. \( k \) is the wavenumber of the Fourier component, such that \( k = 2\pi/\lambda \). Additionally, a non-dimensional (integer) wavenumber \( \mu \) may be defined, being \( L/\lambda \), the number of waves which fit into the domain. Thus, \( k = 2\pi \mu/L \). The expressions for the steady states may be determined by substituting equations (2.23) into equations (2.5 – 2.7).
Table 2.1: Model parameters for the evaluation of steady-state accuracy.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>5500 m</td>
</tr>
<tr>
<td>$\bar{u}$</td>
<td>25 ms$^{-1}$</td>
</tr>
<tr>
<td>$f$</td>
<td>$10^{-4}$ s$^{-1}$</td>
</tr>
<tr>
<td>$g$</td>
<td>9.81 ms$^{-2}$</td>
</tr>
<tr>
<td>$L$</td>
<td>$4 \times 10^6$ m</td>
</tr>
</tbody>
</table>

They are as follows:

\[
\hat{u} = \left( \frac{\bar{u}gk^2}{k^2(gH - \bar{u}^2) + f^2} \right) \hat{h}_0, \quad (2.24)
\]

\[
\hat{v} = i\left( \frac{fgk}{k^2(gH - \bar{u}^2) + f^2} \right) \hat{h}_0, \quad (2.25)
\]

\[
\hat{\eta} = \left( \frac{f^2 - \bar{u}^2k^2}{k^2(gH - \bar{u}^2) + f^2} \right) \hat{h}_0. \quad (2.26)
\]

Note that the expressions for $\hat{u}$ and $\hat{\eta}$ are both real, whereas the expression for $\hat{v}$ is imaginary. This response is illustrated in figures 2.3 and 2.4, calculated for typical ‘atmosphere-like’ values of the parameters, given in table 2.1. Note that figure 2.3 shows the modulus of the complex response; the phase response is shown in figure 2.4. The modulus graph shows a minimum in $|\hat{\eta}/\hat{h}_0|$ response at wavenumber 3. Comparison with figure 2.4 shows that $\hat{\eta}/\hat{h}_0$ changes phase from 0 to $\pi$ at this point, which means that $\hat{\eta}/\hat{h}_0$ becomes negative there. In a non-periodic domain, the $\eta$ response would go through zero between wavenumbers 2 and 3; however, in a periodic domain only integer wavenumbers are possible, so the zero only appears if it happens to fall on an integer wavenumber. From equations (2.24 – 2.26) it can be shown that the zero in $\hat{\eta}$ corresponds to $\hat{u} = \hat{v}$, and occurs when $\mu = fL/2\pi\bar{u} = 1/2\pi Ro$, where $Ro$ is the Rossby number (see chapter 5 for a discussion of the Rossby number).

### 2.3.2 Eulerian, linear model

The steady state of the Eulerian linear scheme may also be found analytically, treating each Fourier component separately, as follows:

\[
u = \hat{v}e^{ijk\Delta x} \quad \eta = \hat{\eta}e^{ijk\Delta x} \quad h_0 = \hat{h}_0e^{ijk\Delta x}. \quad (2.27)\]
**Figure 2.3:** The steady state response magnitude of the continuous equations for typical model parameters, up to $\mu = 64$. The solid line is $|\tilde{u}/\tilde{h}_0|$ (units: $s^{-1}$); the dotted line is $|\tilde{v}/\tilde{h}_0|$ (units: $s^{-1}$); the dashed line is $|\tilde{\eta}/\tilde{h}_0|$ (dimensionless).

**Figure 2.4:** The steady state phase response of the continuous equations. Legend and other details as figure 2.3.
This yields the following expressions for the steady state magnitudes of individual Fourier components:

\[
\hat{u} = \left( \frac{4g\pi \sin^2 \theta \cos \theta}{4 \sin^2 \theta (gH - \pi^2 \cos^2 \theta) + f^2 \Delta x^2 \cos^2 \theta} \right) \hat{h}_0, \quad (2.28)
\]

\[
\hat{v} = i \left( \frac{2g\Delta x \sin \theta \cos \theta}{4 \sin^2 \theta (gH - \pi^2 \cos^2 \theta) + f^2 \Delta x^2 \cos^2 \theta} \right) \hat{h}_0, \quad (2.29)
\]

\[
\hat{\eta} = \left( \frac{f^2 \Delta x^2 \cos^2 \theta - 4\pi^2 \cos^2 \theta \sin^2 \theta}{4 \sin^2 \theta (gH - \pi^2 \cos^2 \theta) + f^2 \Delta x^2 \cos^2 \theta} \right) \hat{h}_0, \quad (2.30)
\]

with \( \theta = k \Delta x / 2 \). By making the approximation that \( \theta \) is small (i.e. the waves are well-resolved), and so \( \sin \theta \sim \theta \), and \( \cos \theta \sim 1 \), these equations may be shown to reduce to those for the continuous equations (2.24 – 2.26).

This steady state response is illustrated in figure 2.5, for the same set of parameters as figure 2.3, and for a domain for 128 points (so that \( \Delta x = 31.25 \) km, and waves up to \( \mu = 64 \) are resolved). Comparison between the two figures shows that the response is underestimated at high wavenumbers in the Eulerian scheme, going to zero at \( \lambda = 2 \Delta x \). This evidence suggests that to improve the representation of the steady state we should amplify the small scale orography, which contradicts the expectation that the accuracy of the flow in the model would be improved by orographic smoothing. Note that these expressions do not depend on the size of the time-step, so that for this scheme the accuracy of the steady state only depends on the resolution of the model. The phase response of the Eulerian scheme is shown in figure 2.6, and, except at \( \mu = 64 \), is the same as for the continuous equations.

### 2.3.3 Semi-Lagrangian, linear model

By applying the same principles as in the previous two sections, it is possible to find expressions for the magnitude of individual Fourier components in the steady state of the semi-Lagrangian scheme. The main expressions are given in equations (2.31 – 2.33); in this case the expressions for \( \hat{v} \) and \( \hat{\eta} \) are given in terms of that for \( \hat{u} \), for
Figure 2.5: The steady state response magnitude of the Eulerian scheme for typical model parameters, which are given in the text. The legend is the same as for figure 2.3.

Figure 2.6: The steady state phase response of the Eulerian scheme for typical model parameters, which are given in the text. The legend is the same as for figure 2.3.
In these equations, $\Psi_1 \ldots \Psi_5$ are defined by equations (2.34 – 2.38).

The parameters $\Gamma_p$, $\Gamma_q$ and $\Gamma_r$ have the same form, but depend on the combination of the grid the equation is discretized on, and the grid the quantity in question is being evaluated on. In the case of the equations discretized on a C-grid, there are four different possible combinations:

1. A variable discretized on grid-points $\{j\}$, in an equation discretized on grid-points $\{j\}$,

2. A variable discretized on grid-points $\{j + \frac{1}{2}\}$, in an equation discretized on grid-points $\{j + \frac{1}{2}\}$,

3. A variable discretized on grid-points $\{j\}$, in an equation discretized on grid-points $\{j + \frac{1}{2}\}$,
4. A variable discretized on grid-points \( \{j + \frac{1}{2}\} \), in an equation discretized on grid-points \( \{j\} \).

These combinations only generate three different grid distances for interpolation: one where the variable and the equation are on the same set of grid points, and two where they are not. Thus, for any particular trajectory, there are three different values of \( \Gamma \), labelled \( p \), \( q \) and \( r \). The expressions for these are given in (2.39), with the \( * \) replaced by \( p \), \( q \) or \( r \) as appropriate. The values of \( A \), \( B \), \( C \) and \( D \) are likewise defined in equations (2.40 – 2.43).

\[
\begin{align*}
\Gamma_* &= e^{-i\theta k \Delta x} \left[ -A_* e^{-i2\alpha \theta k \Delta x} + B_* e^{-i\theta k \Delta x} + C_* - D_* e^{i\theta k \Delta x} \right] \\
A_* &= \frac{\alpha_s(1 - \alpha_s^2)}{6} \\
B_* &= \frac{\alpha_s(1 + \alpha_s)(2 - \alpha_s)}{2} \\
C_* &= \frac{(1 - \alpha_s^2)(2 - \alpha_s)}{2} \\
D_* &= \frac{\alpha_s(1 - \alpha_s)(2 - \alpha_s)}{6}
\end{align*}
\]

In all these equations, \( p \) and \( \alpha_p \) are the integer and fractional parts respectively of \( \pi \Delta t / \Delta x \), \( q \) and \( \alpha_q \) are the integer and fractional parts of \( \pi \Delta t / \Delta x + \frac{1}{2} \), and \( r \) and \( \alpha_r \) are the integer and fractional parts of \( \pi \Delta t / \Delta x - \frac{1}{2} \). Finally, \( \theta = k \Delta x / 2 \), as before.

Unlike the Eulerian case, the semi-Lagrangian scheme steady state depends on the size of the time-step. The time-step, grid-spacing and advection speed can be combined into a non-dimensional quantity called the Courant number, as follows:

\[ C = \frac{\pi \Delta t}{\Delta x}. \]  

The Courant number gives the length of a fluid parcel trajectory as a fraction of the grid length. It is well known from von Neumann stability analysis that Eulerian advection schemes are unstable if the Courant number is greater than a certain value (often 1). This is the Courant-Friedrichs-Lewy stability criterion, alluded to in chapter 1. One of the advantages of the semi-Lagrangian method is that advection remains stable (though not necessarily very accurate) for all time-step lengths. The steady-state response (magnitude and phase) of the semi-Lagrangian
model, calculated for the same model parameters as for the Eulerian model above, and for various Courant numbers, is shown in figures 2.7 – 2.12.

At low Courant numbers (figures 2.7 and 2.8), it can be seen that the response is very similar to the Eulerian scheme, although there is some phase-shift observed at high wave numbers in \( \hat{v} \) and \( \hat{\eta} \). However, as \( C \) becomes larger, more significant differences in behaviour emerge. At \( C = 1 \) (figures 2.9 and 2.10), a noticeable amplification of \( \hat{\eta} \) has developed, relative to the solution for the continuous equations. However, the reduction in amplitude for the other two variables is more significant than at low Courant number. All three variables show an increased phase inaccuracy.

The spuriously amplified response illustrated in figure 2.9 is *Spurious Orographic Resonance*, as described in chapter 1. This effect manifests itself when \( C \gtrsim 1 \). The response at \( C = 2 \) (figures 2.11 and 2.12) shows typical behaviour – two resonance points (at \( \mu \sim 25 \) and \( \mu \sim 35 \)), and a zero response between them. This agrees with the analysis given in Rivest et al. (1994).

The spurious resonances, however, only feature in the \( \hat{\eta} \) response. In the other two variables, the response is always underestimated at small scales in the semi-Lagrangian scheme. So, if the spurious resonances can adequately be dealt with by off-centring, then the results lead to the same conclusion as for the Eulerian scheme, namely that the orography at small scales should be amplified rather than reduced if the accuracy of the steady state is to be improved. Note that this conclusion is only valid for finite-difference scheme; the behaviour of spectral transform models that use semi-Lagrangian advection would be different.

### 2.3.4 Conclusions

Both of the schemes examined above suggest that the amplification, rather than the smoothing, of small scale orography is what is needed to improve the accuracy of the represented flow in finite-difference models. However, this runs contrary to the findings of the studies identified in chapter 1. The reason for this may be that the models used in this chapter are not sufficiently representative of NWP and climate models to be able to reproduce the observed problems with near-grid-scale orography. It may also be that any such effects that might be seen in these simple
Figure 2.7: The steady state response magnitude of the semi-Lagrangian scheme for typical model parameters. In this figure, $\Delta t = 30\,\text{s}$, and $C = 0.024$. The legend is the same as for figure 2.3.

Figure 2.8: The steady state phase response of the semi-Lagrangian scheme for typical model parameters. Other details as figure 2.7.
Figure 2.9: The steady state response magnitude of the semi-Lagrangian scheme for typical model parameters. In this figure, $\Delta t = 1250$ s, and $C = 1.0$. The legend is the same as for figure 2.3.

Figure 2.10: The steady state phase response of the semi-Lagrangian scheme for typical model parameters. Other details as figure 2.9.
Chapter 2  Experiments with a one-dimensional shallow water model  25

Figure 2.11: The steady state response magnitude of the semi-Lagrangian scheme for typical model parameters. In this figure, $\Delta t = 2500\text{ s}$, and $C = 2.0$. The legend is the same as for figure 2.3.

Figure 2.12: The steady state phase response of the semi-Lagrangian scheme for typical model parameters. Other details as figure 2.11.
models are not connected with the representation of the steady part of the motion. With this in mind, it was decided to extend the work to cover transient free waves in the model, generated as a result of changing $\mathbf{u}$ over time. If evidence to support the use of orographic smoothing is still not found, it leads to the conclusion that the model is not sufficiently realistic, a possibility that is discussed at the end of the chapter.

### 2.4 Transients and their accuracy

#### 2.4.1 Introduction

The steady problem described above does not provide any evidence for the use of smoothed orography in NWP models. However, although the steady waves are the only part of the flow which is directly affected by the orography when $\mathbf{u}$ is held constant, this is not true if $\mathbf{u}$ is varied. In this situation, the partition between steady and freely-propagating waves is varied as $\mathbf{u}$ changes, and so the accuracy of the solution depends on both the accuracy of the steady waves, and on the numerical dispersion. With this in mind, it was decided to investigate the accuracy of the evolution of a typical flow, varying $\mathbf{u}$ for part of the run. The aim is to investigate the accuracy of the response to near-grid-scale orography by comparing runs done with well-resolved and just-resolved orography.

#### 2.4.2 Experimental design

The experiment uses the Eulerian scheme, with realistic atmospheric parameters and real orography, in order to render it as useful as possible. The parameters used are almost the same as those in the steady field investigation, but the domain length is altered to the value appropriate to 50°N, which is the latitude of the orography profile chosen for the experiment. The full list of parameters is given in table 2.2. The model is run at two resolutions: one run at high resolution (2048 points, $\Delta x = 12.6$ km), with well-resolved orography, and two runs at low resolution.
### Table 2.2: Parameters used in the study of transient effects – those common to all model runs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>5500 m</td>
</tr>
<tr>
<td>$f$</td>
<td>$10^{-4}$ s$^{-1}$</td>
</tr>
<tr>
<td>$g$</td>
<td>9.81 ms$^{-2}$</td>
</tr>
<tr>
<td>$L$</td>
<td>$2.573 \times 10^7$ m</td>
</tr>
</tbody>
</table>

### Table 2.3: Parameters used in the study of transient effects – those specific to the different model runs.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>run A</th>
<th>run B</th>
<th>run C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$ (points)</td>
<td>2048</td>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td>$\Delta x$ (km)</td>
<td>12.56</td>
<td>201.02</td>
<td>201.02</td>
</tr>
<tr>
<td>$\Delta t$ (s)</td>
<td>100</td>
<td>100</td>
<td>1600</td>
</tr>
<tr>
<td>$C_{\text{min}}$</td>
<td>0.199</td>
<td>0.0124</td>
<td>0.199</td>
</tr>
<tr>
<td>$C_{\text{max}}$</td>
<td>0.398</td>
<td>0.0249</td>
<td>0.398</td>
</tr>
</tbody>
</table>

(128 points, $\Delta x = 201.0$ km), whose orography has components at and near the grid scale. The reason for running the low resolution model twice is explained below.

Each model run is of the same form, and for a total of 5 days each. Initially, $\overline{u} = 25$ ms$^{-1}$; the model is initialized in a steady state, and allowed to run for 1 day. Then, the value of $\overline{u}$ is varied smoothly from 25 ms$^{-1}$ to 50 ms$^{-1}$ over the course of a further day. The model is then allowed to run on for 3 more days at $\overline{u} = 50$ ms$^{-1}$.

The high resolution ‘control’ run has a time-step of 100 s, giving a Courant number of 0.199 at $\overline{u} = 25$ ms$^{-1}$, and one of 0.398 at $\overline{u} = 50$ ms$^{-1}$. The two low resolution runs differ in the time-step used; one has the same time-step as the high resolution run, while the other has the same Courant number (meaning $\Delta t = 1600$ s). The parameters for each run are summarized in table 2.3. Scalar diagnostics (drag, total energy, RMS divergence tendency) are output at every time-step. Field diagnostics ($u$, $v$, $\eta$, spectra of these fields and of energy and drag components) are output every 4 hours.
2.4.3 Orography

It was necessary to construct appropriate orographic fields for the experiment, one well-resolved at high resolution, and one at low resolution. The source of the orography data was the GLOBE\textsuperscript{1} data-set (Hastings et al. (1999)), which has 32768 points in a latitude circle, and therefore has a grid spacing of 785m at 50°N. The procedure for constructing the experimental orography was as follows:

1. Reduce the GLOBE orography to 128-point resolution by averaging over the points in each grid-length;
2. Interpolate the outcome of step 1 using cubic splines, to create a well-resolved orography at 2048-point resolution;
3. Generate a mean orography from step 2 at 128-point resolution.

It is necessary to perform step 3 because the output of step 1 is not the mean orography of the well-resolved 2048-point orography. By re-averaging this orography, the appropriate just-resolved counterpart to the high-resolution data is generated.

The raw GLOBE data, 128-point version, and the 2048-point version are shown in figures 2.13 – 2.15, respectively.

2.4.4 Results

In order to try and characterize some of the effects of the reduction in resolution, comparison was made between the way domain-integrated drag evolved in the different runs. The drag $D$ is defined by equation (2.45), and the appropriate discrete form is given in equation (2.46).

\begin{equation}
D = g \int_{0}^{L} h_0 \frac{\partial \eta}{\partial x} \, dx
\end{equation}

\begin{equation}
D = \frac{g}{2} \sum_{j=1}^{N} h_{0,j}(\eta_{j+1} - \eta_{j-1})
\end{equation}

\textsuperscript{1}Global Land One-kilometre Base Elevation
Figure 2.13: The GLOBE data-set at 50° N, with 32768 points in the domain.

Figure 2.14: The 128-point mean orography.
The evolution of drag for runs A and B is plotted in figure 2.16, while the drag for run C is compared with that for run A in figure 2.17. In the first part of the experiment \((0.0 \text{ s} \leq t < 8.64 \times 10^4 \text{ s})\), when \(\bar{u}\) is constant at \(25 \text{ ms}^{-1}\), all three runs exhibit zero drag. That this is the correct result for a steady state may be confirmed by combining equations (2.5 - 2.7), and substituting into (2.45).

During the second part of the experiment \((8.64 \times 10^4 \text{ s} \leq t < 1.728 \times 10^5 \text{ s})\), when \(\bar{u}\) is increasing from \(25 \text{ ms}^{-1}\) to \(50 \text{ ms}^{-1}\), the drag reaches a maximally negative value half-way through, before returning to zero at the end of the section. During this part, runs B and C underestimate the magnitude of the drag by approximately the same amount compared to run A.

In the third part of the experiment \((t \geq 1.728 \times 10^5 \text{ s})\), when \(\bar{u}\) is held constant again, the drag is non-zero, as the flow is now unsteady. The drag varies between positive and negative, and the magnitude is bounded by the magnitude of the free wave component of the model state. However, over the long term (in a run of 100 days), the drag integrated to zero. During this part, the evolution of the drag in run B is almost identical to that in run A. However, run C, although it is qualitatively similar to run A, does not generally have the same magnitude.

It seems likely that the discrepancy between the three runs in the third part of
Figure 2.16: Evolution of drag for experimental runs A (solid line) and B (dotted line). The 3 sections of the graph (divided by dashed vertical lines) correspond to when $\overline{u} = 25 \text{ ms}^{-1}$, $\overline{u}$ increasing, and $\overline{u} = 50 \text{ ms}^{-1}$ respectively.

Figure 2.17: Evolution of drag for experimental runs A (solid line) and C (dotted line). Other details as figure 2.16.
Figure 2.18: Spectral decomposition of drag for runs A (solid line) and B (dotted line). The magnitude of the drag is shown on a log scale for clarity. Note that the largest resolved wavenumber in run B is 64, with the run A spectrum dropping off to negligibly small values beyond this point.

The experiment is due to the numerical dispersion inherent in the discretization, with the consequence that the shorter the time-step, the more accurate the free wave propagation speeds. However, the behaviour of the drag in part two, as $\mu$ is increased, is harder to explain. With this in mind, the spectral composition of the drag was calculated.

Decomposing the flow into Fourier components, as in section 2.3.2, yields the following expression for the drag associated with wavenumber $k$ in the linear model:

$$\hat{D}(k) = -\frac{gh \sin k \Delta x}{2} (\text{Re}\{\hat{h}_0(k)\} \text{Im}\{\hat{\eta}(k)\} - \text{Re}\{\hat{\eta}(k)\} \text{Im}\{\hat{h}_0(k)\}).$$  (2.47)

Figure 2.18 shows the spectra of drag for runs A and B at $t = 1.296 \times 10^5$ s, approximately when $|D|$ attains its maximum value. It is clear from the plot that the discrepancy between runs A and B is at its greatest at the largest resolved wavenumbers. It was hypothesized that this underestimation of the drag at high wavenumbers might be linked to the misrepresentation of the steady flow at these wavenumbers. In order to test this, the ratio of the spectra of drag for runs A and B ($D_B/D_A$) was compared to the ratio of the spectra of steady $\eta$ response for the two
model configurations ($\hat{\eta}_B/\hat{\eta}_A$). These two ratios are shown in figure 2.19. Comparing run C with run A yields very similar results. Apart from a slight inconsistency at $\mu \sim 16$, which is a numerical artifact resulting from the fact that the steady flow response goes to zero around that point, the two ratios are in good agreement for the whole spectral range. Thus, the underestimation of the drag in the second part of the experiment seems to be entirely attributable to the underestimation of the steady flow response.

However, it is not clear which scales are most significant in the underestimation of drag, as the drag near the grid scale is approximately two orders of magnitude less than at the largest scales. To help understand this, the cumulative drag, $D_C$ was calculated:

$$D_C(k) = \int_1^k D(k) \, dk. \tag{2.48}$$

Figure 2.20 shows the difference between the cumulative drag for runs A and B, at $t = 1.296 \times 10^5$ s. The section with the steepest slope corresponds with the wavenumbers where the underestimation of drag is most significant. Between $\mu \sim 20$ and $\mu \sim 35$ the slope is noticeably steeper; this corresponds to the range $\lambda \sim 4\Delta x$ to $\lambda \sim 6\Delta x$, which is consistent with the results obtained by Davies and Brown.
This behaviour is in marked contrast to the evolution of the flow after $\tau$ has become constant again, in part three of the experiment. Here, the accuracy of the evolution of the drag depends only on the time-step. In run B, the evolution of the drag is almost identical to that in run A; the time-steps of these two runs are the same. However, in run C, when the Courant number is the same as in run A (rather than $\Delta t$ being fixed), the drag evolution is far less accurate, although the qualitative behaviour is captured.

Comparison between the model fields in different runs at various points in the experiment shows that even the discrepancy in the drag evolution in run C does not correspond to a significant error in the fields themselves. The differences between the fields in run A at time-steps 0 and 3600 (i.e. $t = 3.6 \times 10^5$ s) is shown in figure 2.21. This figure gives a snap-shot of the transient waves generated by the change in $\tau$, in the well-resolved control. The difference between the two lower resolution runs (B and C) at $t = 3.6 \times 10^5$ s is shown in figure 2.22. It is clear by comparing these two figures that the difference between runs B and C at $t = 3.6 \times 10^5$ s is at least an order of magnitude smaller than the transient waves being studied. This observation, along with the dependence on $\Delta t$ leads to the conclusion that this is not a
Figure 2.21: The difference between the model fields at $t = 0$ and $t = 3.6 \times 10^5$ s for run A. The solid line is $\eta$, the dotted line $u$, the dot-dashed line $v$.

Figure 2.22: The difference between the model fields at $t = 3.6 \times 10^5$ s for runs B and C. Legend as figure 2.21.
good simulation of the problems which are caused by near grid scale orography in NWP and climate models.

2.5 Summary and conclusions

The conclusions to be drawn from this work may be summarized as follows:

- In general, the steady component of the flow near the grid scale is underestimated in the linear models investigated in this study. In the semi-Lagrangian scheme, there are problems with Spurious Orographic Resonance; in the context of this simple model, these may be best alleviated by the method proposed in Rivest et al. (1994), although experiments would have to be done to check this. Thus, the implication of the steady flow results is that, if anything, the near grid scale component of the orography should be amplified rather than filtered out. This is clearly in conflict with the results of previous studies, and the reasons why filtering should be beneficial, both of which were explained in chapter 1.

- The evolution of the drag when $\pi$ is varied exhibits inaccuracies due to the misrepresentation of the effects of near grid scale orography. The major loss of drag occurs in the range $\lambda \sim 4\Delta x$ to $\lambda \sim 6\Delta x$, which is consistent with the results obtained by Davies and Brown (2001). This error is not corrected by using shorter time-steps, which confirms orographic representation as its source. However, spectral decomposition of the drag suggests that the underestimation of drag here is due to the underestimation of the near grid scale part of the steady flow. Thus, although supporting the findings of previous studies regarding the scales that are badly represented, these results nevertheless seem to suggest that amplifying rather than filtering the orography could be a solution.

- The accuracy of freely-propagating waves generated by varying $\pi$ depends only on the time-step, rather than on the scale of the orography relative to the grid scale. This means that the problem of orographic representation is not manifested in freely-propagating waves.
The overall conclusion of the work in this chapter suggests, perversely, that the optimal orographic ‘smoothing’ for this particular 1-D linear shallow water model is actually an *amplification* of the near-grid-scales. However, this is at odds with the evidence for the need for the removal of these scales in NWP and climate models. Taken together, these conclusions strongly show that the 1-D linear model studied here is not sufficiently similar to the models used in weather prediction and climate research to able to reproduce the problems seen in these more complicated models. The models used in this chapter are linear, and given the nature of the problems encountered with near-grid-scale orography in more complex models, it suggests that non-linear models might be more useful in the study of this problem.

In the following two chapters, the details of a new variational smoothing scheme are discussed. In chapter 5, this scheme is evaluated in a simple numerical model – the results obtained from this present chapter are used to inform the kind of model used.
Chapter 3

Two-dimensional orography – a variational approach to smoothing

3.1 Introduction

As outlined in chapter 1, there is evidence that smoothing orography in numerical models is beneficial. Previous investigations into the benefits of smoothing (e.g. Webster et al. (2003)) have concentrated on linear filtering procedures, which necessarily have two significant drawbacks. Firstly, the smoothing will cause the elevation of grid-points over the sea, and this mismatch between the orography and the land-sea mask can affect the operation of sub-grid parameterizations, since the physics of land-atmosphere and ocean-atmosphere interactions is different. Particularly, the elevated sea-points will act as spurious heat sources. Previous approaches to solving this problem have been to manually set the sea-points to zero (Webster et al. (2003)), or lowering the temperature of elevated sea-points according to some adiabatic lapse rate (Navarra et al. (1994)). Secondly, linear filtering will lower the height of the sharpest features in the orographic field, including dynamically-important barrier ranges such as the Andes. Thus, the removal of near-grid-scale information from the orographic field is likely to have an impact on the large-scale aspects of the flow.

There are parallels between these problems and that of Gibbs ripples in spectral model orography. In that case, the number of spectral coefficients used to represent the model fields is fewer than the number of points in the quadratic Gaussian grid
on which the non-linear terms in the model are evaluated. This mismatch of degrees of freedom means that when the orography is transformed onto the quadratic grid for the non-linear calculations, there are Gibbs oscillations in the areas which are supposed to be flat, such as the sea. One approach to controlling this, suggested by Bouteloup (1995), and extended by Holzer (1996), is to use a variational method; a cost is attached to the non-flatness of the sea, and so the calculation of the spectral coefficients of the orography is constrained accordingly. This method has been tried by ECMWF, but was abandoned.

The variational method described in this chapter allows the calculation of smoothed orography, while constraining the smoothing so that the sea remains flat, and dynamically important features are not lowered. The general variational filter includes a large class of linear filters as special cases. In addition, the method can be extended to include other constraints, if necessary. Most of the chapter is concerned with the general variational smoothing method and the class of linear filters it can emulate; the other constraints are explained briefly at the end of the chapter, before being more thoroughly explored in chapter 4.

### 3.2 Mathematical basis

#### 3.2.1 Preliminaries

In general terms, a variational method involves the minimization of a cost function, $J(h)$, with respect to $h$, which may be a function of one or more variables. The cost function $J(h)$ expresses the penalty attached to different properties of $h$, and so by writing $J$ in the right way, it is possible to generate $h$ so that it has the properties required.

The general cost function for the variational smoothing method has several parts:

$$J(h) = J_{\text{mean}}(h) + J_{\text{smooth}}(h) + J_{\text{sea}}(h) + J_{\text{peaks}}(h) + \ldots$$

The total cost $J(h)$ is the sum of terms penalising the difference of $h$ from the mean orography ($J_{\text{mean}}(h)$), the roughness of $h$ ($J_{\text{smooth}}(h)$), the non-flatness of the
Chapter 3 Two-dimensional orography – a variational approach to smoothing

sea \((J_{\text{sea}}(h))\), the reduction in barrier height \((J_{\text{peaks}}(h))\), and any other desired constraints. These terms in the cost function express \textit{weak constraints}, which constrain \(h\) (the smoothed orography) to differing degrees, depending on their relative magnitude. However, it is also possible to apply \textit{strong constraints}, so that the solution is \textit{required} to have a certain property. The easiest strong constraint to understand is perhaps the flat-sea requirement, which can be implemented by simply setting \(h = 0\) over the sea, and allowing the minimization algorithm to act only on the land points. This particular strong constraint is discussed later in the chapter.

### 3.2.2 A simple example in one dimension

A simple smoothing cost function in one dimension may be written as follows:

\[
J(h) = \int_0^L (h - \bar{h})^2 dx + \alpha \int_0^L \left( \frac{dh}{dx} \right)^2 dx.
\]

The two terms correspond respectively to \(J_{\text{mean}}\) and \(J_{\text{smooth}}\) in (3.1), with \(\bar{h}(x)\) being the unsmoothed mean orography, \(h(x)\) the target orography (the independent variable of the cost function), and \(\alpha\) a parameter which sets the relative importance of the two terms. The domain runs from \(x = 0\) to \(x = L\), and is periodic so that \(h(0) = h(L)\), etc.

Using integration by parts, (3.2) may be rewritten as:

\[
J(h) = \int_0^L (h - \bar{h})^2 dx - \alpha \int_0^L h \frac{d^2h}{dx^2} dx.
\]

We may minimize (3.3) by considering the functional derivative, \(\frac{\delta J}{\delta h}\), defined by

\[
\delta J = \int \frac{\delta J}{\delta h} \delta h \, dx.
\]

In practical terms, this can be evaluated by considering the change in \(J\) caused by a small increment in \(h\); that is, we evaluate \(J(h + \delta h) - J(h)\), and neglect terms in \(\delta h^2\):

\[
\delta J = J(h + \delta h) - J(h) = 2 \int_0^L (h - \bar{h}) \delta h \, dx - \alpha \int_0^L \delta h \frac{d^2h}{dx^2} dx - \alpha \int_0^L h \frac{d^2h}{dx^2} \delta h \, dx.
\]

The final term in this expression can be rewritten in the same form as the second term by integrating by parts twice. Thus, combining all three terms together, we obtain:

\[
\delta J = 2 \int_0^L (h - \bar{h} - \alpha \frac{d^2h}{dx^2}) \delta h \, dx.
\]
In equation (3.6), the bracketed part corresponds directly to $\delta J/\delta h$ in (3.4). When $J$ is minimized, $\delta J = 0$ for all values of $\delta h$; for this to be true, the part in brackets must equal zero:

$$h - \bar{h} - \alpha \frac{d^2 h}{dx^2} = 0.$$  

(3.7)

By considering only single Fourier components (so that $h(x) = \hat{h}(k)e^{ikx}$, etc.), we obtain the spectral response $F(k)$ of this cost function:

$$F(k) \equiv \frac{\hat{h}(k)}{\bar{h}(k)} = \frac{1}{1 + \alpha k^2}.$$  

(3.8)

As confirmation of the form of (3.6), equation (3.8) may be obtained by substituting single Fourier components into (3.3), and then differentiating with respect to $h$.

This is not a very scale-selective filter, as can be seen from figure 3.1, where it is compared with the sixth-order filter of Raymond (1988), used by Webster et al. (2003), and referred to hereafter as the Raymond filter. The spectral response of the Raymond filter in one dimension is given by

$$F(k) = \frac{1}{1 + \delta \tan^6\left(\frac{\delta k}{2}\right)},$$  

(3.9)

where $\delta$ is the grid-length, $k$ is the wavenumber, and $\delta$ is a variable parameter. By substituting $\lambda = \frac{2\pi}{k}$, we can obtain the expressions for $F(\lambda)$ plotted in figure 3.1.

### 3.2.2.1 A note about dimensions

Later, in chapter 4, a method for scaling the coefficients of the cost function (such as $\alpha$) to make them dimensionless is discussed. For the moment, though, the dimensions of the coefficients may be determined by the requirement that the spectral response $F(k)$ is dimensionless. Considering (3.8), and given that the number 1 is dimensionless, it is clear that the dimensions of $\alpha$ must be length squared (denoted $[L]^{2}$). This gives the cost function $J$ dimensions of $[L]^{3}$.

### 3.2.3 A more general linear filter

#### 3.2.3.1 Desired characteristics

It would be advantageous to be able to construct a more general variational filter, to enable the specification of a more scale-selective spectral response. A particularly
Chapter 3  Two-dimensional orography – a variational approach to smoothing

Figure 3.1: Comparison between spectral response functions $F(\lambda)$ of the simple variational filter (3.8) and the Raymond filter.

attractive generalization of the cost function (3.2) would be in the form of a Padé Approximant, which is a ratio of two polynomials. Press et al. (1992) define a Padé Approximant $R(x)$

$$R(x) = \frac{\sum_{k=0}^{M} a_k x^k}{1 + \sum_{k=1}^{N} b_k x^k}, \tag{3.10}$$

and note its usefulness in approximating a wide variety of functions over a greater numerical range than is practical using techniques such as power series expansion. Whereas other series expansion techniques only yield a useful approximation close to a particular value of $x$, Padé Approximants retain their validity over a much wider range – a necessary condition for the representation of orographic smoothing filters. Also, as will be shown below, a very scale-selective filter can be specified as a Padé Approximant with a small number of terms. Lastly, the spectral response of the simple example given above (equation (3.8)) is itself a very simple example of a Padé Approximant, and offers the hope that a generalised cost function whose spectral response is similar to (3.10) could be constructed by including terms in higher orders of $\nabla^2 h$ and $\nabla^2 \tilde{h}$. For a detailed discussion of Padé Approximants, see
Chapter 3 Two-dimensional orography – a variational approach to smoothing

Baker (1975).

The simple example above yields a spectral response function which contains a term in $k^2$, but none in $k$. Therefore, it seems sensible to seek a general cost function whose spectral response is a Padé Approximant in $k^2$ rather than $k$:

$$F(k) = \frac{\alpha_0 + \sum_{n=1}^{N} \alpha_n k^{2n}}{1 + \sum_{n=1}^{N} \beta_n k^{2n}}. \quad (3.11)$$

The first term of the numerator is written separately in (3.11) to emphasise the symmetry between the two polynomials which comprise the fraction.

3.2.3.2 Constructing the cost function

It is not obvious \textit{a priori} what the form of this cost function should be, but it may be determined by beginning from the desired spectral response. First consider a single spectral component of the orography, so that

$$h = \hat{h} e^{ikx} \quad (3.12)$$

$$\hat{h} = \hat{h} e^{ikx}, \quad (3.13)$$

where $\hat{h}$ and $\hat{h}$ are in general complex. We may rewrite (3.11) as

$$\hat{h} = \left[ \frac{\alpha_0 + \sum_{n=1}^{N} \alpha_n k^{2n}}{1 + \sum_{n=1}^{N} \beta_n k^{2n}} \right] \hat{h}, \quad (3.14)$$

which, rearranging and multiplying by $e^{ikx}$, gives:

$$\hat{h} e^{ikx} (1 + \sum_{n=1}^{N} \beta_n k^{2n}) = \hat{h} e^{ikx} (\alpha_0 + \sum_{n=1}^{N} \alpha_n k^{2n}). \quad (3.15)$$

The sum terms in (3.15) may be rewritten using

$$\frac{d^{2n}}{dx^{2n}} \left[ e^{ikx} \right] = (-1)^n k^{2n} e^{ikx}, \quad (3.16)$$
to give
\[
\hat{h} \left( e^{ikx} + \sum_{n=1}^{N} (-1)^n \beta_n \frac{d^{2n}}{dx^{2n}} \left[ e^{ikx} \right] \right) = \hat{h} \left( \alpha_0 e^{ikx} + \sum_{n=1}^{N} (-1)^n \alpha_n \frac{d^{2n}}{dx^{2n}} \left[ e^{ikx} \right] \right).
\]
(3.17)

Substituting (3.12) and (3.13), we get
\[
h + \sum_{n=1}^{N} (-1)^n \beta_n \frac{d^{2n} h}{dx^{2n}} = \alpha_0 h + \sum_{n=1}^{N} (-1)^n \alpha_n \frac{d^{2n} h}{dx^{2n}}.
\]
(3.18)

If we rearrange (3.18) so that all the terms are on the same side, equal to zero, we recover the expression for the gradient of our desired cost function:
\[
\delta J = h - \alpha_0 h + \sum_{n=1}^{N} (-1)^n \left[ \beta_n \frac{d^{2n} h}{dx^{2n}} - \alpha_n \frac{d^{2n} h}{dx^{2n}} \right] = 0.
\]
(3.19)

From the experience of calculating the spectral response of the simple cost function (3.2), we know that the more general cost function will include terms of the form \((Ah - B\bar{h})^2\), where \(A\) and \(B\) may be linear operators such as \(\frac{d^n}{dx^n}\). Supposing that a very general cost function were written in this form:
\[
J = \int_0^L (Ah - B\bar{h})^2 dx,
\]
(3.20)
then its differential with respect to \(h\), and according to (3.4), would be:
\[
\delta J = 2 \int_0^L A(Ah - B\bar{h}) \delta h \, dx.
\]
(3.21)

In the case that \(A\) and \(B\) are scalar constants, then:
\[
\delta J = 2(A h - AB) \delta h
\]
(3.22)
whereas if \(A\) and \(B\) are differentials, so that
\[
A = a \frac{d^n}{dx^n}, \quad B = b \frac{d^n}{dx^n}.
\]
(3.23)
then, by integrating by parts \(n\) times:
\[
\delta J = 2(-1)^n \left( a^2 \frac{d^{2n} h}{dx^{2n}} - ab \frac{d^{2n} h}{dx^{2n}} \right).
\]
(3.24)

So, by using (3.22) and (3.24), and noting that the cost function may be multiplied by a positive scalar without altering its effect, we may finally obtain the desired generalized cost function:
\[
J(h) = \int_0^L (h - \alpha_0 \bar{h})^2 dx + \sum_{n=1}^{N} \int_0^L \left[ \sqrt{\beta_n} \frac{d^n h}{dx^n} - \frac{\alpha_n d^n \bar{h}}{\sqrt{\beta_n} dx^n} \right]^2 dx.
\]
(3.25)
In practice, it is convenient to rewrite this by expanding the brackets, discarding the constant terms that depend only on $h$, and integrating by parts, so that:

$$J(h) = \int_0^L (h^2 - 2\alpha_0 h \overline{h}) \, dx + \sum_{n=1}^N (-1)^n \int_0^L \left[ \beta_n h \frac{d^{2n} h}{dx^{2n}} - 2\alpha_n \frac{d^{2n} \overline{h}}{dx^{2n}} \right] \, dx \quad (3.26)$$

### 3.2.3.3 Dimensional considerations

The determination of the dimensions of the various coefficients of the general cost function of (3.26) follows on from the procedure for the simple cost function, presented above. From (3.26), it is clear that $\alpha_0$ must be dimensionless. This means that the dimensions of $J$ are again $[L]^3$, which means that, for consistency, the dimensions of $\alpha_n$ and $\beta_n$ are both $[L]^{2n}$.

### 3.2.4 Discretization of the cost function

The spectral response given in (3.11) is correct for the continuous one-dimensional problem, but discretization will result in a different response. In the one-dimensional case, we may discretize on the set of points $\{x_j\}$, where $x_j = j \Delta x$. The important difference that this makes is in the evaluation of the differential terms in the cost function. Consider the second derivative of a single Fourier component of $h$, in a centred, second-order approximation, and with $h = \hat{h} e^{ijkx}$. In this case, we have

$$\left. \frac{d^2 h}{dx^2} \right|_j \approx \frac{h_{j+1} - 2h_j + h_{j-1}}{\Delta x^2}, \quad (3.27)$$

and thus if $h = \hat{h} e^{ijkx}$

$$\left. \frac{d^2 h}{dx^2} \right|_j \approx -\hat{h} \left[ 4 \sin^2 \left( \frac{k \Delta x}{2} \right) \right] e^{ijkx}. \quad (3.28)$$

For small $k \Delta x$, when $\sin \theta \approx \theta$, this becomes

$$\left. \frac{d^2 h}{dx^2} \right|_j \approx -\hat{h} k^2 e^{ijkx}, \quad (3.29)$$

the result for the continuous case. So in general, the discretized higher-order derivatives have the following response for a single Fourier component:

$$\left. \frac{d^{2n} h}{dx^{2n}} \right|_j \approx \hat{h} (-1)^n K^{2n} e^{ijkx}, \quad (3.30)$$
with \( K^2 \) given by
\[
K^2 = \frac{4 \sin^2 \left( \frac{k \Delta x}{2} \right)}{\Delta x^2} = \left[ \sin^2 \left( \frac{k \Delta x}{2} \right) \right] k^2. \tag{3.31}
\]
We can think of the second-order differential as a linear operator, in which case the equation (3.31) gives the eigenvalues of the discretized operator as a function of the eigenvalues of the continuous operator. The eigenfunctions (which are the Fourier components) remain unchanged by discretization.

This means that the spectral response of the general variational filter described by the cost function (3.26) becomes:
\[
F(K) = \frac{\alpha_0 + \sum_{n=1}^{N} \alpha_n K^{2n}}{1 + \sum_{n=1}^{N} \beta_n K^{2n}}. \tag{3.32}
\]

### 3.2.5 Determining the Padé coefficients

#### 3.2.5.1 Methods

The determination of the appropriate Padé coefficients for a given filter is far from straight-forward. Press et al. (1992) detail a method for finding Padé coefficients from the coefficients of a power series. However, for typical filters that we might want to emulate in this way, such as the Raymond filter, this is not a useful method. The power series is an expansion about \( k = 0 \), where the derivatives of the response function \( F = F(k) \) are generally all very small, and so the calculation of a useful power series is impractical.

Baker (1975) describes a method for fitting a Padé Approximant to a prescribed set of points. However, the method returns the same number of Padé coefficients as there are input points, which means that while the precision with which a Padé representation of an analytical function may be specified may be arbitrarily high, this leads to Padé Approximants with ever-increasing numbers of coefficients. For the purposes of this variational method it is desirable (for reasons of computational efficiency) to employ Padé Approximants with the smallest possible number of coefficients. However, unless the specified points are chosen with care, the resulting
Approximant may not bear much resemblance to the desired function. In view of the difficulty found in choosing the appropriate points for this fitting process, it was found to be as easy in the case of the Raymond filter to choose the Padé coefficients by trial and error. The use of an appropriately laid-out spreadsheet aided this considerably.

In addition, it may be possible with some filter response functions to use the properties of the functions themselves to obtain a Padé Approximant. For instance, the expansion of trigonometric functions as polynomials may lead to a useful expression.

3.2.5.2 Padé Approximant to the Raymond filter

As described by Webster et al. (2003), the Met Office use the Raymond filter (Raymond 1988) to smooth the orographic field in their Unified Model. They do this by applying the one-dimensional filter in the two spatial directions successively. The value of the variable parameter $\varepsilon$ in (3.9) is set to 1.0, although it is varied with latitude to take into account some of the anisotropy in the grid caused by spherical geometry. This latitudinal variation should not be necessary in the variational scheme if the two-dimensional version is formulated on the correct geometry, so it should be sufficient to determine the Padé coefficients required for a constant value of $\varepsilon = 1.0$.

With some experimentation, it was found that a reasonable 6th-order Padé Approximant to the Raymond filter with $\varepsilon = 1$ in one dimension is specified by the following coefficients:

$$
\begin{align*}
\alpha_0 &= 1 \\
\alpha_n &= 0 \text{ for } n \neq 0 \\
\beta_1 &= 0 \\
\beta_2 &= 0 \\
\beta_3 &= \left(\frac{\delta}{2}\right)^6 \\
\beta_4 &= \frac{1}{5}\left(\frac{\delta}{2}\right)^8 \\
\beta_5 &= 4\left(\frac{\delta}{2}\right)^{10} \\
\beta_6 &= 12\left(\frac{\delta}{2}\right)^{12} \\
\beta_n &= 0 \text{ for } n > 6,
\end{align*}
$$

(3.33)
Figure 3.2: Comparison between spectral response functions $F(\lambda)$ of the Raymond filter with $\varepsilon = 1.0$ and its variational Padé equivalent defined by (3.33).

where $\delta$ is the grid-length. A comparison between the response of the Raymond filter and the response of the variational filter specified above may be made by referring to figure 3.2.

3.2.6 Minimising the cost function

Having formulated the appropriate cost function, it is necessary to devise some method for minimising it. One of the most commonly used and widely applicable is the Conjugate Gradient Method, described in detail by Shewchuk (1994) and Barrett et al. (1994). The conjugate gradient method works well in a wide range of circumstances; particular problems encountered in using it in this case are detailed later in this chapter.
3.3 Minimising the cost function in two dimensions

The orography used in a global model is clearly a two-dimensional field, and it is conceptually straightforward to extend the one-dimensional filtering described above to take place in two dimensions. However, there turn out to be different ways of implementing the smoothing in this case, particularly in the choice of grid geometry. This choice, in turn, has an effect on the nature of the resultant smoothing, and on the means of minimization.

Regardless of the grid geometry chosen, the continuous two-dimensional cost function is:

\[ J(h) = \int (h^2 - 2\alpha_0 h \bar{h}) \, dA + \sum_{n=1}^{N} (-1)^n \int [\beta_n h \nabla^{2n} h - 2\alpha_n h \nabla^{2n} \bar{h}] \, dA, \]  

(3.34)

which is simply an extension of equation (3.26) to two dimensions. Note that, because we are now integrating over the surface of the sphere, the dimensions of the cost function \( J \) are \([L]^4\), with the dimensions of the different coefficients remaining as they were in the 1-D case.

3.3.1 ‘Planar’ smoothing

3.3.1.1 Discretization

Although the Earth is a sphere, and so the latitude-longitude grid used for the representation of orographic data has unevenly-spaced points, it is instructive to initially consider the implementation of the two-dimensional variational smoothing method on a regular grid. Here, we take the grid to be evenly-spaced in \( x \) and \( y \), as though it were a plane. One consequence of this is that the smoothing will be isotropic relative to the grid spacing. In this case we discretize \( \nabla^2 \) as follows:

\[ (\nabla^2 h)_{ij} = \left( \frac{d^2 h}{dx^2} + \frac{d^2 h}{dy^2} \right)_{ij} \approx \frac{h_{i+1,j} - 2h_{ij} + h_{i-1,j}}{\Delta x^2} + \frac{h_{i,j+1} - 2h_{ij} + h_{i,j-1}}{\Delta y^2}, \]  

(3.35)

with \( \Delta x \) and \( \Delta y \) being the \( x \) and \( y \) grid spacing respectively. Higher orders of the Laplacian, such as \( \nabla^{2n} \) are calculated by repeated application of this operator.
3.3.1.2 Eigenfunctions and eigenvalues

The eigenfunctions and eigenvalues of the Laplacian operator are well-known, and are given by
\[ \nabla^2 e^{i(kx + ly)} = -(k^2 + l^2) e^{i(kx + ly)}. \] (3.36)

Thus, the situation is a direct two-dimensional analogue to the one-dimensional case, where the eigenfunctions are a set of complex exponentials, and the eigenvalues are the negatives of the squares of the corresponding wavenumbers.

As demonstrated in section 3.2.3, the spectral response of the smoothing achieved by minimizing the cost function is determined by the eigenfunctions and eigenvalues of the discretized differential operator. If we discretize \( \nabla^2 \) in the way given in (3.35), then by similar analysis to that given in section 3.2.4, we arrive at the discretized eigenvalue equation:
\[ (\nabla^2 e^{i(kp\Delta x + lq\Delta y)})_{pq} = -(K^2 + L^2) e^{i(kp\Delta x + lq\Delta y)}, \] (3.37)

where \( p \) and \( q \) are the grid-point location indices, and \( K^2 \) and \( L^2 \) are given by
\[ K^2 = \frac{4 \sin^2 \left( \frac{k\Delta x}{2} \right)}{\Delta x^2} = \left[ \frac{\sin^2 \left( \frac{k\Delta x}{2} \right)}{\left( \frac{k\Delta x}{2} \right)} \right] k^2, \] (3.38)
\[ L^2 = \frac{4 \sin^2 \left( \frac{l\Delta y}{2} \right)}{\Delta y^2} = \left[ \frac{\sin^2 \left( \frac{l\Delta y}{2} \right)}{\left( \frac{l\Delta y}{2} \right)} \right] l^2. \] (3.39)

Thus, as with the one-dimensional case, the eigenfunctions of the operator remain the same after discretization, while the eigenvalues are changed. So the spectral response of the variational filter becomes:
\[ F(K, L) = \frac{\alpha_0 + \sum_{n=1}^{N} \alpha_n (K^2 + L^2)^n}{1 + \sum_{n=1}^{N} \beta_n (K^2 + L^2)^n}. \] (3.40)

Given the main characteristic which differentiates it from the ‘spherical’ method to be discussed below is the isotropy of the smoothing with respect to the grid length, it makes most sense to set \( \Delta x = \Delta y \) when applying this method in practice.
3.3.2 ‘Spherical’ smoothing

3.3.2.1 Discretization

The second approach is to consider the grid to be a set of points on a sphere, in order to take into account the curvature of the Earth. The choice of the set of points making up the grid depends on the numerics of the model being used – for a grid-point model, the points will most likely be equally spaced in longitude ($\lambda$) and latitude ($\phi$). However for a spectral model, the latitudinal spacing of the grid is not constant – the details of this are explained later. In contrast to the ‘planar’ method, the smoothing here is isotropic with respect to distances on the surface of the sphere, but not in grid-point space. This will become clear in the following analysis. Note that an alternative possibility for the discretization is to use the sine of latitude ($\mu = \sin \phi$) instead of latitude. The advantage of this is that it simplifies to the expression for an element of area to $dA = d\mu d\lambda$; however, the difference that this would make is very small, and the formulation is less intuitive than latitude-longitude coordinates. (Except where indicated, $\mu$ is used to denote sine of latitude from here on, in contrast to chapter 2, where it represents dimensionless wavenumber.)

For ‘spherical’ smoothing, we consider first the form of $\nabla^2$ in spherical polar coordinates:

$$\nabla^2 h = \frac{1}{a^2 \cos^2 \phi} \left[ \frac{\partial^2 h}{\partial \lambda^2} + \cos \phi \frac{\partial}{\partial \phi} \left( \cos \phi \frac{\partial h}{\partial \phi} \right) \right]$$

(3.41)

We may then discretize using centred differences for the derivatives, taking into account the possible latitudinal variation in grid spacing and grid-box boundaries. Figure 3.3 shows the layout of the grid, where in general $\phi_{q+1} - \phi_{q+1/2} \neq \phi_{q+1/2} - \phi_q$. The discretized operator is thus found to be

$$\nabla^2 h_{pq} = A_q (h_{p+1,q} + h_{p-1,q}) + B_q h_{p,q+1} + C_q h_{p,q-1} - (2A_q + B_q + C_q)h_{pq},$$

(3.42)

with $A_q$, $B_q$ and $C_q$ given by

$$A_q = \frac{1}{a^2 \Delta \lambda^2 \cos^2 \phi_q},$$

(3.43)

$$B_q = \frac{\cos \phi_{q+1/2}}{a^2 \cos \phi_q(\phi_{q+1/2} - \phi_{q-1/2})(\phi_{q+1} - \phi_q)},$$

(3.44)

$$C_q = \frac{\cos \phi_{q-1/2}}{a^2 \cos \phi_q(\phi_{q+1/2} - \phi_{q-1/2})(\phi_q - \phi_{q-1})}.$$

(3.45)
An alternative way of making the discretization of the operator is to consider the continuous operator integrated over the grid-box, which approximates to the value of the discrete operator multiplied by the area of the box $S_{pq}$:

$$\int_{\text{cell}} \nabla^2 h \, dA \approx S_{pq} (\nabla^2 h)_{pq}. \quad (3.46)$$

By Stokes’ Theorem, the left-hand side of equation (3.46) may be rewritten as a line integral around the edge of the cell, so that

$$(\nabla^2 h)_{pq} = \frac{1}{S_{pq}} \oint_{\text{cell}} \mathbf{n} \cdot \nabla h \, dl, \quad (3.47)$$

where $\mathbf{n}$ is the normal vector to the grid-box boundary. We may discretize the components of $\nabla h$ in equation (3.47), so that, for instance, $\nabla h$ across side $BC$ becomes

$$\left. \frac{\partial h}{\partial \lambda} \right|_{BC} \approx \frac{h_{p+1,q} - h_{pq}}{\Delta \lambda \cos \phi_q}. \quad (3.48)$$

The expression for $S_{pq}$ is obtained by considering the area of an infinitesimal latitudinal band on the surface of a sphere. The area $\delta A_\phi$ of the band is given by

$$\delta A_\phi = 2\pi a^2 \cos \phi \, d\phi. \quad (3.49)$$
The area $A_{\phi}$ of the latitudinal band between \( \phi_{q-1/2} \) and \( \phi_{q+1/2} \) is therefore

\[
A_{\phi} = \int_{\phi_{q-1/2}}^{\phi_{q+1/2}} 2\pi a^2 \cos \phi \, d\phi = 2\pi a^2 (\sin \phi_{q+1/2} - \sin \phi_{q-1/2}),
\]

which, assuming that the spacing of grid-points is regular (\( \Delta \lambda \)) in longitude, leads to the following expression for \( S_{pq} \), the area of the grid-box:

\[
S_{pq} = \Delta \lambda a^2 (\sin \phi_{q+1/2} - \sin \phi_{q-1/2}).
\]

Combining these equations together, the expression for the spherical Laplacian operator may be found to be

\[
\nabla^2_{pq} = A_q (h_{p+1,q} + h_{p-1,q}) + B_q h_{p,q+1} + C_q h_{p,q-1} - (2A_q + B_q + C_q) h_{pq},
\]

as before, but with these new expressions for \( A_q \), \( B_q \) and \( C_q \):

\[
A_q = \frac{\phi_{q+1/2} - \phi_{q-1/2}}{a^2 \Delta \lambda^2 \cos \phi_q (\sin \phi_{q+1/2} - \sin \phi_{q-1/2})},
\]

\[
B_q = \frac{\cos \phi_{q+1/2}}{a^2 (\phi_{q+1} - \phi_q)(\sin \phi_{q+1/2} - \sin \phi_{q-1/2})},
\]

\[
C_q = \frac{\cos \phi_{q-1/2}}{a^2 (\phi_q - \phi_{q-1})(\sin \phi_{q+1/2} - \sin \phi_{q-1/2})}.
\]

These expressions are different from those obtained by the first method. However, by comparing the two sets of expressions, we can see that the first set of expressions (3.43-3.45) is also obtainable by applying the Stokes’ Theorem method, but using \( S_{pq} = a^2 (\phi_{q+1/2} - \phi_{q-1/2}) \Delta \lambda \cos \phi_q \), which approximates the grid-box as a trapezium, ignoring the curvature of the Earth. For small grid-boxes, the difference between the two sets of coefficients is small, but the analysis clearly demonstrates that the latter is the appropriate one to use.

### 3.3.2.2 Boundary conditions

The basic characteristics of the domain are shown in figure 3.3. Clearly, the computational domain has zonally periodic boundary conditions, but the boundary conditions at the poles need special consideration. For instance, if you consider a grid-point \((p, q)\) adjacent to the north pole, the three grid-points which are adjacent to it to the south, east and west are obvious \( ((p, q - 1), (p + 1, q), \text{ and } (p - 1, q)) \).
Figure 3.4: A schematic showing the part of the spherical grid near the pole. The grid-points are marked with dots, and one grid-box is shaded. The grid points $A$ and $B$ are considered to be adjacent for the purposes of applying the $\nabla^2$ operator. Note that this is not the only possible type of grid, and one could be devised that included a point at the pole.
respectively), but the point adjacent to the north is not. The most obvious choice is the point which is opposite across the pole. Figure 3.4 illustrates the situation schematically.

However, examination of the coefficients $B_q$ and $C_q$ reveals that they equal zero for the grid points adjacent to the north and south poles, respectively. At these locations, the grid-box boundaries to the poles themselves are actually the points of triangles, and so have zero length, which means they do not contribute to the Stokes’ integral. The somewhat counterintuitive consequence of this is that at the grid-points immediately adjacent to both poles, the Laplacian is calculated with contributions from only three points and not four.

### 3.3.2.3 Eigenfunctions and eigenvalues

As discussed above, the spectral response of the variational filter depends on the eigenfunctions and eigenvalues of the differential operator in the cost function. For the Laplacian operator on the sphere, the eigenfunctions are known as *spherical harmonics*, and are denoted $Y_{m;n}$. In this notation, $m$ is the *order* of the spherical harmonic, while $n$ is the *degree*, such that $|m| \leq n$. In addition, $m$ indicates the zonal wavenumber, while the number of nodes between the two poles is given by $n - m$.

For a full description of spherical harmonics, and their relationship to Legendre polynomials, see, for instance, Durran (1999).

The eigenvalues of the Laplacian operator on the sphere are given by the equation

$$
\nabla^2 Y_{m;n} = \frac{-n(n+1)}{a^2} Y_{m;n}.
$$

By analogy with equation (3.36), it is clear that the total wavenumber associated with $Y_{m;n}$ is $(n^2+n)^{\frac{1}{2}}/a$. Note that this depends only on the degree $n$ of the spherical harmonic, and not upon the zonal wavenumber $m$.

Unfortunately, the discrete operator does not retain the same eigenfunctions or eigenvalues as in the continuous case. This is in contrast to the one-dimensional and ‘planar’ cases, where only the eigenvalues change due to discretization. Because of the complicated form of spherical harmonics, it is not possible to see this analytically, but the eigenfunctions and eigenvalues of the discrete operator may be calculated...
This calculation is impractical for the kind of higher resolutions that would be usefully used in modelling, but may be performed at a lower resolution.

We can think of the operation of the discretized $\nabla^2$ operator in terms of the following matrix-vector expression, where $D_q = -(2A_q + B_q + C_q)$:

\[
\nabla^2 h \equiv \begin{pmatrix} D_1 & A_1 & 0 & \cdots & B_1 & 0 & 0 & \cdots & 0 \\
A_1 & D_1 & A_1 & 0 & B_1 & 0 & 0 \\
0 & A_1 & D_1 & 0 & 0 & B_1 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
C_2 & 0 & 0 & D_2 & A_2 & 0 & 0 \\
0 & C_2 & 0 & A_2 & D_2 & A_2 & 0 \\
0 & 0 & C_2 & 0 & A_2 & D_2 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & 0 & 0 & 0 & D_{N_y} \end{pmatrix} \begin{pmatrix} h_{11} \\
h_{21} \\
h_{31} \\
\vdots \\
h_{12} \\
h_{22} \\
h_{32} \\
\vdots \\
h_{N_x,N_y} \end{pmatrix}, \quad (3.57)
\]

Note that the matrix in (3.57) is not symmetric\(^1\), but may be written as the product of two matrices $S^{-1}$ and $D$, where the latter of these is symmetric and $S$ is a diagonal matrix comprising the areas $S_q$. That this decomposition is possible may be seen by considering equations (3.51–3.55). Thus, we can write $\nabla^2 h \equiv S^{-1}Dh$, which becomes important later.

The eigenvectors and eigenvalues of a general real-valued matrix may be found using standard library routines. If the eigenvalues of the $\nabla^2$ operator are arranged in descending numerical order, a series of discrete levels are observed, corresponding to the multiple eigenfunctions which have the same eigenvalue. This well-known property of spherical harmonics is a fundamental part of the quantum mechanical model of the hydrogen atom (see, for example, Rae (1986)). The first few eigenvalues of $\nabla^2$ are shown in table 3.1. So, for each value of $n$, there are $2n+1$ eigenfunctions with eigenvalue $-n(n+1)/a^2$.

If we calculate the eigenfunctions of the discrete operator, there will be as many eigenfunctions as there are points in the domain (i.e. $N_xN_y$), but which spherical harmonics should we expect those eigenfunctions to correspond to? The answer to this can be found by considering the spherical harmonics which may be represented

---

\(^1\)A square matrix is symmetric if it is equal to its own transpose
Table 3.1: Values of the first few eigenvalues of $\nabla^2$ on the unit sphere ($a = 1$).

<table>
<thead>
<tr>
<th>$n$</th>
<th>$m$</th>
<th>eigenvalue ($= \frac{-n(n+1)}{a^2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>-2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-2</td>
</tr>
<tr>
<td>2</td>
<td>-2</td>
<td>-6</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>-6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-6</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>-6</td>
</tr>
</tbody>
</table>

by a grid of a particular resolution. If there are $N_x$ grid-points zonally, and $N_y$ grid-points between the poles, then these are the maximum numbers of nodes that may be resolved in the two directions. For a spherical harmonic $Y_{m,n}$, the number of nodes in the zonal direction is equal to $m$, while the number of nodes between the poles is equal to $n - m$. Thus, in $m$-$n$ space, the region containing the spherical harmonics which may be represented on the spherical grid satisfies $|m| \leq N_x/2$ and $n - |m| \leq N_y$, in addition to $|m| \leq n$, as specified above. The resulting region is shown in figure 3.5, and is in the form of two contiguous parallelograms.

The grid used in the a typical numerical model will have $N_x = 2N_y$, so that on average the angular spacing of the grid points will be the same in latitude and longitude. As a basis for comparing the idealized and discretized operators, consider a 10×20 grid of this type. The grid is a so-called Gaussian grid, which means that the meridional grid spacing depends on the location of the nodes of the highest resolved spherical harmonic. This is the type of grid usually used in spectral models of the atmosphere, which will be explained in more detail later on.

For a 10×20 grid, there are 200 degrees of freedom, and so 200 spherical harmonics may be resolved. For the rhomboidal truncation described above, the eigenvalues of the resolved spherical harmonics, and the number present for each eigenvalue, are shown in table 3.2. The eigenfunctions and eigenvalues of the discretized operator were calculated, and sorted into descending numerical order for comparison. The
Figure 3.5: The region in $m$-$n$ space corresponding to the spherical harmonics resolved by an idealized discretization of $\nabla^2$. This is the same so-called rhomboidal truncation used in some spectral atmospheric models. Note that $N_y = N_x/2$. Also, note that this representation is not isotropic on the sphere, whereas triangular truncation, to be described in section 4.2.1, is isotropic.
### Table 3.2: Number of eigenfunctions for each eigenvalue present in a rhomboidal truncation with 200 degrees of freedom, ranked in descending numerical order.

<table>
<thead>
<tr>
<th>Value of $n$</th>
<th>Eigenvalue</th>
<th>Number of eigenfunctions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>-6</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>-12</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>-20</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>-30</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>-42</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>-56</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>-72</td>
<td>17</td>
</tr>
<tr>
<td>9</td>
<td>-90</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>-110</td>
<td>19</td>
</tr>
<tr>
<td>11</td>
<td>-132</td>
<td>17</td>
</tr>
<tr>
<td>12</td>
<td>-156</td>
<td>15</td>
</tr>
<tr>
<td>13</td>
<td>-182</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>-210</td>
<td>11</td>
</tr>
<tr>
<td>15</td>
<td>-240</td>
<td>9</td>
</tr>
<tr>
<td>16</td>
<td>-272</td>
<td>7</td>
</tr>
<tr>
<td>17</td>
<td>-306</td>
<td>5</td>
</tr>
<tr>
<td>18</td>
<td>-342</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>-380</td>
<td>1</td>
</tr>
</tbody>
</table>
results are shown in figures 3.6 and 3.7. For the first few eigenvalues (i.e. the largest-scale spherical harmonics) the discretized operator shows similar behaviour to the continuous operator. However, the correspondence does not extend beyond $n = 5$ (the 36$^{th}$ eigenvalue). For most of the eigenfunctions resolved, the eigenvalues are of less than the correct magnitude, but beyond the 170$^{th}$ eigenvalue the eigenvalues of the discrete operator are of greater magnitude than they should be. This disparity grows rapidly over the higher numbered eigenvalues, so that for the last placed one, the eigenvalue of the discretized operator is over $2\frac{1}{2}$ times that expected.

Examining the eigenfunctions themselves demonstrates that their accuracy mirrors that of the corresponding eigenvalues. To illustrate this, figures 3.8–3.10 compare selected eigenfunctions of the discrete and idealized operators. At higher eigenvalues, the eigenfunctions are less and less recognisable as spherical harmonics, and so a one-to-one correspondence is not possible; figure 3.11 shows two examples of this kind of eigenfunction.

As explained above, the nature of the smoothing effected by minimising the cost function depends on the eigenfunctions and eigenvalues of the differential operator. When the eigenfunctions are the same for the discretized operator as they are for the continuous operator, this means that we can think of the smoothing in spectral terms. But when the eigenfunctions of the discretized and continuous operators are not the same, as is the case with the spherical Laplacian, it is wrong to think of the smoothing being applied spectrally. Here, we must think of the orographic field decomposed into the eigenfunctions of the discretized Laplacian. The amount of attenuation applied to each of these components is determined by their respective eigenvalues. Put formally, the mean orographic field may written as a sum

$$h(x, y) = \sum_{i=1}^{N_x N_y} c_i \psi_i,$$

(3.58)

where $\{\psi_1 \ldots \psi_{N_x N_y}\}$ is the set of eigenfunctions of the discretized spherical Laplacian at a resolution of $N_x \times N_y$. If the corresponding set of eigenvalues is denoted $\{\hat{k}_1^2 \ldots \hat{k}_{N_x N_y}^2\}$, then the smoothed orography $h(x, y)$ is related to the mean orography
Chapter 3  Two-dimensional orography – a variational approach to smoothing

Figure 3.6: Comparison between the eigenvalues of the discretized and idealized-continuous Laplacian operators (solid and dashed lines, respectively) shown over the whole range. In both cases, the eigenvalues are arranged in descending numerical order.

Figure 3.7: As figure 3.6, but restricted to the first 100 eigenvalues.
Chapter 3  Two-dimensional orography – a variational approach to smoothing  62

Figure 3.8: Comparison between the spherical harmonic \( m = 1, n = 1 \) (eigenvalue \(-2\)) and the corresponding eigenfunction of the discrete \( \nabla^2 \) operator (number 3 in the sequence, with eigenvalue \(-1.98\)). The contouring is arbitrary; the \( x \) and \( y \) axes show longitude and latitude, respectively, in grid-points.

Figure 3.9: Comparison between the spherical harmonic \( m = 3, n = 10 \) (eigenvalue \(-110\)) and the corresponding eigenfunction of the discrete \( \nabla^2 \) operator (number 102 in the sequence, with eigenvalue \(-61.94\)). The contouring is arbitrary; the \( x \) and \( y \) axes show longitude and latitude, respectively, in grid-points.
Figure 3.10: Comparison between the spherical harmonic $m = 4$, $n = 11$ (eigenvalue $-132$) and the corresponding eigenfunction of the discrete $\nabla^2$ operator (number 120 in the sequence, with eigenvalue $-77.31$). The contouring is arbitrary; the $x$ and $y$ axes show longitude and latitude, respectively, in grid-points.

Figure 3.11: Two eigenfunctions of the discrete $\nabla^2$ operator with high eigenvalues: (a) is 171 in the sequence, and has eigenvalue $-221.77$; (b) is 200 in the sequence, and has eigenvalue $-1001.94$. The contouring is arbitrary; the $x$ and $y$ axes show longitude and latitude, respectively, in grid-points.
\( \tilde{h}(x, y) \) as follows:

\[
h(x, y) = \sum_{i=1}^{N_x N_y} \left[ \alpha_0 + \sum_{n=1}^{N} \alpha_n \hat{k}_i^{2n} \right] \left( \frac{1}{1 + \sum_{n=1}^{N} \beta_n \hat{k}_i^{2n}} \right) c_i \psi_i, \\
\]  

so that the spectral response function derived for the continuous case (3.11) becomes a scaling factor in the expansion of the orographic field as a sum of the discrete operator’s eigenfunctions.

The response of this filter expressed in terms of a spherical harmonic decomposition (i.e. the spectral response) is not easily related to this. Consider that the eigenfunctions \( \{\psi_i\} \) may each be expressed as a sum of spherical harmonics:

\[
\psi_i = \sum_{p=1}^{P} s_{p,i} Y_p, \\
\]  

where \( \{Y_p\} \) is the set of \( P \) spherical harmonics which may be resolved at the given resolution, and \( \{s_{p,i}\} \) are the corresponding expansion coefficients for eigenfunction \( \psi_i \). By combining (3.58) and (3.60), we can see how the two sets of coefficients \( \{c_i\} \) and \( \{s_{p,i}\} \) are related to the spectral coefficients:

\[
\tilde{h} = \sum_{p=1}^{P} \left[ \sum_{i=1}^{N_x N_y} c_i s_{p,i} \right] Y_p. \\
\]  

Note that actually \( P = N_x N_y \). By similarly combining (3.58) and (3.60) with (3.59), the smoothed orography \( h(\lambda, \phi) \) may be expressed in terms of spherical harmonics:

\[
h(\lambda, \phi) = \sum_{p=1}^{P} \sum_{i=1}^{P} \left[ s_{p,i} c_i \left( \frac{\alpha_0 + \sum_{n=1}^{N} \alpha_n \hat{k}_i^{2n}}{1 + \sum_{n=1}^{N} \beta_n \hat{k}_i^{2n}} \right) \right] Y_p. \\
\]  

So the determination of the spectral response of the filter depends on knowing the decomposition of the eigenfunctions of the discretized \( \nabla^2 \) operator in spherical harmonics. Since this involves computing the eigenfunctions of the discretized operator to start with, it is only practical for low resolutions, and would be computationally too expensive to do for the kinds of resolutions typical of an operational NWP or climate model. It is true, though, that the eigenfunctions which are on the smallest scales have the highest eigenvalues, which means that qualitatively the effect
of the smoothing should be similar to that which would result if the $\nabla^2$ operator was ‘perfect’. However, the larger-scale eigenfunctions of the continuous operator may contain some components of small-scale spherical harmonics, and so this could result in the small-scale spherical harmonics not being attenuated as effectively as desired.

One possibility not discussed above is the use of a spectral transform to apply the $\nabla^2$ operator. This would be possible, and would lead to the correct spectral response by the variational method. However, it is likely that the eigenfunctions and eigenvalues of the discrete operator are more relevant to the properties of a grid-point numerical scheme than are those of the continuous operator.

### 3.3.3 Practical minimization

#### 3.3.3.1 Matrix formulation

The first step in the practical minimization of the cost function (3.34) is to rewrite it in matrix form. We do this by considering writing the mean and smoothed orographies $\overline{h}$ and $h$ as column vectors $\overline{h}$ and $h$. The ordering of the grid-points as elements within the vectors is not strictly important, but it is helpful to think of them as being ordered in rows: $h_{11} \ldots h_{N_1}$, then $h_{12} \ldots h_{N_2}$, etc. The effect of integrating over the sphere needs to be taken into account, and this is done by introducing a diagonal matrix $S$, whose elements are the areas of the grid boxes. Finally, the operator $\nabla^2$ is represented by $S^{-1}D$, as described in section 3.3.2.3 above. Thus, (3.34) may be recast in matrix notation as follows:

$$J(h) = h^TSh - 2\alpha_0h^TS\overline{h} + \sum_{n=1}^{N} \left( \beta_n h^T S \left[ S^{-1}D \right]^n h - 2\alpha_n h^T S \left[ S^{-1}D \right]^n \overline{h} \right) \quad (3.63)$$

The gradient vector $\nabla J$ (that is, the gradient of $J$ in $h$-space) may be calculated by considering a small arbitrary increment $\varepsilon h'$ to $h$, and ignoring terms in $\varepsilon^2$:

$$J(h + \varepsilon h') - J(h) = \varepsilon h'^TSh + \varepsilon h'^TSh' - 2\alpha_0\varepsilon h'^T S \overline{h}$$

$$+ \sum_{n=1}^{N} \left[ \beta_n \left( \varepsilon h'^T S \left[ S^{-1}D \right]^n h + \varepsilon h'^T S \left[ S^{-1}D \right]^n h' \right) - 2\alpha_n \varepsilon h'^T S \left[ S^{-1}D \right]^n \overline{h} \right]. \quad (3.64)$$
For a matrix $A$, then $x^T Ay = y^T A^T x$. $A$ is self-adjoint if $\tilde{A}^T = A$, where $\tilde{A}$ denotes the complex conjugate of $A$, and $A^T$ denotes the transpose of $A$. If $A$ is self-adjoint, then clearly $x^T Ay = y^T A x$. In addition, for matrices $A$ and $B$, in general $(AB)^T = B^T A^T$. Taking these things together, also noting that $S$ is diagonal, and that $D$ is self-adjoint, we can use (3.64) to obtain

$$
\delta J = 2Sh - 2\alpha_0 S\tilde{h} + \sum_{n=1}^{N} (2\beta_n S[S^{-1}D]^n h - 2\alpha_n S[S^{-1}D]^n \tilde{h}).
$$

(3.65)

### 3.3.3.2 Conjugate Gradient Method

The Conjugate Gradient Method (CG) is a well-known iterative method for solving systems of linear equations. It is described in depth in, for instance, Shewchuk (1994), but there are some aspects of it which are worth exploring here, as they illuminate practical issues in the use of the method to minimize $J$.

The cost function $J$ is a quadratic function of the form

$$
J(h) = \frac{1}{2} h^T Ah - b^T h + c,
$$

(3.66)

where $A$ is given by

$$
A = 2S \left( 1 + \sum_{n=1}^{N} \beta_n [S^{-1}D] \right),
$$

(3.67)

$b$ is given by

$$
b = 2S \left( \alpha_0 + \sum_{n=1}^{N} \alpha_n [S^{-1}D] \right) \tilde{h},
$$

(3.68)

and with $c = 0$. Minimising $J(h)$ is equivalent to solving

$$
Ah = b.
$$

(3.69)

The method will be able to solve (3.69) if $A$ is *positive-definite*, that is, if all the eigenvalues of $A$ are positive. The speed of convergence of CG depends on the clustering of the eigenvalues, and the condition number $\kappa$ of $A$. The condition number of a matrix is defined as the ratio of the largest and smallest eigenvalues; the maximum number of iterations $i$ required by CG for convergence is

$$
i \leq \left\lceil \frac{1}{2} \sqrt{\kappa} \ln \left( \frac{2}{\varepsilon} \right) \right\rceil,
$$

(3.70)
where $\epsilon$ is a convergence parameter (the factor by which the iterative process reduces the norm of the error). A matrix with a very large condition number is said to be *ill-conditioned*.

If the convergence of CG is impaired by matrix $A$ being ill-conditioned, it is sometimes possible to overcome this by using the technique of *preconditioning*. This involves multiplying (3.69) by the matrix $M^{-1}$, which is chosen so that $M^{-1}A$ has a smaller condition number, or more favourably-clustered eigenvalues, than $A$. This means that instead of having to solve (3.69), the problem now becomes the solution of

$$M^{-1}Ah = M^{-1}b. \quad(3.71)$$

In general, making a good choice for $M^{-1}$ is not straightforward. An ideal preconditioner would be to use $M = A$, but to do this involves inverting $A$, which is the same problem that we are trying to solve in the first place. At the other extreme, $M = I$ (I is the identity matrix) simply leaves the problem as it is. It is sometimes possible to make some improvement to the convergence properties of a problem by using a diagonal preconditioner. In this case, $M$ is a diagonal matrix whose diagonal entries are those of matrix $A$, so the calculation of $M^{-1}$ is straightforward.

In applying CG to the present problem, the method of Shanno and Phua (1976) is used. The convergence required before the algorithm terminates is specified by the convergence parameter $\epsilon$; If the current estimate of the solution is $x$, and its gradient is $\tilde{x}$, then convergence occurs when

$$|\tilde{x}| \leq \epsilon \max(1, |x|). \quad(3.72)$$

### 3.4 Other constraints

The main purpose of formulating the smoothing as a cost function to be minimized, rather than as an explicit spectral filter, is to enable the addition of extra constraints upon the smoothing. The two extra constraints which were implemented in this case concerned the flatness of the sea, and the height of dynamically important barriers. They were chosen because, as explained in chapter 1, they are potentially the most significant drawbacks with orographic smoothing.
3.4.1 Two different types of constraint

Additional constraints may be added to a cost function minimisation process in two ways: by devising the minimisation process to ensure that the constraint is necessarily fulfilled, or by introducing an extra term in the cost function which penalizes the extent to which the constraint is unfulfilled. The first of these is known as a \textit{strong constraint}, while the second is a \textit{weak constraint}. Using a weak constraint allows the desired properties of the orography to be balanced against each other, whereas a strong constraint is an unyielding condition under which the other constraints in the cost function are applied. The strong constraint approach is only suitable for the application of a limited range of constraints, as it involves adjusting the minimisation method in some way. The difference between these two constraints may also be understood in terms of the formalism of slack functions and Lagrange multipliers, respectively; for details, see, for instance, Daley (1991) or Sasaki and McGinley (1981).

3.4.2 Constraining barrier height

Orographic barriers can have a profound effect on the motion of the atmosphere. Particularly long and high barrier ridges may divert the flow to be parallel to the ridge, rather than allowing it to flow over. Clearly, if the smoothing process lowers the height of an orographic barrier in a model, this may cause the flow in the model to differ fundamentally from the flow in the real world. It may be possible to lower the barrier height a little without effecting such a change in flow regime, and for this reason the barrier height constraint is implemented in this variational scheme as a weak constraint. This is done by adding this term to the cost function (3.26):

\[ J_{\text{peaks}}(h) = \gamma \sum_{\text{peaks}} S_{pq}(h - \bar{h})^2, \]  

(3.73)

where $S_q$ is the area of a grid-box at latitude $q$, $\gamma$ is a dimensionless weighting parameter, and the sum is carried out only over those points previously determined to be the points whose heights we wish to additionally constrain. Note that this term has an identical form to the $J_{\text{mean}}$ term, except for applying to a limited range
of points. To incorporate this term into the matrix formulation of the variational scheme, it may be rewritten as

$$J_{\text{peaks}}(h) = \gamma(h^T P sh - 2h^T P \overline{sh}),$$

(3.74)

where $P$ is a diagonal matrix whose diagonal elements are set to one for elements corresponding to peaks, and zero otherwise.

It is clear that for this additional term to be used in the cost function, it is necessary to have some objective procedure for determining the location of the peaks whose heights are to be constrained. Ideally, this would be done by studying flow regimes in a general circulation model, experimenting with reduced barrier heights, and noting those locations where such a reduction had a particularly detrimental effect. However, such a study would in itself be a major undertaking, and is beyond the scope of the present work. A more feasible approach is to identify significant ridges solely as a property of the orography itself, rather than of its effect on the flow. A variety of possible methods for accomplishing this are discussed in chapter 4. Note also that while the extra term given above constrains the barrier points to be close to their height in the mean orography, this is not the only possibility; if desired, the mean orography in this term could be replaced by the field of sub-grid-scale orographic maxima, so countering the barrier-height loss caused in the generation of the mean orography.

3.4.3 Constraining the flatness of the sea

In an NWP or climate model, elevated sea-points act as spurious sources of heat, so it might be good to prevent their generation by orographic smoothing. The requirement that the sea be flat lends itself to being applied as a strong constraint, since it is obvious what the elevation of the sea points should be, and only a completely flat sea will avoid the spurious generation of atmospheric waves over its surface.

As noted in above, the cost function used in the variational method may be written in matrix form (see equation (3.63)). Where the flatness of the sea is unconstrained, the vectors $h$ and $\overline{h}$ contain all the points in the domain; their order in the vector is not important. If we wish to apply a flat-sea constraint, however, this may simply
be done by only including the land points in the vector. Clearly, the contents of the matrices \( S \) and \( D \) need to be adjusted to take account of that; however, the formulation of \( \nabla^2 \) (on which \( S \) and \( D \) depend) is straightforward in these circumstances. The points missing from the vector are simply assumed to be zero when applying the discretized form of \( \nabla^2 \), given by (3.52), and substituting the appropriate coefficients.

In the implementation of the flat sea constraint, object-oriented programming (OOP) techniques have been used. The use of OOP concepts allows the writing of code with greater clarity and reliability, as explained in, for example, Schildt (1998). These concepts are not so easy to implement in FORTRAN 90, but possible ways of doing so are explored in Decyk et al. (1997). Of relevance to the flat-sea constraint is the concept of data-hiding; by defining a new data type, corresponding to a masked data-set, and then defining the operations that may act on that type (such as the arithmetic operators), the need to know how the masked data is handled is removed from the main program. Consequently the FORTRAN expression

\[
\text{gamma*sum(areas*peak_mask*(h*h-2.0*h*hbar)*precon)}
\]

may be written for the peak-constraint term given in (3.73), with \( h, hbar, precon, areas \) and \( peak\_mask \) all being of type \texttt{masked\_array}. This code will work whether or not we are imposing a flat sea, providing the \texttt{masked\_array} data-type’s appropriate built-in function is first used to fix the set of points in use.

### 3.5 Summary

In this chapter, a new, flexible variational filter has been described. It is easily able to emulate a wide range of linear filters, while also applying additional constraints. These include, but are not limited to, the maintenance of the flatness of the sea and the height of important orographic ridges. This combination of attributes makes it ideally suited to addressing some of the major concerns surrounding the use of smoothed orography in numerical models of the atmosphere. Furthermore, the method has been subjected to a detailed mathematical analysis, dealing both with the practical aspects of its implementation, and also the predicted properties.
of the filter. It is the practical implementation of the method, and the verification and description of its effect, that are the subjects of chapter 4.
Chapter 4

Applying the variational method to real orography

4.1 Introduction

In this chapter, the variational method described in chapter 3 is applied to a real orographic dataset. The context of this is the spectral shallow water model used in chapter 5 to test the variational method. So, in this chapter, the aspects of the model pertaining to the choice of grid on which the orography is defined are discussed. Practical aspects of the method are expanded, particularly the question of the identification of ridges whose height should be constrained. In addition, the effects of applying the variational method to the orography are explored in detail.

4.2 Grids in spectral models

In a spectral model of the atmosphere, the model fields are, at a fundamental level, represented as sets of coefficients of the terms in a spherical harmonic expansion. However, to make the problem computationally tractable, the non-linear terms of the equations of motion are calculated in grid-space. The choice of grid is determined by considering the resolution and truncation of the model. These well-known aspects of spectral model design are discussed in detail in, for instance, Durran (1999), but it is worth reviewing them here, as they are relevant to the later parts of the thesis.
4.2.1 Model truncation

*Truncation* refers to the spectral resolution of the model, particularly the shape of the region in $n$-$m$ space which contains the resolved spherical harmonics. There are two common types of truncation, *rhomboidal* and *triangular*. (*Rhomboidal* is something of a misnomer, as the regions in question are actually parallelograms.) For a discussion of truncation, see, for example, Bourke (1988).

Rhomboidal truncation was discussed in section 3.3.2.3 above, and illustrated in figure 3.5. Note, however, that this figure was constructed starting from the properties of the grid. Strictly, the truncation of the model stands apart from the choice of grid, so it is better to talk about the truncation in terms of a truncation limit $M$. In this notation, the region containing the resolved spherical harmonics is given by $|m| \leq M$, $n \leq |m| + M$, and $|m| \leq n$. Rhomboidal truncation is isotropic and homogeneous in *latitude-longitude* space, which means that the smallest-scale spherical harmonics contained in the truncation are also the smallest-scale waves which may be resolved by a suitable latitude-longitude grid. Another way of understanding this is to consider the resolution limit of a 1D grid. The smallest resolvable wave has $\lambda = 2\Delta x$ – the two-grid wave. For rhomboidal truncation, the smallest-scale spherical harmonic varies as a two-grid wave both in the meridional and zonal directions. The number of degrees of freedom in rhomboidal truncation is $M^2$.

Triangular truncation consists of the spherical harmonics in the region $|m| \leq M$ and $n \leq M$, in addition to $|m| \leq n$. The resulting region is shown in figure 4.1. This truncation is homogeneous and isotropic in terms of physical space – the surface of the sphere – which means that the size of the smallest resolvable features doesn’t depend on their location or orientation. For a spherical harmonic $Y_{m,n}$, the number of nodes zonally is $|m|$, while the number between the poles is $n - |m|$. Therefore, the smallest-scale spherical harmonics resolved by the truncation range from having $M$ nodes between the poles, with no zonal variation ($Y_{0,M}$), to having $M$ nodes zonally, with no nodes between the poles ($Y_{M,M}$). The number of degrees of freedom in triangular truncation is $\frac{1}{2}M^2$.

The spectral shallow-water model used in chapter 5 uses triangular truncation, and the following discussion is relevant to that situation.
4.2.2 Choosing a grid for a spectral model

The grid in a spectral model is used for the calculation of non-linear products in the equations of motion. Such products will, in general, generate spectral components which have a wavenumber equal to the sum of the wavenumbers of the two parts of the product. Therefore, to be able to resolve all the waves generated by the calculation, and so avoid errors due to aliasing, the grid must have a higher spectral resolution than the model itself. It is not necessary for the grid to be able to resolve all the waves generated by the calculation, as, beyond the truncation limit of the model, they will be discarded. So, the grid needs only to have sufficient resolution to avoid this aliasing affecting the modes which are to be retained. This grid is, by convention, known as a quadratic grid, and has \( N_x \geq 3M \) and \( N_y \geq (3M + 1)/2 \). The grid was originally proposed by Eliassen et al. (1970), while Orszag (1970) and Machenhauer (1979) demonstrated that it has sufficient resolution to resolve non-linearities that are the product of two terms (i.e. quadratic). If non-linearities resulting from the product of three terms were present in the model equations, then this would require a cubic grid, with \( N_x \geq 6M \) and \( N_y \geq (6M + 1)/2 \).

Orography is usually supplied to a spectral model as a grid-point field, and using a quadratic grid causes problems. Because there are more degrees of freedom on the grid than in the spectral representation, the spectrum of the orography represented on the grid will be truncated when it is used in the model. This results in the
Gibbs effect occurring at sudden changes in the orographic height, for instance at the land/sea interface. Thus, to perform any meaningful kind of experiment involving changes to the orography, it is necessary to counter this effect. One possible way to do this is to use a grid of a lower resolution, so that the number of degrees of freedom in the grid more closely matches the number of resolved spherical harmonics. Côté and Staniforth (1988) describe what they term a reduced-resolution grid, sometimes known as a linear grid, a more detailed exposition of which is given in Hortal (2002). This grid has \( N_x \geq 2M + 1 \) and \( N_y \geq (2M + 1)/2 \), and can be thought of as a quadratic grid for spectral resolution \( M' \), where \( M' = 2M/3 \). The number of degrees of freedom in the grid is still approximately twice the number of resolved spherical harmonics in the triangular truncation, but it would be impossible to construct a latitude-longitude grid where there was an exact correspondence with the resolved spherical harmonics in a triangular truncation. As has already been demonstrated, such a grid necessarily resolves a set of spherical harmonics which form a rhomboidal truncation.

In the experiments detailed in the next chapter, the orography is supplied to the model on such a linear grid. The grid orography is transformed to spectral space, and the spectral coefficients are then used in the model calculations. This process is used to reduce the truncation error resulting from defining the orography on a grid.

An important aspect of the spectral model grid is the meridional spacing of the grid-points. These are not quite spaced equally in latitude; their spacing is generally chosen to allow the use of Gaussian quadrature; this is exact for the Legendre polynomials which comprise the latitudinal variation of the spherical harmonics. Hence, the grid is known as a Gaussian grid. Details of Gaussian quadrature may be found in, for instance, Durran (1999). Although the latitudinal spacing of a Gaussian grid is not equal, as the resolution increases, the spacing tends asymptotically towards equality.

Due to the use of Gaussian quadrature, it is also not generally the case that the latitudinal grid-box boundaries lie halfway between the grid-points. As explained in Williamson and Laprise (2000), the Gaussian weight \( w_q \) at latitude \( q \) is given by:

\[
  w_q = \frac{2(1 - \mu_q^2)}{|Q P_{Q-1}(\mu_q)|^2},
\]  

(4.1)
where $Q$ is the number of points latitudinally between the poles, $\mu_q = \sin \phi_q$, and $P_Q$ is the Legendre polynomial of order $Q$. The location of the boundaries between the grid boxes, and thus the areas of those boxes, may be calculated using the fact that $w_q = \Delta \mu_q$.

The grid used in the work presented in the remainder of this chapter is discretized on a Gaussian grid with $N_x = 200$ and $N_y = 100$. This is the quadratic grid for a triangular truncation with $M = 66$ (denoted T66), and a linear grid for a triangular truncation with $M = 99$ (T99).

### 4.3 Mean orography

#### 4.3.1 Calculating the mean orography

In order to apply the variational method, an unsmoothed, mean orography is required to start with. This was obtained from the GLOBE dataset (Hastings et al. (1999)), and averaged over each grid-box on the T99 linear grid. The manner of averaging – simply taking the mean of all the high-resolution points that fall within the T99 grid-box – is not isotropic in terms of the surface of the sphere, and consequently there will be more smaller-scale features at high-latitudes. The alternative would be to use some kind of kernel averaging, but this would involve the specification of the kernel function, and would itself be a form of smoothing. Despite the more noisy result, it seems better to use a simple mean to generate the low resolution orography in order to avoid confusion between the smoothing effected by the averaging process, and that resulting from the variational method under test. The resulting mean orography is shown in figure 4.2.

#### 4.3.2 Orographic spectra

In spherical polar coordinates, and using triangular truncation $M$, a field such as orography may be decomposed spectrally into spherical harmonics so that:

$$h_0(\lambda, \mu) = \sum_{m=0}^{M} \sum_{n=m}^{M} c_{m,n} Y_{m,n}(\lambda, \mu), \quad (4.2)$$
where $\lambda$ is longitude, $\mu$, sine of latitude, $\{c_{m,n}\}$, a set of constant coefficients that may be complex, and $\{Y_{m,n}(\lambda, \mu)\}$, the set of spherical harmonics\(^1\). As explained in chapter 3, the total wavenumber $k$ associated with spherical harmonic $Y_{m,n}$ is given by $k = (n^2 + n)^{1/2}/a$, where $a$ is the radius of the Earth. Here, for convenience, we set $a = 1$.

One way of understanding the decomposition of a field into spherical harmonics is to take the mean of the magnitudes for each wavenumber, so that:

$$\hat{h}_0(n) = \frac{1}{n+1} \sum_{m=0}^{n} |c_{m,n}|,$$  \hspace{1cm} (4.3)

which may then be plotted against $k$. Although this is a slightly unconventional way of calculating the spectrum, it is convenient because it allows the correct calculation of the spectral response of a filter by dividing one spectrum by another. The spectral response of the variational filter in the previous chapter is given in terms of the effect on individual spherical harmonics of wavenumber $k$. By simply adding the magnitudes of these components together, rather than taking a sum of

\(^1\)The sum in (4.2) could also be written with $m$ going from $-M$ to $M$, in which case $\{c_{m,n}\}$ would be real. For our purposes, however, it is clearer to write it in the way shown.
the squares of the magnitudes, as is often done (see, for instance, Balmino (1993)),
the verification of these predicted responses may be done easily. The spectrum of
the mean orography calculated in this manner is shown in figure 4.3. Note that,
although the spectrum decreases with wavenumber, it only does so slowly, leaving a
significant component at the truncation limit.

4.4 Methods for identifying important ridges

A number of different possible criteria for the identification of ridges have been
devised; they are explained and illustrated in the following sections. The set of
points which are found by applying these methods is termed a peak mask. In all the
examples given, the orography used to generate the peak mask is the unsmoothed
mean orography on the T99 linear grid.

The evaluation of the different methods is difficult to do objectively, since it would
either rely on some pre-existing objective method for identifying ridges or a sound
knowledge of the dynamically important orographic features. So, probably the only
way to evaluate the possibilities is by eye. A successful method should pick up
the major orographic ridges of the world – the Andes, the Rockies, the Urals, etc. However, it is unclear what is desirable for larger areas of elevated orography, such as the Tibetan Plateau. While these areas are certainly important for atmospheric dynamics, it might not be beneficial to constrain the height of the whole upland region, since that would attenuate the spectral smoothing in that area. Where the upland area encompasses a small range of heights, it may be appropriate to constrain the heights of the orography at the grid-points around the edge of the region. Taking these problems into account, the identification of the major ridges is used in the following sections to evaluate the various methods.

This problem is similar to some of those encountered in the field of machine vision, particularly in binary image processing. Correspondingly, some of the methods described below share properties with the basic methods used in this field. For a thorough discussion, see, for instance, Jain et al. (1995).

4.4.1 Local maxima

Perhaps the simplest method for locating important orography is to check each point to see whether it is higher than its surroundings. In the case of a rectangular grid, there is a choice to be made regarding the points deemed to be adjacent to the point under consideration. One possibility is to consider only those grid points which lie to the north, south, east and west of the current location. The second option is to include not only these, but also the points adjacent at the corners of the grid-box. The two possibilities are shown in figure 4.4, labelled (a) and (b) respectively. The results of applying these two methods to the mean orography are shown in figures 4.5 and 4.6. Of these two plots, figure 4.5 has more features that look like ridges, but, of course, they are still made up of disconnected points. Figure 4.6 seems much more to be composed of randomly distributed points. Both plots contain mask points which are very low lying, as would be expected. It would be possible to introduce a height requirement to the selection criteria, but this would merely reduce the number of points in the set, and not bring about a more contiguous mask.
Figure 4.4: Two possible domains of adjacent points. In these plots, the black square is the point under consideration, and the grey squares are points held to be adjacent to it.

4.4.2 Perpendicular second derivatives

This idea stems from a consideration of the properties of a ridge, particularly the horizontal second derivatives at a point on the ridge. At such a point, the second derivative along the ridge may be of either sign, but will be small. The second derivative perpendicular to the ridge will be negative, and large. So, to determine whether a given point is a ridge point, one only needs to determine the second derivatives, and find out whether they conform to this description. Unfortunately, the ridges are not in general aligned with the grid, but this can be taken into account by calculating the eigenvalues of the matrix of second derivatives (known as the **Hessian**):

$$
\begin{vmatrix}
\frac{\partial^2 h}{\partial x^2} & \frac{\partial^2 h}{\partial x \partial y} \\
\frac{\partial^2 h}{\partial y \partial x} & \frac{\partial^2 h}{\partial y^2}
\end{vmatrix}
$$

These eigenvalues are then the second derivatives parallel and perpendicular to the ridge. The way to understand this is as follows: if the ridge was aligned with one of the axes, then the two cross derivative elements in the matrix would be zero, and the eigenvalues would simply be the two second derivatives \( \frac{\partial^2 h}{\partial x^2} \) and \( \frac{\partial^2 h}{\partial y^2} \). In this situation, these are clearly the second derivatives parallel and perpendicular to the ridge. But the eigenvalues of a matrix are invariant under rotation\(^2\), so regardless of the orientation of the ridge, the eigenvalues will always give the second derivatives parallel and perpendicular to it.

\(^2\text{This is because rotation is a similarity transformation; see, for instance, Press et al. (1992)\)}
Figure 4.5: Peak mask consisting of points higher than four surrounding grid points. The dotted line indicates the land-sea boundary. The axes are labelled in grid-points.

Figure 4.6: Peak mask consisting of points higher than eight surrounding grid points. The axes are labelled in grid-points.
So, to determine which are ridge points, it is necessary to specify the values of two parameters: the maximum allowed magnitude of the along-ridge second derivative (which may be of either sign), denoted $\varepsilon_1$, and the maximum allowed value of the across-ridge second derivative (which must be negative), denoted $\varepsilon_2$. If the point in question is a ridge then the magnitude of the negative eigenvalue will always be larger than the other eigenvalue, allowing easy distinction between the two.

In implementing this method, it was decided to calculate the derivatives as though the orography was defined on an $x$-$y$ grid, with $\Delta x = \Delta y = 1.0$ m. The problem with trying to make the method work in spherical geometry is that as the grid converges towards the poles, it becomes less clear how to take derivatives in perpendicular directions. However, one consequence of using planar geometry here is that north-south aligned ridges near the poles will seem less sharply defined than those near the equator, because the $x$ grid-length has been artificially stretched. This will mean that a ridge of particular dimensions may be flagged as a ridge by the method if it lies near the equator, but not if it lies near the pole.

The method was tried with several different values of the two parameters. The outcome for one of the more successful combinations is shown in figure 4.7, compared with a less successful combination in figure 4.8. Given the theory behind this method, the choice of parameters used to generate figure 4.7 seems to be very generous in terms of what it will consider to be a ridge. The maximum allowed size of the along-ridge second derivative is over twice as big as the minimum size of the across-ridge component. This means that isolated peaks and deep cols will be identified by the method. That these parameters are necessary to capture the contiguous nature of the mountain ranges reflects the fact that, at the scale of the grid, there is considerable variation of height along the length of a range.

By setting a more stringent test to generate figure 4.8, a large number of points on well-defined ranges have been left out, with the result that even the most striking ranges (the Andes and Rockies), appear as rather uneven and disjointed collections of points. However, the edges of the Greenland and Antarctic ice sheets are captured well in both figures, perhaps due to the smooth profile of these features.

The behaviour of the method over rough upland areas, as distinct from the relatively smooth ice-domes of Greenland and Antarctica, may be less promising. Over the
Figure 4.7: Peak mask generated by the method of perpendicular second derivatives, with $\varepsilon_1 = 1000 \, \text{m}^{-1}$ and $\varepsilon_2 = -400 \, \text{m}^{-1}$. The dotted line indicates the land-sea boundary. The axes are labelled in grid-points.

Figure 4.8: Peak mask generated by the method of perpendicular second derivatives, with $\varepsilon_1 = 200 \, \text{m}^{-1}$ and $\varepsilon_2 = -400 \, \text{m}^{-1}$. The dotted line indicates the land-sea boundary. The axes are labelled in grid-points.
Chapter 4 Applying the variational method to real orography

Figure 4.9: Peak mask generated by selecting all points above a given threshold (1000 m, in this case). The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.

Tibetan Plateau and the mountains of Turkey and Iran, the method picks up large regions of points, which, if used in the variational method, might adversely affect the smoothing in that area.

As a procedure for identifying the ridges to be constrained, the method of perpendicular second derivatives has some merits. However, it generates quite noisy fields, and could result in a maintaining of roughness over upland areas. This may be a result of using the unsmoothed mean orography as input; better results might be obtained starting from previously smoothed orography.

4.4.3 Height threshold

One very simple way to identify mountainous areas for constraint in the variational method would be to select all points over a certain height. An example of this method in use is given in figure 4.9.

It is worth noting that many areas of orography above 1000 m in elevation are not identified by this method. This is because the mean orography itself has lost some of
its height through the process of areal averaging which was used in its creation. So, some smaller upland areas are smoothed by the averaging to the extent that their maximum height is reduced below 1000 m. This is why there is no contribution from the Urals, for instance, whose actual height reaches 1894 m.

The main advantage of this method is its simplicity, but it does have several disadvantages. Firstly, it is not obvious that the importance of a given orographic ridge to atmospheric dynamics depends on its height alone. Relatively low ridges which are surrounded by flat land may still be important. Secondly, as with the method of perpendicular second derivatives, large upland areas are flagged in their entirety by this method, and this could lead to a reduction in the effectiveness of the variational smoothing across these regions.

4.4.4 Ridge searches

The examples above suggest that a more sophisticated method of finding orographic ridges is needed. A suggestion of what that method might be is given in this section. At first glance, the method described in the first part below is merely another method for determining whether a given point should be constrained by looking at the relationship of that point to its surroundings. However, the development of this idea in the second part extends the method to consider ridges as complete entities, rather than tackling the problem on a point-by-point basis.

4.4.4.1 Simple ridge search

The method for determining whether a point \((i, j)\) should be tagged as a ridge point may be summarized as follows:

- Calculate average height \(h_{av}\) of orography in area surrounding \((i, j)\).

- If point \((i, j)\) and at least one other adjacent point is higher than \(\tau h_{av}\) then tag all such points as ridge points (\(\tau\) is a parameter), but only if point \((i, j)\) is higher than some minimum height \(h_0\).
Figure 4.10: Possible sets of points (coloured) adjacent to the central point (black). The grey points are not considered adjacent, and the numbers and different colours indicate the degree of the search.

So we can characterize this as a search of the immediately adjacent points to see whether the point under consideration is a ridge point, but we only begin the search if the starting point is above some minimum height.

Which points should be counted as adjacent in this case? A narrow ridge running neither parallel nor at 45° to the grid will not directly intersect two points which are immediately adjacent to each other (see figure 4.4 for a reminder). Therefore, to include other possibilities, it was decided to allow the possibility of searching points further away from the point in question. Possible adjacent points now include any point which does not have another point directly between it and the starting point, within a $9 \times 9$ square. These adjacent points are shown in figure 4.10, and are divided into groups of points at different distances from the centre. The range of points searched is termed the degree of the search; for instance, if the points labelled 1, 2, 3 and 4 are searched, then this search has degree 4. Also, the area over which the average height $h_{av}$ is calculated depends on the degree: $h_{av}$ is the average height of all the points contained in a square which just encloses the points being searched. So, for a search of degree 1, the average is evaluated over a $3 \times 3$ square, while for a search of degree 5, it is evaluated over a $9 \times 9$ square. For simplicity, the search does not cross the edges of the domain at the poles. So, there are three parameters
in this scheme: the minimum height ($h_0$), the threshold scale ($\tau$), and the degree of search.

Once again, the method was tried with many different combinations of these parameters, and some of the results of this are shown in figures 4.11–4.15. These plots illustrate the wide range of outcomes possible with this method. They all show some success – in all five figures, the Andes and Rockies are clearly identified. However, there is a fair amount of noise in some of the plots, caused by the identification of very short low-lying ridges. The application of a minimum height $h_0$ helps to alleviate this, and figure 4.14 is particularly effective at capturing recognizable mountain ranges.

As remarked above, there is a problem with the spurious identification of very short ridges in this method. It would be much better if it were possible to take into account the length of an orographic feature in deciding what to include in the mask. This possibility is addressed in the next section.
Figure 4.12: Peak mask generated by simple ridge search of degree 4, with $\tau = 1.2$ and $h_0 = 0 \text{ m}$. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.

Figure 4.13: Peak mask generated by simple ridge search of degree 4, with $\tau = 1.4$ and $h_0 = 0 \text{ m}$. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.
Figure 4.14: Peak mask generated by simple ridge search of degree 4, with $\tau = 1.4$ and $h_0 = 500$ m. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.

Figure 4.15: Peak mask generated by simple ridge search of degree 6, with $\tau = 1.8$ and $h_0 = 500$ m. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.
4.4.4.2 Recursive ridge search

In order to select points for inclusion in the peak mask on the basis of the length of the feature being examined, it is necessary, on a computational level, to record the complete extent of the ridge before determining whether to include it. One way to do this is to employ a recursive search method.

A recursive method is one whose outcome may depend on a further application of itself. This is explained in detail in, for instance, Nyhoff and Leestma (1997); in computational terms, a recursive method is implemented as a subroutine which is allowed to call itself. This idea is relevant to the problem of finding ridges because of the tree-like structure of the features we wish to identify. The question can be posed in ordinary language; at each point in the domain, in turn, we ask:

1. Is the current point significantly higher than the average of its surroundings? Has it yet to be marked as part of the ridge? If the answer to both of these is yes, then continue; otherwise, stop.

2. Mark this point as part of a ridge

3. Apply this algorithm to each of the neighbouring ridge points in turn

When all the instances of the algorithm have returned we are left with a set of the points in a ridge. We can then choose whether to accept these into the peak mask depending on the number of points in the ridge.

As with the simple ridge search, the average of surrounding points is calculated for each point considered, and this calculation is made for the square set of points which just encloses the adjacent points. Also, adjacent points are specified in the same manner as for the simple ridge search, with the degree of the search being specified. The threshold scale $\tau$ is again the factor that determines the extent to which a point has to be higher than its surroundings to be considered. In addition, the minimum height condition $h_0$ may be applied to the search. So, there is just one extra parameter in this method, compared with the simple ridge search, and that is the minimum ‘length’ of the ridge, $L_0$, expressed in grid-points. This is the minimum total number of points necessary for the ridge to be included in the mask.
The recursive ridge search method was tested for a range of different parameter combinations, and the results for some of these are shown in figures 4.16–4.19. In all four plots, the Andes are identified very clearly; other mountain ranges are identified with varying degrees of success. Among the ranges captured in some of the four plots are the Rockies, the Appalachians, the Scandinavian range, the Urals, the Zagros Mountains (in the Middle East), the southern and east African highlands, and the Tibetan plateau. Compared to the simple ridge search, this method produces less-noisy masks, especially when a minimum height is specified.

One potential problem with this method concerns the treatment of the grid geometry. By specifying the minimum length of features to be included in terms of grid-points, there is an implicit assumption being made that the grid consists of equally-spaced points on a plane. However, near the poles, the grid-points are spaced much more closely together longitudinally than they are at the equator, and much more closely than they are latitudinally everywhere on the globe. The consequence of this is that if a ridge is aligned zonally, and located near a pole, the length it is required to be for inclusion (in metres) is less than if the ridge was located at the equator, or oriented meridionally.

### 4.4.5 Choosing a ridge-finding method

The choice of a ridge-finding method for use with the variational smoothing method is not easy to make in an objective way. The methods described above apply certain rules consistently to obtain a mask of peaks and ridges, but it is not immediately obvious which set of rules should be used. In choosing the mask to be used in the experiments described below, a few simple ideas were applied. In general terms, the following properties were required:

- The taller and longer a ridge is, the more important it is to include it.

- Noisy masks are bad, because they impede the smoothing process at the grid-scale.

- An included ridge should only have its summit points identified, again because constraining a broad ridge will have a negative impact upon the effectiveness of the smoothing.
Figure 4.16: Peak mask generated by recursive ridge search of degree 1, with $\tau = 1.2$, $L_0 = 5$, and $h_0 = 0\,\text{m}$. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.

Figure 4.17: Peak mask generated by recursive ridge search of degree 2, with $\tau = 1.3$, $L_0 = 10$ and $h_0 = 300\,\text{m}$. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.
Figure 4.18: Peak mask generated by recursive ridge search of degree 6, with $\tau = 1.6$, $L_0 = 10$ and $h_0 = 0 \text{ m}$. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.

Figure 4.19: Peak mask generated by recursive ridge search of degree 6, with $\tau = 1.4$, $L_0 = 10$ and $h_0 = 500 \text{ m}$. The dotted line indicates the land-sea boundary, and the axes are labelled in grid-points.
Where possible, the peak mask should include those mountain ranges which are known to be important climatologically. For instance, the Andes are well-known for their role blocking the zonal flow in the southern hemisphere, while the East African mountains are important for the development of the Monsoon.

Consideration of these criteria led to the decision to use the peak mask shown in figure 4.17, which was a search of degree 2, with $\tau = 1.3$, $L_0 = 10$ and $h_0 = 300$ m. Undoubtedly, there is considerable room for further work here. The parameter space of this method alone is large, and given that the method itself could be made more sophisticated, the possibilities are very great indeed. However, given the practical limits to the investigation possible in this area, the chosen mask represents a reasoned choice, which should allow the application of constrained smoothing, as described earlier, to be investigated.

### 4.5 The variational smoothing method in practice

The theory behind the variational smoothing method has been discussed in the preceding chapter, and various general aspects of its practical application have been examined in the earlier sections of this one. Now, we turn to the specific application of the method to the real orography described in section 4.3. Of particular interest is the verification of spectral response predicted by the analysis in chapter 3, and the effect on that spectral response of the imposition of extra constraints. As described in chapter 3, the variational method can be used with two different grid geometries: ‘planar’ and ‘spherical’. Both the examples below and the subsequent testing of the orography in the shallow water model concentrate on using the method with spherical geometry. However, because of its comparative simplicity, and because it has less relevance to the problem of smoothing real global orography in the sphere, the planar method is also examined in the first part below.
4.5.1 ‘Planar’ geometry

4.5.1.1 Smoothing coefficients for ‘planar’ geometry

In the case of ‘planar’ geometry, the domain is considered to be a set of grid-points equally-spaced on a plane. Since, for the Gaussian grids we are considering, \( N_x = 2N_y \), this means that the plane is twice as big in the \( x \) direction as it is in the \( y \) direction. However, the grid-length used is not of central importance to the way the smoothing is specified. The spectral response of the variational filter is given in (3.40), where \( K^2 \) and \( L^2 \) depend on the grid-length and are defined by (3.38) and (3.39). Thus, the coefficients of (3.40) are not non-dimensional, and the spectral response of the filter, for a given set of coefficients, depends on the grid-length specified. To counter this, and thus enable the specification of coefficients which result in the same spectral response, regardless of grid-length, it is necessary to specify the coefficients as products of a non-dimensional part, and a part depending on the grid-length. In these coefficients, the non-dimensional part is distinguished by a hat. For instance,

\[
\alpha_n = \hat{\alpha}_n \delta^{2n},
\]

(4.4)

where \( \delta \) is the grid-length. The coefficients \( \hat{\alpha}_0 \ldots \hat{\alpha}_n \) and \( \hat{\beta}_1 \ldots \hat{\beta}_n \) correspond with \( \alpha_0 \ldots \alpha_n \) and \( \beta_1 \ldots \beta_n \) in this way.

4.5.1.2 A test case

In order to verify the spectral response of the planar variational smoothing method, a simple test case is used. The coefficients used are as follows:

\[
\begin{align*}
\hat{\alpha}_0 &= 1 \\
\hat{\alpha}_n &= 0 \text{ for } n > 0 \\
\hat{\beta}_1 &= 1 \\
\hat{\beta}_n &= 0 \text{ for } n > 1.
\end{align*}
\]

(4.5)

In the whole of this section, the variational method is applied using the Conjugate Gradient (CG) method, as described in chapter 3, with a convergence parameter of \( 10^{-6} \). In this case, the minimization was completed in 9 CG iterations, and the resulting orography is shown in figure 4.20.
Comparing with figure 4.2, it is clear that smoothing has taken place. The spectral response is calculated, by taking the ratio of the Fourier transforms of the filtered and unfiltered fields, as a function of the total wavenumber $|\mathbf{K}|$, where $|\mathbf{K}|^2 = K^2 + L^2$, and shown in figure 4.21. However, as in chapter 3, it is plotted against non-dimensional wavenumber $\mu = \frac{2\pi|\mathbf{K}|}{L}$, where $L$ is the length of the domain in the $x$-direction. The calculation of the spectrum on a rectangular grid (as distinct from calculating the spherical spectrum described earlier) is not straightforward, since the maximum resolved wavelength in the $y$-direction is half that in the $x$-direction. To counter this, the spectra are calculated on a square grid, of dimensions $N_x \times 2N_y$. The part of the grid $0 \leq j < N_y$ is occupied by the orography as usual, while the remainder of the grid is occupied by a copy of the orography arranged so that an over-the-north-pole boundary condition is fulfilled. The spectrum of this square domain is the one that is calculated.

Figure 4.21 shows that the agreement between the predicted and measured spectral responses is very good, and verifies the analysis of the variational method given in chapter 3.
Figure 4.21: Comparison between the predicted and actual responses of the planar variational filter, at $200 \times 100$ resolution, with the smoothing coefficients given in 4.5.

### 4.5.1.3 Flat sea and barrier height constraints

As expected, the orography in figure 4.20 has large areas of elevated sea points. To test the effectiveness of the flat sea constraint, the variational method is here applied with the same coefficients, but with the extra constraint that the sea must remain flat. Again, the minimization was completed in 9 iterations. The resulting orography is shown in figure 4.22. Comparison with figure 4.2 shows that the orography has been smoothed, and the flat sea maintained.

Similarly, the operation of the barrier-height constraint was tested. Figure 4.23 shows the result of applying the variational method with the same coefficients, while constraining the sea to be flat, and applying the barrier height constraint with $\gamma = 20$. The peak mask used is that shown in figure 4.17, and the minimization was completed in 15 iterations. Comparison between the two figures 4.22 and 4.23 shows a clear effect from the additional constraint. Both figures are generally much smoother than the original mean orography of figure 4.2, but the height of several regions has been clearly enhanced in figure 4.23 by the barrier height constraint. Note particularly the additional height of the Andes, the Antarctic peninsula, and the East African highlands.
Chapter 4 Applying the variational method to real orography

Figure 4.22: Orography resulting from applying planar variational filter at $200 \times 100$ resolution, with the smoothing coefficients given in 4.5, and with the flat sea constraint imposed. The contour interval is 250 m, the unmarked contour is 0 m, and the axes are labelled in grid-points.

To enable better comparison between the smoothed orographies, two further figures are provided. In both cases, the comparison is also made with the result of applying the variational method with the barrier-height constraint, but not with the flat sea constraint. For this, $\gamma = 20$, and convergence was achieved in 16 iterations. Figure 4.24 shows the spectral responses of all four parameter combinations, while figure 4.25 shows a cross-section through the Andes for these orographies, and for the mean orography. From the second figure (4.25), it is clear that the constrained smoothing is working as expected. The orography resulting from unconstrained smoothing (denoted ‘smoothing only’) shows elevated sea-points, and reduced barrier heights. The orography resulting from the three types of constrained smoothing shows that the height of the barrier can be restored, and the flatness of the sea maintained, while still undertaking the smoothing on unconstrained points. This may be seen most clearly between $60^\circ$W and $50^\circ$W, where all the smoothed orographies are considerably smoother than the mean orography. However, the first figure (4.24) shows the consequences of the constrained smoothing for the spectral properties of the filter. Applying only the flat sea constraint does not affect the spectral
Figure 4.23: Orography resulting from applying planar variational filter at $200 \times 100$ resolution, with the smoothing coefficients given in 4.5, and with both the flat sea constraint and the barrier-height constraint imposed. For the barrier-height constraint, the peak mask is shown in figure 4.17, and $\gamma = 20$. The contour interval is $250 \text{m}$, the unmarked contour is $0 \text{m}$, and the axes are labelled in grid-points.

Figure 4.24: Spectral responses of the four versions of the planar variational method tested. Details in the text.
Figure 4.25: Cross-sections through the Andes at $22^\circ$ S, showing the mean orography and the outcomes of the four tested parameter combinations. The green lines indicate the location of the four points included in the peak mask, and thus constrained by the barrier-height constraint.

response very significantly. The loss of attenuation is only a small fraction of the total attenuation, and this is true at all wavenumbers, including the grid-scale. However, applying the barrier-height constraint considerably reduces the attenuation of the filter at all wavelengths. This effect is particularly bad at large wavenumbers, which is understandable when the effect on the cross-section of these constraints is taken into account. The resolution of this orography is such that major features such as the Andes are only one or two grid-points wide, so in reinforcing them, the near-grid-scale components of the orography are bound to be enhanced.

4.5.1.4 Approximation to the Raymond filter

One of the key properties of the variational method is its ability to imitate any filter whose spectral response function may be written as a Padé Approximant. This enables the use of the filter by Raymond (1988), which has been employed by the UK Met Office in their Unified Model (see section 3.2.5.2), but with the extra constraints applied.
Figure 4.26: Spectral response of the variational version of the Raymond filter, compared with the predicted response, and the actual Raymond filter response. Also shown is the theoretical response of the 1-D Raymond filter.

Before examining the effectiveness of the variational version of the Raymond filter with spherical geometry, it is tested here with planar geometry. The coefficients used are those given in 3.33, which may be rewritten in non-dimensional form as

\[
\begin{align*}
\hat{\alpha}_0 &= 1 \\
\hat{\alpha}_n &= 0 \text{ for } n \neq 0 \\
\hat{\beta}_1 &= 0 \\
\hat{\beta}_2 &= 0 \\
\hat{\beta}_3 &= \frac{1}{64} \varepsilon \\
\hat{\beta}_4 &= \frac{1}{320} \varepsilon \\
\hat{\beta}_5 &= \frac{1}{256} \varepsilon \\
\hat{\beta}_6 &= \frac{3}{1024} \varepsilon \\
\hat{\beta}_n &= 0 \text{ for } n > 6, \\
\end{align*}
\]

(4.6)

where \( \varepsilon \) is the Raymond filter parameter. The variational method with these coefficients was applied to the mean orography. The spectrum of the resulting field is plotted in figure 4.26, where it is compared with the predicted response and the Raymond filter response for \( \varepsilon = 1 \). Also shown is the theoretical 1-D response of the Raymond filter, as given by equation (3.9). The actual spectral response of the
variational version of the Raymond filter agrees well with the prediction, and can also been seen to be a good approximation to the Raymond filter itself. The discrepancy between these curves and the theoretical response of the Raymond filter may be explained by the fact that the Raymond filter is implemented in 2-D by applying the filter twice: once zonally and once meridionally. Thus, the 2-D response function of the Raymond filter is actually

\begin{equation}
F(k, l) = \frac{1}{(1 + \varepsilon \tan^6 \left( \frac{k \Delta x}{2} \right))(1 + \varepsilon \tan^6 \left( \frac{l \Delta y}{2} \right))}.
\end{equation}

This test is sufficient to determine that the variational method can effectively imitate the spectral response of the Raymond filter. The application of the other constraints within the context of this filter is considered in the next section, where the method is applied on spherical geometry.

### 4.5.2 ‘Spherical’ geometry

In this section, the variational smoothing method is applied to the mean orography with a range of parameter combinations. The emphasis here is on the effect of these parameter combinations upon the properties of the orography, and the technical aspects of using the method. In all cases, Padé coefficients which imitate the Raymond filter have been used, with \( \varepsilon = 1.0 \).

#### 4.5.2.1 Technical note

In the previous section, covering the use of the variational method with planar geometry, the method was applied without the use of a preconditioner, and with the same CG convergence parameter in all cases. However, for spherical geometry, this was found not to be successful. Owing to the ill-conditioned nature of the CG matrix \( A \) in the case of spherical geometry (see section 3.3.3.2), it was found necessary to use a diagonal preconditioner to enable convergence to happen. This was true of all examples presented below. The diagonal preconditioner is, however, not an ideal preconditioner, and the degree of convergence which was possible varied, depending on the parameter combination in use. In the specifications of smoothed orographies in the following section, the values of the convergence parameter and the number of iterations required for convergence are given in each case.
4.5.2.2 The Raymond filter on the sphere

Before considering the effects of extra constraints on the smoothed orography, the effectiveness of the variational method at imitating the Raymond filter on the sphere needs to be assessed.

The coefficients used in this are those given in (4.6), with the grid-length $\delta = 2\pi/N_x$. The reason for this is that in the spherical variational method, the points are taken to be located on a sphere of radius $a = 1$ m. This means that for a regular latitude-longitude grid (which the Gaussian grid almost is), the meridional grid-length is $\pi/N_y (= 2\pi/N_x)$, and at the equator the zonal grid-length is also $2\pi/N_x$.

As in the previous section, the variational method has been applied to the mean orography with $\varepsilon = 1$, and the spectral response calculated. The convergence parameter was $10^{-6}$, and the conjugate gradient method converged after 1228 iterations. Note that this a considerably larger number of iterations than was required for planar geometry, and, when the use of the diagonal preconditioner is taken into account, emphasizes the ill-conditioned nature of the problem. The measured spectral response of the variational filter in this case is shown in figure 4.27, where it is compared with the theoretical response of the filter. Also shown in figure 4.27 is the measured spectral response of the actual Raymond filter in a version for spherical geometry. Webster et al. (2003) used this version of the Raymond filter which takes into account the reduction of zonal grid-length with increasing latitude by varying the value of $\varepsilon$ with latitude.

The spectra used to calculate the response functions shown in figure 4.27, and in subsequent plots concerned with the spherical version of the variational filter, were calculated by decomposing the orography into spherical harmonics using Fourier and Legendre transforms. The spectra were the calculated according to equation (4.3).

The spectral response of the spherical variational version of the Raymond filter is not very similar to the form of the response calculated from the coefficients used. This is for the reason given in chapter 3, namely that the eigenfunctions of the discretized Laplacian operator on the sphere are not the same as the eigenfunctions of the continuous operator. The difference between these two sets of eigenfunctions is greatest at the smallest scales, and so this is the reason why the discrepancy
Figure 4.27: Spectral response of the spherical variational Raymond filter, compared with the Padé approximant to the response function, and the response of the spherical version of the Raymond filter.

between the two spectral response functions is greatest there. However, as pointed out in chapter 3, it may well be that the eigenfunctions of the discretized operator are more relevant for the problems associated with near-grid-scale orography in global grid-point models than the spherical harmonics, so this is not necessarily a bad thing. Also of significance is the spectral response of the spherical Raymond filter used by Webster et al. (2003). In terms of spherical harmonics, the attenuation of high-wavenumber components is considerably less than that accomplished by the variational filter. This is because the spherical Raymond filter is not designed to filter in terms of spherical harmonics, but to filter the perpendicular Fourier components of the orography, taking into account the local grid-length on the sphere. Figure 4.27 shows clearly that these two approaches are not equivalent.

4.5.2.3 The effect of the barrier-height constraint

Having established the spectral response of the variational version of the Raymond filter, this section looks at the effect upon the orography of applying the other
barrier-height constraint. The flat-sea constraint was tested in the preceding section, and will not be considered again here.

For this part of the investigation, and for the subsequent experiments with the shallow water model, a large set of model orographies were generated. These were the result of applying the variational filter, with the Raymond filter coefficients, and with different values of $\gamma$. The peak mask used remains the same as previously, and is shown in figure 4.17.

The different values of $\gamma$ used are listed in table 4.1. Note that, in all cases, the diagonal preconditioner was used. The values of the convergence parameter listed in the table above were determined by trial and error as the minimum necessary for the convergence of the conjugate gradient method to succeed. Their variation, and the variation in the number of iterations necessary for convergence, gives some indication of the complex properties of the problem being solved.

In order to examine the effects of applying this constraint, various bulk properties of the orography have been calculated for each value of $\gamma$. In addition, the variation with $\gamma$ of the heights of a number of individual points in the domain have been collated. These are shown in figures 4.28–4.34. Note that, where a quantity is plotted against $\gamma$, the scale of the horizontal axis varies between figures.

Figure 4.28 shows the way the spectral response varies with $\gamma$. It is clear that as $\gamma$ increases, the spectral response converges towards a particular form. At almost all wavenumbers this response is greater than the response for $\gamma = 0$, so that the barrier-height constraint inevitably brings an amplification of lower wavenumbers compared to the mean orography. This is the same as the behaviour seen in the simple, planar case, and shown in figure 4.24. However, the convergence to this spectral response function is not uniform for all wavenumbers; the lower wavenumbers converge faster than the higher wavenumbers to the extent that there is still a noticeable difference in the spectral response at high wavenumbers between $\gamma = 1000$ and $\gamma = 5000$. Despite this behaviour, it is important to remember two things about these spectral responses. Firstly, the application of the barrier-height constraint renders the variational filter non-linear. Thus, the spectral response of the filter will depend on the orography to which it is applied. Secondly, the non-ideal nature of the way the Conjugate Gradient method has been applied means that calculated
<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>convergence parameter</th>
<th>number of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>$1.0 \times 10^{-6}$</td>
<td>1228</td>
</tr>
<tr>
<td>1.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>958</td>
</tr>
<tr>
<td>2.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>1450</td>
</tr>
<tr>
<td>3.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>1533</td>
</tr>
<tr>
<td>5.5</td>
<td>$1.5 \times 10^{-6}$</td>
<td>455</td>
</tr>
<tr>
<td>8.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>500</td>
</tr>
<tr>
<td>10.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>571</td>
</tr>
<tr>
<td>15.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>562</td>
</tr>
<tr>
<td>20.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>486</td>
</tr>
<tr>
<td>25.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>632</td>
</tr>
<tr>
<td>30.0</td>
<td>$1.5 \times 10^{-6}$</td>
<td>540</td>
</tr>
<tr>
<td>40.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>406</td>
</tr>
<tr>
<td>60.0</td>
<td>$1.5 \times 10^{-6}$</td>
<td>365</td>
</tr>
<tr>
<td>80.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>572</td>
</tr>
<tr>
<td>100.0</td>
<td>$1.5 \times 10^{-6}$</td>
<td>336</td>
</tr>
<tr>
<td>150.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>705</td>
</tr>
<tr>
<td>200.0</td>
<td>$1.5 \times 10^{-6}$</td>
<td>414</td>
</tr>
<tr>
<td>300.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>580</td>
</tr>
<tr>
<td>500.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>563</td>
</tr>
<tr>
<td>1000.0</td>
<td>$1.8 \times 10^{-6}$</td>
<td>447</td>
</tr>
<tr>
<td>2000.0</td>
<td>$1.8 \times 10^{-6}$</td>
<td>592</td>
</tr>
<tr>
<td>5000.0</td>
<td>$5.0 \times 10^{-6}$</td>
<td>539</td>
</tr>
</tbody>
</table>

Table 4.1: Convergence parameters used in the generation of orography for $\gamma$ sensitivity experiments.
Figure 4.28: Spectral response of the variational filter with different values of $\gamma$.

Figure 4.29: Variation of the height of point (63,38) with $\gamma$. This is a constrained point, situated at 22° S, and the dashed line shows the height of this point in the mean orography.
Figure 4.30: Variation of the height of point (62,38) with $\gamma$. This is an unconstrained point, adjacent to the constrained point in figure 4.29, situated at 22° S. The dashed line shows the height of this point in the mean orography.

Figure 4.31: Variation of the height of point (56,79) with $\gamma$. This is an unconstrained point close, but not adjacent, to constrained points in North America, at 51° N. The dashed line shows the height of this point in the mean orography.
Figure 4.32: Variation of the height of point (133,86) with $\gamma$. This is a constrained point at high latitude (64° N), and the dashed line shows the height of this point in the mean orography.

Figure 4.33: Variation of the volume of the orography with $\gamma$. The dashed line shows the value for the mean orography.
Chapter 4 Applying the variational method to real orography

Figure 4.34: The total variation of the orography plotted against $\gamma$. The dashed line shows the value for the mean orography.

spectra may differ slightly from those that would be obtained if the minimization had been perfectly converged in all cases.

Figures 4.29–4.32 show the way the heights of various individual points vary with $\gamma$. These particular points were chosen because they illustrate different types of behaviour seen in different classes of points. The point shown in figure 4.29 is a constrained point at quite a low latitude, and as the value of $\gamma$ increases, so its height smoothly approaches the height of that point in the mean orography. This smooth behaviour indicates that this point is well-converged, and the clear approach to the mean orography value shows that the barrier-height constraint is working properly here.

Figure 4.30 shows the behaviour of a point which is not constrained, but is situated adjacent to one that is. It seems to illustrates the changing balance between the parts of the cost function concerned with linear spectral filtering (those containing parameters $\alpha_0 \ldots \alpha_n$ and $\beta_1 \ldots \beta_n$) and the term concerned with the barrier-height constraint (containing $\gamma$). A more detailed explanation is not known at present, but it could be due to the changing amplitude of Gibbs ripples in the smoothed orography.
Both the points so far examined are located quite close to the equator. However, the points whose behaviour is shown in figures 4.31–4.32 are located in higher-latitude regions. The first of these is an unconstrained point, not placed adjacent to a constrained point, but quite close to one. The height of this point does not evolve smoothly with $\gamma$, but shows an erratic development, especially for $25 \leq \gamma \leq 300$. Comparison with the table on page 106 shows that, in this region, the variation in height is correlated with the variation in convergence parameter. The values of $\gamma$ which resulted in the greater heights were also those where the convergence parameter used was larger. If these points ($\gamma = 30, 60, 100$ and $200$) were to be removed, then the evolution of the height of this point would be much smoother. This illustrates the extent to which unconstrained points at this latitude are not necessarily well-converged. By contrast, small changes to the value of the convergence parameter have no noticeable effect on the height of points near the equator. The reason for this concerns the eigenfunctions and eigenvalues of the $\nabla^2$ operator. The convergence of the Conjugate Gradient method is hampered by the poor clustering of the eigenvalues, which is particularly due to the very large eigenvalues associated with particular eigenfunctions; these eigenfunctions generally have greatest magnitude in the polar regions. These components of the field are least likely to converge fully because of the large magnitudes of their eigenvalues, and consequently the resulting orography will be most sensitive to the convergence parameter in the higher latitudes. Similar behaviour is seen in figure 4.32, but in this case it is largely masked by the effect of the barrier-height constraint, which forces the height of the point ever closer to that in the mean orography.

Figure 4.33 shows the variation in orographic volume with $\gamma$. This is the same as the coefficient of the 0th wavenumber spherical harmonic multiplied by the area of the surface of the Earth. The volume remains unchanged from the mean orography when $\gamma = 0$, but converges rapidly to a new value as $\gamma$ increases. This confirms some of the conclusions reached from figure 4.28.

The final plot (figure 4.34) shows the evolution of the total variation of the orography as $\gamma$ increases. Total variation ($TV$) is defined as

$$TV = \frac{1}{2N_xN_y} \sum_{ij} (|h_{i,j} - h_{i+1,j}| + |h_{i,j} - h_{i,j+1}|),$$

so that it is the average of the magnitudes of all the differences between immediately
adjacent points. Intuitively, we would expect this quantity to be lower for smoothed orography than for rough orography. This expectation is confirmed by the fact that the total variation when $\gamma = 0$ is considerably less than it is for the mean orography. Then, as $\gamma$ is increased, the roughness increases, and thus the total variation does too. The total variation reaches the value for the mean orography at $\gamma \approx 300$, indicating that, by this measure, the smoothing and barrier-height parts of the cost function balance each other at this point.

### 4.6 Summary and conclusions

The new variational smoothing scheme, outlined in chapter 3, was presented there in on a largely theoretical basis. This chapter, in contrast, has shown that the method can be implemented successfully, and it has the properties expected. In applying the method to real orography, much consideration has been given to the choice of peak mask to use when constraining the height of ridge points. However, although the methods explored here have led to a reasoned choice for that mask, much more work could yet be done in this area.

Having first covered the generation of the mean orography, and discussed the choice of a grid of the spectral model, the variational method was thoroughly tested, and its spectral response characterized. The main conclusion was that the possibility of emulating a class of linear filters, specified by Padé Approximants, is realized as expected. This makes the use of the variational method to refine a linear smoother a practical possibility.

The other significant feature of the method’s spectral response is that, with the additional constraints in place, there is still significant attenuation of near-grid-scale components. This means that the imposition of a flat sea, or the maintenance of barrier height does not undo the work of smoothing the orography performed by the other terms in the cost function. We might, therefore, hope that the variational method will be useful in an NWP context, since it seems to offer exactly the characteristics required at the outset.

Having explained and tested the variational scheme, the next chapter evaluates it in the context of a simple global model.
Chapter 5

Experiments with a global shallow water model

5.1 Introduction

5.1.1 Background

In the preceding two chapters, a new variational smoothing method for use with model orography has been developed and tested. It has been demonstrated that the method is successful in smoothing the orographic field, while also allowing for the possibility of maintaining the flatness of the sea and the height of important ridges. The expectation is that this will have benefits for NWP and climate prediction applications. This chapter describes some experiments done with a global shallow water model to determine the effect of the smoothed orography on the representation of the flow. In particular, the experiments are designed to determine whether the smoothing does have a positive effect on the accuracy of the flow, and to investigate the relative usefulness of various parameter combinations.

The chapter begins with a discussion of the aims of the experiments, and the hypotheses they are designed to test. After outlining the rationale behind the basic experimental procedure, the numerical model used is described, followed by a detailed explanation of the experimental method. Finally, the results are given, and discussed in the light of the initial aims and hypotheses.
5.1.2 Aims and hypotheses

The hypotheses underpinning all the work in this thesis may be summarized in three points:

1. Problems in NWP and climate models concerned with near-grid-scale orography may be mitigated by filtering the orography to remove the smallest scales;

2. The negative impact that orographic smoothing has on large-scale aspects of the flow may be alleviated by constraining the smoothing process;

3. The constrained smoothed orography will result in more accurate model integrations than the smoothed orography alone, but without the original near-grid-scale problems being manifest.

From the start, and for reasons explained in chapter 1, the approach has been to investigate these questions with simpler models than a full NWP model. The work presented in chapter 2 led to the conclusion that a linear, one-dimensional model was insufficiently representative of NWP models to be useful in investigating these hypotheses, and was the reason for choosing the non-linear, global shallow water model described below. The spectral model was chosen because it was readily available, simply constructed, and well-understood.

Nevertheless, it remains the case that the major problems resulting from the presence of near-grid-scale orography are not simply connected with the dynamics of the model. Grid-point storms, for instance, are thought to be produced by the interactions between the resolved flow, the vertical grid definition, the moisture parameterization, and the orographic field. It is clearly over-optimistic to hope that they will be reproduced in a simple model such as this one. This means that the questions to be asked have to focus on the resolved flow and the effect of smoothing the orography upon the representation of that flow in the model.

One of the main questions which arises out of the hypotheses listed above concerns the best filter to use in smoothing the orography. In the light of work already done on this question, and the potential range of possible answers, certain assumptions
are made for the purposes of the present work. As already explained, the study of Davies and Brown (2001) provided a rationale for the filtering of orography, and further work by Webster et al. (2003) identified the Raymond filter (Raymond 1988) as a suitable orographic smoothing filter. The starting point of the experiment described below is, therefore, that the Raymond filter is a suitable filter for orographic smoothing in this context, with $\varepsilon = 1.0$. The spectral response of the Raymond filter is given in equation (3.9).

Finally, the hypotheses given above contain the assumption that the accurate representation of at least some aspects of the flow will be hindered by orographic smoothing, and that the further constraints introduced in chapter 3 will help to counteract this. With this in mind, the specific questions to be answered by this experiment were formulated:

- Given the use of the Raymond filter, is there an optimal combination of values of $\varepsilon$ and $\gamma$ in the variational scheme, which minimizes the errors in the flow?
- Where there are errors, can we identify which aspects of the flow are adversely affected, and quantify that effect?

These questions have been formulated so that it is tractable to answer them with a series of closely-linked experiments; it is important, though, not to lose sight of the broader context of the questions, which comprises not only reasonably established understanding, but also a large number of unresolved problems. These unresolved problems are beyond the scope of the present work to address, but possible ways of tackling them are discussed in chapter 6.

5.2 The model

5.2.1 Description

The model used in the experiment is a global spectral shallow water model developed at the University of Reading. The principles underlying the construction of spectral
models are discussed in, for instance, Durran (1999); spectral shallow water models in particular are covered in Williamson and Laprise (2000).

The governing equations of the model are framed in terms of vorticity $\zeta$, divergence $\delta$ and fluid depth perturbation $h^*$. (The fluid depth perturbation is related to the perturbation variables used in chapter 2 by the relationship $h^* = \eta - h_0$.) The vorticity and divergence are defined thus:

$$\zeta \equiv \mathbf{k} \cdot (\nabla \times \mathbf{v}),$$

$$\delta \equiv \nabla \cdot \mathbf{v},$$

where $\mathbf{k}$ is the unit vector perpendicular to the Earth’s surface, and $\mathbf{v}$ is the velocity vector. The shallow water equations in spherical polar coordinates are:

$$\frac{\partial \zeta}{\partial t} = -\frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} [(\zeta + f)u] - \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} [(\zeta + f)v \cos \phi],$$

$$\frac{\partial \delta}{\partial t} = \frac{1}{a \cos \phi} \frac{\partial}{\partial \lambda} [(\zeta + f)v] - \frac{1}{a \cos \phi} \frac{\partial}{\partial \phi} [(\zeta + f)u \cos \phi] - \nabla^2 \left[ gh^* + \frac{1}{2}(u^2 + v^2) \right],$$

$$\frac{\partial h^*}{\partial t} = -\mathbf{v} \cdot \nabla h^* - \frac{h^* + H}{a \cos \phi} \left( \frac{\partial u}{\partial \lambda} + \frac{\partial v \cos \phi}{\partial \phi} \right).$$

As before, $a$ is the radius of the Earth, $f$ the Coriolis parameter, $\lambda$ longitude, and $\phi$ latitude. Note also that $\mathbf{v}$ appears in these equations; this is calculated diagnostically, at each time-step, from the vorticity and divergence. Expressions for the divergence $\nabla \cdot$, gradient $\nabla (\cdot)$ and curl $\nabla \times$ operators in spherical polar coordinates may be found in Williamson et al. (1992). The expression for the Laplacian $\nabla^2$ is given in (3.41).

The truncation of the model is triangular, and the product terms in the governing equations are calculated on a quadratic Gaussian grid (see chapter 4). The time-stepping is a semi-implicit leapfrog scheme, with two-level time-stepping being used for every 36th time-step, to damp the computational mode.

The combination of triangular truncation and a quadratic Gaussian grid is a compromise. The number of degrees of freedom in the quadratic grid is larger than that in the truncation, which means that there will be some loss of small scales in the transformation from grid to spectral space. For this reason, the orography is specified on a linear Gaussian grid (and supplied to the model in the form of spectral
coefficients). The linear grid comprises fewer points, and so more nearly matches the number of degrees of freedom in the truncation. This ensures that the effect of this inherent smoothing on the model orography is mitigated to some extent.

This combination of grid and truncation used in this experiment is a very commonly used configuration in spectral models, because it prevents the unwanted aliasing of small-scale non-linear products. Using it here gives the experiment greatest relevance to this type of model. Secondly, it is the differences between runs which are important in this experiment, and so the effect of the truncation smoothing will be cancelled out in the comparison.

Throughout the experiment, the smoothing under investigation is carried out isotropically, with the smallest scales being removed by the Raymond filter. Essentially, this means that the spherical harmonics present in the gridded orography but not resolved by the triangular truncation do not contain any significant data. Furthermore, the mean orography, though not smoothed isotropically, has those scales effectively truncated when supplied to the model. Thus, the outcome of the experiment necessarily depends only on those scales which the model can resolve.

The model is initialized by specifying the vorticity and divergence distributions; at \( t = 0 \), the divergence is set to \( \delta = 0 \) over the whole domain, and the height field is then calculated to fulfil the condition that the divergence tendency \( \frac{\partial \delta}{\partial t} = 0 \). In all the experiments and tests carried out, solid-body rotation is specified as the initial condition, so that

\[
\begin{align*}
u &= \frac{\pi \cos \phi}{a}, \\
 \delta &= 0.
\end{align*}
\]

Application of (5.1) and (5.2) leads to

\[
\begin{align*}
z &= \frac{2\pi}{a} \sin \phi, \\
 \delta &= 0.
\end{align*}
\]

The spherical harmonic \( Y_{0,1} = \sqrt{\frac{3}{2}} \sin \phi \), has the same form as the RHS of (5.8), so, in order to specify the vorticity in the model, it is only necessary to specify a single coefficient in the spectral representation. The impact of the model initialisation upon the state of balance of the flow is discussed below; this initialisation method
was chosen because it fulfils the requirements of the test-case specified by Williamson et al. (1992), used to confirm the correct operation of the model below.

In addition to the terms in the continuous equations (5.3–5.5), the numerical discretization includes scale-selective dissipation to prevent the accumulation of energy at the grid-scale. This is accomplished by the addition of terms proportional to \( \frac{1}{\tau} \nabla^n \delta \) and \( \frac{1}{\tau} \nabla^n h^* \) on the right-hand-sides of equations (5.4) and (5.5), respectively, where \( \tau \) is the diffusion time-scale, and \( n \), the order of the hyperdiffusion, is even. The particular choice of dissipation used here is a little unorthodox; the impact of this is considered in section 5.4.6, where it is found not to affect the conclusions of this chapter.

### 5.2.2 Testing

In order to confirm that the model was working correctly, and could be initialized properly, a test-case was required. A complete suite of test-cases for the evaluation of shallow water models has been developed by Williamson et al. (1992); one of these is concerned with flow over orography, making it the most suitable for use in this case.

The Williamson et al. orographic test-case is specified in terms of the domain geometry, the initial flow, the orography, and the time for which the model is run. The domain is the surface of the Earth, so that

\[
a = 6.37122 \times 10^6 \text{m},
\]

\[
\Omega = 7.292 \times 10^{-5} \text{s}^{-1},
\]

\[
g = 9.80616 \text{ms}^{-2}.
\]

The initial flow is zonal solid-body rotation, with an equatorial velocity of \( \overline{u} = 20 \text{ms}^{-1} \), and surface height determined as described above. In this case, the mean depth of the fluid is such that the initial equatorial fluid depth is 5960 m.

The orography \( h_0 \) is an isolated conical hill defined by

\[
h_0 = \overline{h}_0 \left(1 - \frac{r}{R}\right),
\]
with \( \bar{h}_0 = 2000 \) m, \( R = \pi/9 \), and \( r^2 = \min[R^2, (\lambda - \lambda_c)^2 + (\phi - \phi_c)^2] \). The location of the centre of the mountain is specified by \( \lambda_c = 3\pi/2 \) and \( \phi_c = \pi/6 \).

The model is run for 15 days, and the surface height output every 5 days for comparison with high-resolution reference solutions given in Jakob et al. (1993). These reference solutions are shown in figure 5.1; the model output is presented in figure 5.2. The model time-step was 345.6 s, which is equivalent to 250 time-steps per day. Sixth-order hyper-diffusion was used, and the diffusion time-scale was 4 hours. The simulation was conducted at a spectral resolution of T99, with the orography being specified on a linear grid of 200 \( \times \) 100 points. The same model configuration is used in the experiments detailed below.

Comparison between figure 5.1 and figure 5.2 shows that although there are small differences between the reference solution and the model output, the overall pattern of contours is substantially the same, showing that the model is functioning correctly.

5.3 Experimental method

5.3.1 Basic considerations

The question to be answered by the experiment hinges on an assessment of the accuracy of the representation of the flow. Two things are necessary for this to be possible: criteria for assessing accuracy, and a reference solution to be taken as ‘truth’. The testing of the model, described above, provides a template for this process. For this model, it is not possible to calculate the flow of air over orography, even an idealized mountain, analytically. Therefore, the best estimate of ‘truth’ is a high-resolution model run, with well-resolved orography. It is important that the orography is well-resolved, as the orographic forcing, to a great extent, determines the spectral composition of the waves present in the flow. Well-resolved waves are represented most accurately in the model, and so the exclusion of poorly-resolved waves will ensure that the flow is as accurate as possible.

The second consideration is the choice of criteria for comparison of the different lower-resolution model runs with this ‘truth’. For the testing of the model, the
Figure 5.1: Contour plots of the surface height for the Williamson orographic test case – supplied reference solution (in Jakob et al. (1993)). The contour interval is 100 m.
Figure 5.2: Contour plots of the surface height for the Williamson orographic test case – solution calculated by the spectral shallow water model described in the text. The contour interval is 100 m.
comparison made was between the height fields of the test and reference model runs, and this comparison was done by eye. However, for the purposes of this experiment, something more sophisticated and quantitative is needed. The methods employed are described in detail below, but in essence comprise the quantitative comparison, on a global basis, of all model fields, rather just the height field.

These considerations lead to the identification of three different types of model run:

- **Reference run:** high resolution, well-resolved mean orography;
- **Control run:** low resolution, unsmoothed mean orography;
- **Experimental run:** low resolution, smoothed orography.

We distinguish between the experimental and control runs as it allows the discrimination between effects due to the change in resolution and those due to the application of smoothing. The model parameters for these runs are discussed in detail below.

In seeking an optimal combination of values of $\varepsilon$ and $\gamma$, it is important remember that the smoothed orography will be used in a wide variety of flow regimes. It is possible that different combinations of $\varepsilon$ and $\gamma$ might be appropriate under different flow regimes, and it should be remembered that not all terrestrial flow regimes occur over all parts of the globe equally. It is important, therefore, to carry out the experiment under a range of representative flow regimes; the specifications of these regimes are given below.

A more sophisticated smoothing method might take the prevailing local flow regimes into account when smoothing the orography. This could well form a natural extension of the present work, but, currently, such a smoothing method has not been developed, and so, in its absence, the smoothing is here applied equally to all parts of the global orography.

### 5.3.2 Model configuration

In determining the model configuration to be used for the experiments, the model resolution and length of run chosen is limited by considerations of computational
cost. The high-resolution reference solution has to be calculated, and this is what limits the length of run which is possible. In addition, the reference run has to have significantly higher resolution than the experimental and control runs, but these lower-resolution runs have to have enough resolution to resolve the large-scale flow well.

Taking these considerations into account, and having run some trial integrations to determine the computational cost of running the model, the parameter combination shown in table 5.1 was arrived at. Note that the time-step is equivalent to 1000 and 250 time-steps per day for the reference and experimental runs, respectively. The time-steps were chosen so that the two runs have the same Courant number (see equation (2.44)), and so that the experimental run would have the same time-step as the Williamson test-run described earlier. In common with that test-run, the diffusion time-scale of both the reference and experimental runs was 4 hours, and 6th-order hyper-diffusion was used.

In order to fulfil the requirement, explained in the preceding section, that the orography in the reference run be well-resolved, the mean orography at a resolution of 200 $\times$ 100 is used. This is spectrally transformed, and the resulting coefficients are padded with zeros, so that it may be supplied to the model at T399 resolution. This permits a direct comparison between the model running at T399 and T99 with the same mean orography.

<table>
<thead>
<tr>
<th>type</th>
<th>truncation</th>
<th>quadratic grid</th>
<th>linear grid</th>
<th>length of run (days)</th>
<th>$\Delta t$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference</td>
<td>T399</td>
<td>1200 $\times$ 600</td>
<td>800 $\times$ 400</td>
<td>10</td>
<td>86.4</td>
</tr>
<tr>
<td>control and experimental</td>
<td>T99</td>
<td>300 $\times$ 150</td>
<td>200 $\times$ 100</td>
<td>10</td>
<td>345.6</td>
</tr>
</tbody>
</table>

Table 5.1: Parameter combination used with the spectral shallow water model for the various different types of model run.
Table 5.2: The nine flow regimes initially considered, and identified by letters. Note that, for clarity, there is no regime I.

5.3.3 Flow regimes

For an Earth-like world in a quasi-atmospheric numerical model, such as this one, the two model parameters which may be used to adjust the flow regime are the mean fluid depth $H$ and the initial equatorial flow speed $\overline{u}$. It is taken that the radius of the planet $a$, the planetary rotation rate $\Omega$ and the acceleration due to gravity $g$ are not candidates for variation in the experiment because they are fundamental aspects of the Earth system.

5.3.3.1 Selection of flow regimes

In order to test the smoothed orography in a wide range of flow regimes, three values each of $\overline{u}$ and $H$ were selected, encompassing a realistic range of atmospheric values. All nine possible combinations of these parameters are identified by letters, and these are shown in table 5.2. The values of $\overline{u}$ are chosen to be separated by a factor of five, and, originally, it had been intended that the lowest value of $H$ should be 100 m, giving a separation factor of ten. However, it was found that this led to the height of the fluid surface being negative at high latitudes. Being deeply unphysical, this was rejected, and the lowest value changed to 200 m.

Having ensured that the initial conditions did not result in negative fluid depth, the same condition must be applied to disturbances to the flow. For a disturbance of length scale $L$ and velocity scale $U$, the associated height disturbance may be calculated to a good approximation by assuming geostrophic balance; this gives a height scale of $fLU/g$. Clearly, this must be less than $H$, the mean depth of the
Table 5.3: Values of $L_{\text{max}}$ for the nine different flow regimes.

<table>
<thead>
<tr>
<th>$H$/m →</th>
<th>200</th>
<th>1000</th>
<th>10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi$/ms$^{-1}$ ↓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>$2.69 \times 10^6$</td>
<td>$1.34 \times 10^7$</td>
<td>$1.34 \times 10^8$</td>
</tr>
<tr>
<td>25</td>
<td>$5.38 \times 10^5$</td>
<td>$2.69 \times 10^6$</td>
<td>$2.69 \times 10^7$</td>
</tr>
<tr>
<td>125</td>
<td>$1.08 \times 10^5$</td>
<td>$5.38 \times 10^5$</td>
<td>$5.38 \times 10^6$</td>
</tr>
</tbody>
</table>

fluid, which leads to the condition $gH > fLU$. We may rearrange this to give a maximum permitted length scale $L_{\text{max}}$:

$$ L < L_{\text{max}} = \frac{gH}{fU}, \quad (5.14) $$

This condition is most stringent at the pole, where $f$ has its maximum value. It may also be understood in terms of the dimensionless numbers explained in section 5.3.3.2 below. In terms of the Rossby and Froude numbers, the condition is $Ro > Fr$, while, in terms of the Rossby and Burger numbers, it is $Ro < Bu$.

Taking $f = 1.46 \times 10^{-4}$ s$^{-1}$, and $g = 9.81$ ms$^{-2}$, the values of $L_{\text{max}}$ (given in metres) for the nine flow regimes are given in table 5.3. To ensure that negative fluid depth does not occur, it is necessary to arrange that $L_{\text{max}}$ is larger than any disturbances that may exist in the flow. A likely order of magnitude for this maximum disturbance size is the radius of the Earth ($6.37 \times 10^6$ m). However, since the above values of $L_{\text{max}}$ were calculated for the poles, they represent an upper bound on the values of $L_{\text{max}}$ appropriate for the whole globe. On these grounds, it was decided to reject flow regimes for which $L_{\text{max}} < 10^6$ m, in other words, regimes D, G and H. To test this reasoning, 10-day model runs were made under each flow regime, with the mean orography, at T99. As expected, the integration failed for regimes D, G and H, but was successful for the other regimes. Thus, the experiment was confined to the six regimes A, B, C, E, F and J.

Although the above-discussed criterion prevents the surface disturbances from being greater than the depth of the fluid, it does not take into account of the height of the orography. For this reason, it was decided to scale the orography with $H$, so that there is no danger of generating negative fluid depth. The scaling factors chosen were based on the maximum height of the mean orography at $200 \times 100$ resolution, which is 5343 m. The height of the orography was scaled so that the maximum
height of the field was $\frac{1}{10}H$, which leads to factors of $3.74 \times 10^{-3}$, $1.87 \times 10^{-2}$ and 0.187 for the three different fluid depths used. The parameters for each flow regime are summarized in table 5.4.

### 5.3.3.2 Characterisation of flow regimes with dimensionless numbers

The properties we might expect the flow to exhibit in the six flow regimes may be understood to some degree by the consideration of various dimensionless numbers. It is possible to construct several different dimensionless numbers from the parameters in the model, but here we will focus on four examples. In all cases, $U$ is a velocity scale and $L$ a length scale, while the other symbols have their usual meanings.

- **The Rossby number:**
  
  $$ Ro = \frac{U}{fL} $$

  The Rossby number quantifies the importance of rotation in the flow. If $Ro \ll 1$, then the Coriolis force is significant, and the flow is largely in geostrophic balance. The other extreme, when $Ro \gg 1$, indicates that rotation has an insignificant effect on the flow.

- **The Froude number:**
  
  $$ Fr = \frac{U^2}{gH} $$

  The Froude number is the square of the ratio between the velocity scale and the speed of gravity wave propagation $\sqrt{gH}$. If $Fr > 1$ then the flow speed is
greater than the gravity wave speed, and so the gravity waves cannot play a role in establishing balance in the flow. This is known as *supercritical* flow. In *subcritical* flow, where \( \text{Fr} < 1 \), the flow is able to attain a balanced state by the radiation of gravity waves. See, for instance, Gill (1982), and the discussion of balance later in this chapter.

A further application of the Froude number comes from the fact that it is roughly equal to the ratio of the kinetic energy of the flow to the energy needed to raise a fluid parcel a height of \( H \). By replacing \( H \) with \( h \), an orographic height scale, we obtain the orographic Froude number:

\[
\text{Fr}_o = \frac{U^2}{gh}.
\]

If \( \text{Fr}_o > 1 \) then there is enough energy in the flow to raise fluid parcels over the orography. The converse, that the flow goes around the orography when \( \text{Fr}_o < 1 \), is not strictly true, as in a shallow water model the flow must, in all circumstances, go over hills. However, \( \text{Fr}_o \) does allow the characterisation of the flow into types which are relevant to more representative models.

In a three-dimensional model, the Froude number gives a measure of the importance of stratification to the flow. This is discussed by, for instance, Cushman-Rosin (1994).

- The Burger number:

\[
\text{Bu} = \frac{gH}{f^2 L^2} = \frac{\text{Ro}^2}{\text{Fr}}.
\]

The Burger number is the square of the ratio between the Rossby radius of deformation \( \left( = \sqrt{gH/f} \right) \) and the length-scale \( L \). The Rossby radius of deformation is a natural length-scale associated with adjustments to a balanced state. See, for instance, Salmon (1998) for information.

In addition, the Burger number is approximately equal to the ratio between kinetic and potential energy in a balanced flow of length scale \( L \). This can be seen by considering that kinetic energy \( KE \sim HU^2 \) and potential energy \( PE \sim g\eta^2 \), with \( \eta \sim fLU/g \) (from geostrophic balance).

Table 5.5 shows the values of these dimensionless numbers for the six flow regimes. Because both the Rossby number and the Burger number contain the Coriolis parameter \( f \), it is important to consider their values at different latitudes. At the
Chapter 5 Experiments with a global shallow water model

equator, \( f = 0 \), so the Rossby number and the Burger number go to infinity. Therefore, to get representative values at high and mid-latitudes, table 5.5 includes values of both numbers at \( \phi = 45^\circ \) and \( 90^\circ \). The value of \( U \) is set to the initial equatorial x-velocity (\( \overline{u} \)), and \( L = 2000 \) km, a typical horizontal length scale of a mountain range. In calculating \( F_{ro} \), the ratio \( h/H = 0.0468 \); this is the appropriate value for a mountain of height 2500 m when the orographic scaling described above is taken into account.

The values of the dimensionless numbers given in table 5.5 allow some conclusions to be drawn about the nature of the different flow regimes. Firstly, with the exception of J, all flow regimes have \( Ro \ll 1 \) in mid and high latitudes; this means that flow in these areas will be close to geostrophic balance. Secondly, all regimes have \( Fr < 1 \), which means that the flow will be sub-critical. This is a good thing, as it reduces the likelihood of non-linear wave-breaking in the model, which would probably cause the simulation to fail. In addition, the flow will be closest to balance when \( Fr \ll 1 \), which is true for all regimes, except perhaps regime J.

The Burger number and the orographic Froude number present a more complex picture. The orographic Froude number is greater than 1 for flow regimes E and J, which indicates that we might expect to see flow-over behaviour for features less than \( \approx 2500 \) m in height. The other regimes have smaller values of \( F_{ro} \), and so are more likely to exhibit flow around orographic features. The Burger number also divides the regimes into two categories, where \( Bu \) is greater or less than one.

The relationships between the different flow regimes in terms of \( Ro \), \( Fr \) and \( Bu \) is illustrated in figure 5.3, which locates them in the \( Ro-Fr \) plane.

5.3.4 The effect of initialisation on model balance

There is a large body of evidence to suggest that, in general, the atmosphere is, on large scales, close to a balanced state most of the time. The concept of balance is not, though, entirely straightforward to understand; indeed there is some controversy over its more fundamental aspects, including the postulated existence of the so-called slow manifold (see, for instance, Lorenz (1986) and Lorenz and Krishnamurthy (1987)). In general, however, the idea is that the flow may be decomposed into
<table>
<thead>
<tr>
<th>Regime</th>
<th>Ro ((\phi = 45^\circ))</th>
<th>Ro ((\phi = 90^\circ))</th>
<th>Bu ((\phi = 45^\circ))</th>
<th>Bu ((\phi = 90^\circ))</th>
<th>Fr</th>
<th>Fr_o</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>(2.42 \times 10^{-2})</td>
<td>(1.71 \times 10^{-2})</td>
<td>(4.61 \times 10^{-2})</td>
<td>(2.31 \times 10^{-2})</td>
<td>(1.27 \times 10^{-2})</td>
<td>(2.72 \times 10^{-1})</td>
</tr>
<tr>
<td>B</td>
<td>(2.42 \times 10^{-2})</td>
<td>(1.71 \times 10^{-2})</td>
<td>(2.31 \times 10^{-1})</td>
<td>(1.15 \times 10^{-1})</td>
<td>(2.55 \times 10^{-3})</td>
<td>(5.45 \times 10^{-2})</td>
</tr>
<tr>
<td>C</td>
<td>(2.42 \times 10^{-2})</td>
<td>(1.71 \times 10^{-2})</td>
<td>(2.31)</td>
<td>(1.15)</td>
<td>(2.55 \times 10^{-4})</td>
<td>(5.45 \times 10^{-3})</td>
</tr>
<tr>
<td>E</td>
<td>(1.21 \times 10^{-1})</td>
<td>(8.57 \times 10^{-2})</td>
<td>(2.31 \times 10^{-1})</td>
<td>(1.15 \times 10^{-1})</td>
<td>(6.37 \times 10^{-2})</td>
<td>(1.36)</td>
</tr>
<tr>
<td>F</td>
<td>(1.21 \times 10^{-1})</td>
<td>(8.57 \times 10^{-2})</td>
<td>(2.31)</td>
<td>(1.15)</td>
<td>(6.37 \times 10^{-3})</td>
<td>(1.36 \times 10^{-1})</td>
</tr>
<tr>
<td>J</td>
<td>(6.06 \times 10^{-1})</td>
<td>(4.29 \times 10^{-1})</td>
<td>(2.31)</td>
<td>(1.15)</td>
<td>(1.59 \times 10^{-1})</td>
<td>(3.40)</td>
</tr>
</tbody>
</table>

Table 5.5: Values of the dimensionless parameters defined in the text for each flow regime, and for different latitudes, where appropriate.
Figure 5.3: Diagram of Rossby/Froude number space, showing the location of the six flow regimes A, B, C, E, F and J that are detailed in the text. The length of the lines for each regime shows the range of values between 45° and 90° latitude. The two other lines indicate $Bu = 1$ and $Ro = Fr$. Note the logarithmic scales of both axes.

a more slowly varying part (the balanced part), and a more quickly varying part (the unbalanced part, consisting of gravity waves); a flow with no (or almost no) unbalanced part is said to be balanced, and will remain so (or nearly so) over a long time period – this is the slow manifold. In terms of modelling simplified systems of equations, balance can be defined by the judicious approximation of the equations on the basis of scaling arguments. Typical balance states include geostrophic, hydrostatic and cyclostrophic balance; for a detailed discussion, see, among others, Gill (1982) and Holton (1992).

For the purposes of this experiment, it is sufficient to note that, because the atmosphere is close to balance, the amount of gravity wave activity is small compared to the other dynamical processes which are present. This is important from the perspective of model initialisation, for when flow is initialized in an unbalanced state it will, in general, undergo a process of adjustment to a balanced state, during which gravity waves are emitted\(^1\). The failure of L. F. Richardson to properly initialize his model, and so avoid the spurious generation of large-amplitude gravity waves, was one of the reasons for the failure of his historic first attempt at numerical weather

\(^1\)For a discussion of this process, see Ford et al. (2000).
prediction (see Richardson (1922) and Lynch (1999)). For the present experiment to be most relevant to simulations of the real atmosphere, it is necessary to make sure that gravity wave activity is only present at a level that does not swamp the evolution of the balanced flow.

In the case of the shallow water equations, a common means of specifying balance is to set $\frac{\partial \delta}{\partial t} = \delta = 0$ initially, which is appropriate in the absence of orography. This is the method of initialisation used in this experiment. However, when orography is present, the balanced flow will include a quasi-stationary, non-zero divergence field. Thus, the flows generated by the initialisation employed here will include an unbalanced component. The presence of inertia-gravity waves, emitted by this unbalanced flow, shows up as a time-varying component of the divergence. Consequently, the decomposition of the divergence into steady and time-varying components can help diagnose the state of balance of the flow. In order to assess the relative sizes of the balanced and unbalanced divergent flow, the ratio of the domain integrated divergence squared of the two components was calculated for flow regime A:

$$\frac{\int (\delta - \overline{\delta})^2 dA}{\int \overline{\delta}^2 dA}$$

The time-mean divergence field $\overline{\delta}$ is calculated from fields output after each day of model time, and $\delta$ is the divergence field at the end of the run (after ten days). Flow regime A was chosen for this study as it should be the most balanced of the six, as may be seen by considering the values of the Rossby and Froude numbers, and thus any spurious imbalance will show up most clearly. The above calculation was performed for the reference solution at T399 resolution, and for the control solution at T99 resolution. In both cases, the integrals are calculated on the unit sphere. The resulting value, and the values of the numerator and denominator in each case, are given in table 5.6.

In both cases, the value of the ratio is less than one, showing that the balanced flow dominates the integration, which demonstrates that the initialisation procedure is adequate for our purposes. The order-of-magnitude difference between the values for the two runs is curious; table 5.6 shows this is due to the control run having both a greater balanced component ($\int \delta^2 dA$) and a smaller unbalanced component
Chapter 5  Experiments with a global shallow water model

Table 5.6: Global integrals of unsteady and steady divergence squared, and their ratio, for flow regime A, calculated for both reference and control runs.

<table>
<thead>
<tr>
<th></th>
<th>$\int (\delta - \overline{\delta})^2 , dA$</th>
<th>$\int \overline{\delta}^2 , dA$</th>
<th>$\frac{\int (\delta - \overline{\delta})^2 , dA}{\int \overline{\delta}^2 , dA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference run</td>
<td>$9.05 \times 10^{-14} , s^{-2}$</td>
<td>$1.45 \times 10^{-13} , s^{-2}$</td>
<td>0.625</td>
</tr>
<tr>
<td>control run</td>
<td>$1.81 \times 10^{-14} , s^{-2}$</td>
<td>$2.94 \times 10^{-13} , s^{-2}$</td>
<td>0.062</td>
</tr>
</tbody>
</table>

$(\int (\delta - \overline{\delta})^2 \, dA)$. The difference in the unbalanced part can be attributed to the action of the scale-selective dissipation at different resolutions; because of the higher resolution of the reference run, the scale selective dissipation acts largely on scales that are not resolved in the control run, and whose amplitude is negligible. The consequence of this is that the dissipation is effectively weaker in the reference run. Because this effect is solely a result of the difference in resolution, it will not have an impact on the experiment; comparing the errors in the experimental run with the errors in the control run eliminates these effects from the results. The much smaller difference in the magnitude of the balanced part of the divergent flow seems likely to be also due to the difference in resolution, and the treatment of well-resolved and just-resolved spectral components, though the mechanism is not understood at the moment. The effect of the scale-selective dissipation upon the steady, forced part of the flow would be less than on the unforced part, as the former has its amplitude constantly reinforced by the forcing.

5.3.5  Quantitative comparison between model runs

A large amount of data is produced by a model run, and given that the model is to be run many times, the volume of data will be quite considerable. Consequently, it is necessary to establish beforehand some procedures for comparing the output of different model runs.

Two measures are used to compare model fields: the $L_2$ norm (sometimes called the RMS, or root-mean-squared, error) and the $L_\infty$ norm (the maximum error). These norms are defined thus:

$$L_n(\phi) = \sqrt[n]{\frac{\sum_{i,j} (\phi_{i,j} - \hat{\phi}_{i,j})^n}{N}},$$  \hspace{1cm} (5.19)
where $\widehat{\phi}$ is the reference field anomaly after ten days (i.e. $\widehat{\phi} = \phi_{10\text{days}} - \phi_0$), and $\hat{\phi}$ is an instance of the same field anomaly from another model run. Note that $\overline{\phi}$ denotes the reference solution, and $\phi_0$ is the initial state; the summation over $i$ and $j$ indicates the whole domain. On a more practical level, the $L_\infty$ norm may be written

$$L_\infty(\widehat{\phi}) = \sqrt{\max[(\hat{\phi}_{i,j} - \overline{\phi}_{i,j})^2]}.$$ (5.20)

In this calculation, the field anomalies are used instead of the actual values of the field for reasons of consistency: the change in orographic volume caused by applying the variational smoother could, otherwise, introduce a systematic error to the results. These two norms may be used to compare any of the model fields, both prognostic ($\zeta, \delta$ and $h$) and diagnostic ($u, v$ and potential vorticity). Potential vorticity ($PV$) may be defined for the shallow water case as

$$PV = \frac{\zeta}{H + \eta}.$$ (5.21)

One complication is that the reference and experimental runs are at different resolutions, and thus are output on different grids. This problem is overcome by transforming the reference solutions to spectral space, truncating at the resolution of the experimental run (T99), and transforming back onto the appropriate quadratic grid ($300 \times 150$). The experimental solutions may then be compared with the reference solutions in a consistent manner.

To aid the interpretation of the $L_2$ and $L_\infty$ norms, relative errors are calculated. The idea behind these is to compare the $L_2$ and $L_\infty$ norms for each run with the same norms for the control run. If the control run is denoted $\phi_C$, then the relative error $R_n$ is given by:

$$R_n = \frac{L_n(\phi)}{L_n(\phi_C)}.$$ (5.22)

The hypotheses given at the start of this chapter lead to the expectation that the error attached to the model runs with particular smoothed orographies will be less than that attached to the control run. If this is the case, the relative error will be less than 1. A very clear result would show a minimum in $R_n$ for a particular parameter combination.
5.3.6 Procedure for exploring the parameter space

The question posed in section 5.1.2 requires the study of a two-dimensional parameter space, in $\gamma$ and $\varepsilon$. Given that both parameters may take a wide range of values, some rationale needs to lie behind the manner of exploration of the parameter space, in order to constrain the experiment within the bounds of tractability.

The starting point for the exploration is the value of $\varepsilon = 1.0$, which was suggested by Webster et al. (2003) for use with the Raymond filter, and is employed by the Met Office in their Unified Model. Keeping the Raymond filter parameter fixed at $\varepsilon = 1.0$, the model is run with orography produced by the variational method with $\gamma$ varying from 0 to 5000. The output of the model for these different smoothed orographies is compared with the reference and control solutions, in the manner described above.

Having examined the sensitivity of the model solution to $\gamma$, the sensitivity to $\varepsilon$ will be explored. The foregoing experiment will inform the selection of a non-zero value of $\gamma$, which will be held constant as $\varepsilon$ is varied between 0.1 and 10.0. The same process of comparison is then undertaken, both with the reference and control runs.

Finally, the model will be run with orography generated by the variational smoothing method at $\varepsilon = 1.0$ and the previously chosen value of $\gamma$, but with the sea also constrained to be flat. This will allow the effects of this constraint to be studied. Although it is very likely to be a worthwhile feature of the variational scheme from the point of view of full NWP and climate models, it is not expected that it will show any particular benefit in the shallow-water model. Therefore, it was decided that only a single model run would be done with it in place, in order to provide an illustration of its effect on the flow.

Figure 5.4 summarizes the exploration of parameter space described above. The values of $\varepsilon$ and $\gamma$ shown on the axis of the figure were determined as the experiment progressed. The reasoning behind these choices is explained in the following sections.
5.4 Results

5.4.1 The control run

Before discussing the sensitivity of the model solution to $\gamma$ and $\varepsilon$, it is important to have an appreciation of the size and nature of the errors present in the control run. As explained above, the control run is made with the same $200 \times 100$ resolution mean orography used in the T399 reference run, but the model resolution in this case is T99. Thus, the difference between the control run and the reference run is that in the reference run the orography is well-resolved, whereas in the control run it contains components at the truncation limit of the model. Although the emphasis of the experiment is on global error measures, it is worth noting that the spatial distribution of the control run error varies considerably among the different flow regimes. Figures 5.5 and 5.6 show the different spatial distributions of the control run errors in the height field for flow regimes A and F, respectively; regime A has its largest errors over polar mountains, while, in regime F, they are in the tropics and mid-latitudes. Regime F has larger Rossby and Froude numbers than regime A, which suggests that the errors in regime F may be unbalanced motions (i.e. propagating waves), whereas the errors in regime A may be in the steady response. In addition, the more noisy appearance of the regime F plot is suggestive of this. Another possibility is that the global distribution of different flow regimes...
differs between the two parameter combinations sufficiently to change the global distribution of errors.

Table 5.7 shows the $L_2$ error for the control run. The striking thing about these values is that they are very small in comparison to the typical values for the flow. Taking the surface height as an example, the maximum value of $L_2$ is for regime J (15.44 m). However, the mean fluid depth is 10,000 m for this regime, which makes the error only 0.15% of that depth. The other $L_2$ height errors are a comparable fraction of their respective fluid depths. To get a more consistent indication of the size of the errors in comparison to the magnitude of disturbances of the fields concerned, they are divided by the root-mean-squared value of that field's anomaly after ten days integration; the resulting values are shown in table 5.8.

The largest normalized errors shown in table 5.8 are those for divergence, which are of order 1. This means that the size of the error is comparable with the size of the disturbance caused by the orographic forcing. Apart from the vorticity in flow regime A, the other errors are of order $1/10$.

The importance of these errors is that they are the starting-point for the evaluation of the smoothing. The fundamental hypothesis of this work is that the control run solutions differ from the reference solutions due to the fact that the orography contains just-resolved components in the control run, in contrast to the reference run. Orographic smoothing is suggested as a means of combating these errors. Thus, the normalized errors shown in table 5.8 quantify the importance of the near-grid-scale effect on each of the fields, and indicate that in evaluating the smoothing, the reduction of errors in some fields is more important than the reduction of errors in others.

This analysis leads to the conclusion that the effect of orographic smoothing on the divergence is more significant than its effect on the other fields. The need to focus on the divergence may also be arrived at from a different perspective, upon consideration of the possible mechanisms for the formation of spurious ‘grid-point storms’. Grid-point storms are likely to be caused by the triggering of moisture parameterizations by small-scale vertical motion. In a shallow water model, there is no vertical motion, but divergence may be seen as a proxy for it. Thus, for our purposes, the accuracy of the near-grid-scale component of the divergence is very
Figure 5.5: Difference between control and reference runs after 10 days integration: height field, flow regime A. The axes are labelled in degrees of longitude and latitude, the colour bar in metres.

Figure 5.6: Difference between control and reference runs after 10 days integration: height field, flow regime F. Other details are as figure 5.5.
important. Since this is the component of the divergence field most affected by the orographic smoothing, the sensitivity of the divergence error to the parameters $\gamma$ and $\varepsilon$ is perhaps the most important measure of the success of that smoothing.

5.4.2  Sensitivity of the model solution to $\gamma$

The generation of the smoothed orography used in this section has been detailed in chapter 4. These smoothed orographies were generated with the variational smoothing scheme, with $\varepsilon = 1.0$, and $\gamma$ varied between 0 and 5000. However, for reasons of computational expense, model runs were not undertaken for all the generated orographies; the runs done were for $\gamma = 1, 2, 3, 8, 20, 30, 40, 100, 300, 1000$ and 5000. In each case, model runs were made for each of the six flow regimes detailed above, giving 66 runs in total. The resulting sensitivity to $\gamma$ is shown for all six fields in figures 5.7–5.12 ($R_2$) and figures 5.13–5.18 ($R_1$). The $y$-axes of these plots do not cover the whole range of $\gamma$ investigated, as the most interesting behaviour lies in the range shown.

A number of things are clear from the first set of figures (5.7–5.12), which show the variation of $R_2$ with $\gamma$:

- Although in some cases the value of $R_2$ decreases as $\gamma$ reaches its very highest values, in general, $R_2$ increases with $\gamma$.
- Most of the increase in $R_2$ occurs in the range $0 < \gamma < 500$.
- Only the divergence and height have $R_2 < 1.0$ for some values of $\gamma$, and this is not for all six flow regimes. This indicates that, by this measure and in almost all cases, the model more accurately represents the flow with the mean orography than with any of the smoothed orographies. The results for the divergence are quite striking, however, as they divide the flow regimes into two groups, one showing a benefit from smoothing, the other not. By examining the evolution of the global integral of divergence squared ($\int \delta^2 dA$) during the course of the model runs, we can gain some insight into why this happens; the relevant data is shown in figure 5.19 (regime A) and figure 5.20 (regime E).
<table>
<thead>
<tr>
<th>Regime</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height /m</td>
<td>0.23</td>
<td>0.15</td>
<td>0.67</td>
<td>0.81</td>
<td>3.20</td>
<td>15.44</td>
</tr>
<tr>
<td>PV /m⁻¹s⁻¹</td>
<td>3.98 x 10⁻⁸</td>
<td>4.75 x 10⁻¹⁰</td>
<td>3.64 x 10⁻¹¹</td>
<td>1.00 x 10⁻⁸</td>
<td>4.63 x 10⁻¹¹</td>
<td>1.40 x 10⁻¹⁰</td>
</tr>
<tr>
<td>Vort /s⁻¹</td>
<td>1.18 x 10⁻⁶</td>
<td>3.96 x 10⁻⁷</td>
<td>3.53 x 10⁻⁷</td>
<td>1.82 x 10⁻⁶</td>
<td>4.25 x 10⁻⁷</td>
<td>8.59 x 10⁻⁷</td>
</tr>
<tr>
<td>Div /s⁻¹</td>
<td>1.87 x 10⁻⁷</td>
<td>1.16 x 10⁻⁷</td>
<td>1.20 x 10⁻⁷</td>
<td>6.27 x 10⁻⁷</td>
<td>6.29 x 10⁻⁷</td>
<td>2.25 x 10⁻⁶</td>
</tr>
<tr>
<td>U /ms⁻¹</td>
<td>0.0837</td>
<td>0.0388</td>
<td>0.0383</td>
<td>0.2017</td>
<td>0.1522</td>
<td>0.5330</td>
</tr>
<tr>
<td>V /ms⁻¹</td>
<td>0.0907</td>
<td>0.0328</td>
<td>0.0333</td>
<td>0.1722</td>
<td>0.1259</td>
<td>0.3730</td>
</tr>
</tbody>
</table>

Table 5.7: $L_2$ error for the control run, as compared with the high-resolution reference run. Note that these values are not normalized in any way.
<table>
<thead>
<tr>
<th>Regime</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>$8.571 \times 10^{-2}$</td>
<td>$3.233 \times 10^{-2}$</td>
<td>$9.928 \times 10^{-2}$</td>
<td>$2.733 \times 10^{-2}$</td>
<td>$8.741 \times 10^{-2}$</td>
<td>$1.059 \times 10^{-1}$</td>
</tr>
<tr>
<td>Pv</td>
<td>$2.824 \times 10^{-1}$</td>
<td>$1.835 \times 10^{-1}$</td>
<td>$1.886 \times 10^{-1}$</td>
<td>$9.617 \times 10^{-2}$</td>
<td>$1.009 \times 10^{-1}$</td>
<td>$1.436 \times 10^{-1}$</td>
</tr>
<tr>
<td>Vort</td>
<td>$5.319 \times 10^{-1}$</td>
<td>$2.242 \times 10^{-1}$</td>
<td>$1.932 \times 10^{-1}$</td>
<td>$2.726 \times 10^{-1}$</td>
<td>$1.079 \times 10^{-1}$</td>
<td>$2.019 \times 10^{-1}$</td>
</tr>
<tr>
<td>Div</td>
<td>$9.143 \times 10^{-1}$</td>
<td>$8.345 \times 10^{-1}$</td>
<td>$8.679 \times 10^{-1}$</td>
<td>$8.875 \times 10^{-1}$</td>
<td>$1.130$</td>
<td>$8.562 \times 10^{-1}$</td>
</tr>
<tr>
<td>U</td>
<td>$2.439 \times 10^{-1}$</td>
<td>$8.557 \times 10^{-2}$</td>
<td>$7.258 \times 10^{-2}$</td>
<td>$8.916 \times 10^{-2}$</td>
<td>$6.514 \times 10^{-2}$</td>
<td>$1.282 \times 10^{-1}$</td>
</tr>
<tr>
<td>V</td>
<td>$1.898 \times 10^{-1}$</td>
<td>$6.182 \times 10^{-2}$</td>
<td>$6.146 \times 10^{-2}$</td>
<td>$7.615 \times 10^{-2}$</td>
<td>$5.267 \times 10^{-2}$</td>
<td>$9.721 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

*Table 5.8: $L_2$ error for the control run, normalized by the RMS value of the field anomaly after 10 days integration.*
Figure 5.7: The sensitivity of the $R_2$ error of the height field to $\gamma$.

Figure 5.8: The sensitivity of the $R_2$ error of the PV field to $\gamma$. 
Figure 5.9: The sensitivity of the $R_2$ error of the vorticity field to $\gamma$.

Figure 5.10: The sensitivity of the $R_2$ error of the divergence field to $\gamma$. 
Figure 5.11: The sensitivity of the $R_2$ error of the $x$-velocity field to $\gamma$.

Figure 5.12: The sensitivity of the $R_2$ error of the $y$-velocity field to $\gamma$. 
Figure 5.13: The sensitivity of the $R_\infty$ error of the height field to $\gamma$.

Figure 5.14: The sensitivity of the $R_\infty$ error of the PV field to $\gamma$. 
Chapter 5  Experiments with a global shallow water model

Figure 5.15: The sensitivity of the $R_{\infty}$ error of the vorticity field to $\gamma$.

Figure 5.16: The sensitivity of the $R_{\infty}$ error of the divergence field to $\gamma$. 
Figure 5.17: The sensitivity of the $R_{\infty}$ error of the $x$-velocity field to $\gamma$.

Figure 5.18: The sensitivity of the $R_{\infty}$ error of the $y$-velocity field to $\gamma$. 
The reason for the division of the flow regimes into those which show a positive effect from smoothing and those that do not is clear from these figures: in those flow regimes that show a positive effect, such as regime A, the control run shows greater divergence than the reference, and so smoothing, which reduces the overall divergence, has a positive effect; in the other regimes, where the reference run has more divergence than the control, smoothing only widens the gap between them. Interestingly, after an initial transient, the reference run has much more variability over the course of the integration than the lower resolution control and experimental runs. This is probably because of the different actions of the scale-selective dissipation discussed above; the variability of $\int \delta^2 dA$ in the reference runs is probably due to the scales which are not resolved in the lower resolution runs. (Note that, despite appearances, the variability of the reference run in regime E is actually, in fractional terms, less than that of regime A – approximately $\frac{1}{4}$ of the mean value, rather than $\frac{1}{3}$). In both plots, the evolution of the four lower resolution runs is identical in form, the only difference between then being their mean value of $\int \delta^2 dA$.

- The minimum value of $R_2$ occurs at $\gamma = 0$, for all flow regimes and all fields.

Additionally, there are, evident in the results, a number of relationships between the various dimensionless numbers and the magnitudes of $R_2$ for different fields:

- The PV and vorticity plots (figures 5.8–5.9) show $R_2$ increasing with increasing $Ro/Fr$. The trend in the vorticity error is illustrated in figure 5.21.

The dimensionless number $Ro/Fr = gH/UfL$, and may be interpreted by recalling that, in a geostrophically balance flow with length-scale $L$, $\eta \sim fLU/g$. This means that $Ro/Fr \sim H/\eta$, and so may be taken as an indication of the linearity of the flow, since $\eta \ll H$ is one of the assumptions made when linearising the shallow water equations; see chapter 2 for details. If $Ro/Fr \gg 1$, the flow is very linear; but as $Ro/Fr$ decreases, it becomes increasingly non-linear. So figure 5.21 shows a clear relationship between the effects of smoothing and the linearity of the flow.

- The divergence plot (figure 5.10) shows $R_2$ increasing with increasing $Ro$. The Rossby number gives a measure of the extent to which the flow is in geostrophic
Figure 5.19: Evolution of $\int \delta^2 dA$ for flow regime A, comparing the reference run with the control run and three experimental runs with different values of $\gamma$. Note that the $x$-axis is labelled in time-steps for the reference run, with 1000 time-steps per day.

Figure 5.20: Evolution of $\int \delta^2 dA$ for flow regime E, comparing the reference run with the control run and three experimental runs with different values of $\gamma$. Note that the $x$-axis is labelled in time-steps for the reference run, with 1000 time-steps per day.
balance, so this figure shows that there is a clear relationship between the effect of the smoothing and the balance of the flow, namely that the smoothing only has a beneficial effect in the most geostrophically-balanced cases.

The set of figures relating to $R_\infty$ (figures 5.13–5.18) is more complex:

- Some regimes exhibit minima at particular values of $\gamma$, in certain flow regimes. The minima are concentrated in the region $0 < \gamma < 500$.

- Some of the regimes show a sizeable fall-off in $R_\infty$ for some fields as $\gamma$ increases to its maximum value.

- All but one of the fields ($y$-velocity) shows an improvement in accuracy over the control run for some values of $\gamma$ in some flow regimes.

- There are no clear relationships evident between the sizes of the $R_\infty$ errors and the various dimensionless numbers. A possible reason for this is that the location of the maximum error may not be the same in all flow regimes, and may also vary with $\gamma$.
Since $R_2$ is a measure of the error over the whole domain, rather than the error at one point, it is reasonable to attach more importance to it when interpreting the results. In addition, the assertion, given above, that the divergence may be the most important field in this assessment, leads to the following set of conclusions:

- A universal benefit from smoothing is not demonstrated by this experiment. However, for the most geostrophically-balanced flows, smoothing has a positive impact on the divergence field.
- There is less of a negative effect due to smoothing on the vorticity and PV fields in more non-linear flows.
- Constraining the orographic height at certain points in the domain generally has a negative effect on the $R_2$ measure of the accuracy of the integration.

The conclusions relating to the effect of smoothing on the vorticity and divergence fields are important. They indicate that smoothing is most beneficial (or least detrimental) to the representation of the flow when it is at its most geostrophically balanced and most non-linear. The implications of this are discussed later in this chapter.

From the $R_\infty$ results, the most significant challenge to the broader conclusions that smoothing does not show a benefit comes from the presence of minima in some of the datasets. The most striking examples of this are seen in regime F PV, regime J vorticity, and regimes B, C and J divergence.

In order to clarify the overall picture concerning the $R_\infty$ results, they were averaged over the six flow regimes, to give one dataset for each model field. These are shown in figure 5.22. There is a minimum for the PV at $\gamma = 3$, and minima for vorticity and divergence at $\gamma = 8$. For height, $u$ and $v$, the minimum value of the mean $R_\infty$ falls at $\gamma = 0$. As evidence that there is a non-zero optimal value of $\gamma$, this is not very convincing, but it does provide a basis for choosing a value of $\gamma$ to use while we investigate the sensitivity of the model to $\varepsilon$. 
Figure 5.22: Sensitivity of $R_{\infty}$ to $\gamma$. For each field, the mean of all six flow regimes is shown.

<table>
<thead>
<tr>
<th>$\varepsilon$</th>
<th>CG convergence parameter</th>
<th>number of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>$1.0 \times 10^{-6}$</td>
<td>149</td>
</tr>
<tr>
<td>0.3</td>
<td>$1.0 \times 10^{-6}$</td>
<td>2706</td>
</tr>
<tr>
<td>1.0</td>
<td>$1.1 \times 10^{-6}$</td>
<td>500</td>
</tr>
<tr>
<td>3.0</td>
<td>$2.0 \times 10^{-6}$</td>
<td>1828</td>
</tr>
<tr>
<td>10.0</td>
<td>$2.7 \times 10^{-6}$</td>
<td>3750</td>
</tr>
</tbody>
</table>

Table 5.9: Convergence details for smoothed orography with different values of the Raymond parameter $\varepsilon$. All were generated with $\gamma = 8.0$. 
5.4.3 Sensitivity of the model solution to $\varepsilon$

The next stage in the exploration of parameter space is to investigate the sensitivity of the model solution to $\varepsilon$, the Raymond filter parameter. In this, the value of $\varepsilon$ is varied, while keeping the other parameters (including $\gamma$) constant. Given the desire to examine this sensitivity to $\varepsilon$ in the context of the ridge-height constraint, it is necessary to choose a non-zero value of $\gamma$ to be used for this part of the work. However, since there is no clear evidence of the existence of a non-zero optimal value of $\gamma$, it was decided to use the $R_\infty$ results given in the preceding section to inform the choice. Since the minimum of $R_\infty$ falls at $\gamma = 8$ for both the mean divergence data and the mean vorticity data, that was the value of $\gamma$ chosen. All other aspects of the experiment remain the same as for the previous section.

The smoothed orographies used in this part of the work are listed, along with their CG convergence details, in table 5.9. The entry for $\varepsilon = 1$ refers to the same orography used in the previous section, with $\gamma = 8$.

The model was run with these five orographies, with the six flow regimes used before; $R_2$ and $R_\infty$ were calculated, and these results are shown in figures 5.23–5.28 ($R_2$) and 5.29–5.34 ($R_\infty$). Although no points are plotted on these figures at $\varepsilon = 0$, orography calculated with this parameter would be identical to the unsmoothed mean orography. So, by the definition of $R_2$ and $R_\infty$, all datasets must have $R_2 = R_\infty = 1$ at $\varepsilon = 0$. The results bear many similarities to those discussed in relation to the sensitivity of the solution to $\gamma$. Taking the $R_2$ results first:

- Only four datasets (Regimes A, B and C in divergence, and regime C in height) show values less of $R_2$ less than 1 for any values of $\varepsilon$.
- All the remaining datasets show minima at $\varepsilon = 0$.
- Trends in the size of $R_2$ with $Ro/Fr$ and $Ro$ are seen in the vorticity and divergence respectively.

The results for $R_\infty$ also have similar features to those in the previous section, and are more complex than those for $R_2$:
Figure 5.23: The sensitivity of the $R_2$ error of the height field to $\epsilon$.

Figure 5.24: The sensitivity of the $R_2$ error of the PV field to $\epsilon$. 
Figure 5.25: The sensitivity of the $R_2$ error of the vorticity field to $\varepsilon$.

Figure 5.26: The sensitivity of the $R_2$ error of the divergence field to $\varepsilon$. 
Figure 5.27: The sensitivity of the $R_2$ error of the $x$-velocity field to $\epsilon$.

Figure 5.28: The sensitivity of the $R_2$ error of the $y$-velocity field to $\epsilon$. 
Figure 5.29: The sensitivity of the $R_{\infty}$ error of the height field to $\varepsilon$.

Figure 5.30: The sensitivity of the $R_{\infty}$ error of the PV field to $\varepsilon$. 
Figure 5.31: The sensitivity of the $R_{\infty}$ error of the vorticity field to $\epsilon$.

Figure 5.32: The sensitivity of the $R_{\infty}$ error of the divergence field to $\epsilon$. 
Figure 5.33: The sensitivity of the $R_{\infty}$ error of the $x$-velocity field to $\epsilon$.

Figure 5.34: The sensitivity of the $R_{\infty}$ error of the $y$-velocity field to $\epsilon$. 
• Most datasets show minima at \( \varepsilon = 0 \), though fewer than is the case for \( R_2 \).

• Some datasets (e.g. regime F PV) show obvious minima at \( \varepsilon \neq 0 \).

• Most datasets show \( R_\infty \) generally increasing as \( \varepsilon \) increases. However, in a significant number of these cases, the increase in \( R_\infty \) is very small.

• Some datasets (e.g. regime A divergence) show a very small steady decrease in \( R_\infty \) with \( \varepsilon \).

The broad conclusions to be drawn from these results are exactly those drawn in section 5.4.2 above. The implications of this, and possible reasons for it will be discussed in the light of the original hypotheses towards the end of this chapter.

As with the results relating to the sensitivity of the solution to \( \gamma \), the results for \( R_\infty \) contrast with the results for \( R_2 \) by exhibiting minima in some fields under particular flow regimes with \( \varepsilon \neq 0 \). As argued above, the \( R_\infty \) results are of less significance to the choice of an optimal combination of smoothing parameters than the \( R_2 \) results. However, to gain some sense of whether these results do offer a basis for choosing a non-zero value of \( \varepsilon \), the \( R_\infty \) results were averaged over all flow regimes, and are shown in figure 5.35. It is clear that this figure does not offer much evidence of a non-zero optimal value of \( \varepsilon \). There is a minimum in the mean height dataset at \( \varepsilon = 0.1 \), and there is a very slight minimum in the mean PV dataset at \( \varepsilon = 0.3 \), but the overall conclusion is that the optimal value of \( \varepsilon \) in this experimental configuration is zero.

5.4.4 Sensitivity of the model solution to sea flatness

For reasons outlined at the start of this chapter, it was not expected that the imposition of a flat sea upon the smoothed orography would confer a clear benefit upon the \( R_2 \) and \( R_\infty \) measure of the accuracy of the model integration. However, the sensitivity of the model to the flatness of the sea is still important, as any deterioration in the representation of the flow would have to be considered, along with potential benefits, when determining whether to use the method operationally. To this end, an additional experiment was done to compare the accuracy of the flow for orography with and without the flat sea constraint being applied.
Figure 5.35: Sensitivity of $R_\infty$ to $\varepsilon$. For each field, the mean of all six flow regimes is shown.

The parameter combination used for this experiment was $\gamma = 8, \varepsilon = 1$, which is the point of intersection in parameter space of the lines corresponding to the two previous sections (see figure 5.4). The generation of the orography with this combination of parameters, but without a flat sea, was detailed in chapter 4. The orography with the flat sea constraint applied was generated in the usual way, with a convergence parameter of $10^{-6}$; the convergence took 571 iterations.

The values of $R_2$ and $R_\infty$ were calculated in the same way as before, but, since only two experimental runs are involved, only two values were produced for each error measure, for each flow-regime/field combination. To save space, these are tabulated in tables 5.10 and 5.11, along with the difference due to the imposition of the flat sea constraint. The flow regime/field combinations where there is a decrease in the error due to the flat sea constraint are highlighted in bold.

The differences between the two runs are, in general, quite small, mostly in the third significant figure of the error. Very few flow regime/field combinations show a reduction in $R_2$ with the imposition of the flat sea. Noticeably, the divergence under regimes A, B and C are among those that do. In the $R_\infty$ data, a larger number of rows show an improvement, but, as with the $R_2$ data, the changes are small.
Table 5.10: $R_2$ errors for the six different fields under each of the six different flow regimes. The value of $R_2$ is shown with and without the flat sea constraint being in operation, and with $\varepsilon = 1$ and $\gamma = 8$. The third column shows the change in $R_2$ due to the imposition of the flat sea. Negative values of that change, denoting an improvement in accuracy, are highlighted in bold.
<table>
<thead>
<tr>
<th>Field</th>
<th>flow regime</th>
<th>without flat sea</th>
<th>with flat sea</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>A</td>
<td>1.059</td>
<td>1.237</td>
<td>$1.779 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2.334</td>
<td>2.310</td>
<td>$-2.406 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>$7.429 \times 10^{-1}$</td>
<td>$7.713 \times 10^{-1}$</td>
<td>$2.834 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.034</td>
<td>1.030</td>
<td>$-4.493 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.061</td>
<td>1.052</td>
<td>$-9.014 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>1.082</td>
<td>1.046</td>
<td>$-3.591 \times 10^{-2}$</td>
</tr>
<tr>
<td>PV</td>
<td>A</td>
<td>1.001</td>
<td>$9.968 \times 10^{-1}$</td>
<td>$-4.362 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.041</td>
<td>1.602</td>
<td>$5.614 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1.032</td>
<td>1.614</td>
<td>$5.820 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.004</td>
<td>1.017</td>
<td>$1.284 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>$9.531 \times 10^{-1}$</td>
<td>1.486</td>
<td>$5.327 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>1.033</td>
<td>1.104</td>
<td>$7.069 \times 10^{-2}$</td>
</tr>
<tr>
<td>Vort</td>
<td>A</td>
<td>$9.997 \times 10^{-1}$</td>
<td>$9.910 \times 10^{-1}$</td>
<td>$-8.786 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.020</td>
<td>1.569</td>
<td>$5.498 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1.001</td>
<td>1.566</td>
<td>$5.647 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.004</td>
<td>1.013</td>
<td>$8.886 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.021</td>
<td>1.526</td>
<td>$5.047 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>$9.705 \times 10^{-1}$</td>
<td>1.122</td>
<td>$1.519 \times 10^{-1}$</td>
</tr>
<tr>
<td>Div</td>
<td>A</td>
<td>$9.898 \times 10^{-1}$</td>
<td>$9.963 \times 10^{-1}$</td>
<td>$6.432 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>$8.393 \times 10^{-1}$</td>
<td>$8.172 \times 10^{-1}$</td>
<td>$-2.212 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>$9.331 \times 10^{-1}$</td>
<td>$9.691 \times 10^{-1}$</td>
<td>$3.601 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.046</td>
<td>1.074</td>
<td>$2.849 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.126</td>
<td>1.110</td>
<td>$-1.627 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>1.726</td>
<td>1.716</td>
<td>$-1.053 \times 10^{-2}$</td>
</tr>
<tr>
<td>U</td>
<td>A</td>
<td>$9.905 \times 10^{-1}$</td>
<td>$9.936 \times 10^{-1}$</td>
<td>$3.088 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.446</td>
<td>1.488</td>
<td>$4.190 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1.581</td>
<td>1.635</td>
<td>$5.461 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.042</td>
<td>1.068</td>
<td>$2.579 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.012</td>
<td>1.040</td>
<td>$2.869 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>1.025</td>
<td>1.025</td>
<td>$4.167 \times 10^{-5}$</td>
</tr>
<tr>
<td>V</td>
<td>A</td>
<td>1.001</td>
<td>$9.947 \times 10^{-1}$</td>
<td>$-6.475 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.243</td>
<td>1.974</td>
<td>$7.317 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1.363</td>
<td>1.885</td>
<td>$5.212 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1.016</td>
<td>1.036</td>
<td>$1.910 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>1.009</td>
<td>1.036</td>
<td>$2.710 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>1.023</td>
<td>1.016</td>
<td>$-7.650 \times 10^{-3}$</td>
</tr>
</tbody>
</table>

Table 5.11: $R_\infty$ errors for the six different fields under each of the six different flow regimes. The value of $R_\infty$ is shown with and without the flat sea constraint being in operation, and with $\varepsilon = 1$ and $\gamma = 8$. The third column shows the change in $R_\infty$ due to the imposition of the flat sea. Negative values of that change, denoting an improvement in accuracy, are highlighted in bold.
These results illustrate that there is a cost, in terms of the accuracy of the flow, in imposing a flat sea. However, the change is not universally negative, and is generally of a small size. Thus, the possible benefits of imposing the flat sea, discussed previously, may well outweigh any adverse effects.

### 5.4.5 Discussion

The central question to be answered by the experiments detailed in this chapter is the first of the two given in section 5.1.2:

- Given the use of the Raymond filter, is there an optimal combination of values of $\varepsilon$ and $\gamma$, in the variational scheme, which minimizes the errors in the flow?

The results presented in the preceding sections answer this fairly emphatically: yes, there is an optimal combination of $\varepsilon$ and $\gamma$, namely $\varepsilon = \gamma = 0$. This is an unexpected result. The experiment was designed in the expectation that orographic smoothing would damage the representation of the flow, but that constraining the smoothing process would mitigate this to some degree. However, the message of the results is not only that orographic smoothing does damage the representation of the flow, but also that constraining the smoothing usually makes things worse.

Does this mean that the constrained orographic smoothing described in chapters 3 and 4 is of no use in an NWP or climate modelling context? In order to answer this question, it is necessary to consider several aspects of the experiment, and of the problem we were trying to solve.

The near-grid-scale orography problem is fundamentally non-linear. It involves both the impact of small-scale forcing on the large-scale flow, and the interaction between this forcing and the various physical parameterizations in the model. If this were not the case, and we were only concerned with reducing the presence of near-grid-scale noise in the model because of, say, numerical dispersion errors, we could smooth the orography with impunity. In this case, the near-grid-scales would not be forced, while the well-resolved scales would be unaffected.
Table 5.12: Values of Ro/Fr for the six flow regimes, calculated at $\phi = 90^\circ$ and with $L = 2000$ km.

<table>
<thead>
<tr>
<th>Flow regime</th>
<th>Ro/Fr</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, E</td>
<td>1.35</td>
</tr>
<tr>
<td>J</td>
<td>2.69</td>
</tr>
<tr>
<td>B</td>
<td>6.73</td>
</tr>
<tr>
<td>F</td>
<td>13.5</td>
</tr>
<tr>
<td>C</td>
<td>67.3</td>
</tr>
</tbody>
</table>

The model used in these experiments is a dynamics-only model, and the dynamics which it simulates are very much simpler than those of the real atmosphere. As such, many of the highly non-linear aspects of a full NWP or climate model are not present. The only remaining non-linearity is due to the product terms in the equations of motion. Also, the model is a spectral transform model. This means that the linear evolution of individual spectral components is well-represented. Thus, if the flow is very linear, then all scales within it are well-represented, and so smoothing the orography will cause a reduction in the accuracy of the model solution. Consequently, we can only expect orographic smoothing, and particularly constrained smoothing, to have a beneficial effect if the flow is non-linear in character.

As discussed above, the linearity of a given flow may be characterized by the dimensionless number Ro/Fr, which is roughly equal to the ratio of the mean depth of the fluid to the height of the surface disturbance, assuming geostrophic balance and surface features of a particular length. The smaller the value of this number, the more non-linear the flow is. The values of Ro/Fr for each of the six flow regimes are given in table 5.12. These values represent the upper bound for this number in each regime, since they are calculated with $\phi = 90^\circ$, and for fairly large-scale features.

These features of the near-grid-scale orography problem, and of the model used in the experiment, go a great way towards explaining the behaviour seen in the results, especially those discussed in relation to the sensitivity to $\gamma$. Here, the smoothing was found to be least detrimental to the accuracy of the model solution when the flow was at its most non-linear, which is exactly what is to be expected in the light of the arguments presented above.
The other striking trend in the results is towards greater benefit from smoothing, as measured in the relative divergence error, with more geostrophically-balanced flows. As outlined above, the divergence is of particular interest in this experimental evaluation because it can be thought of as a proxy for the vertical motion, which is the input to important parameterizations in NWP and climate models. Moheshballhojeh and Dritschel (2000) noted that numerical schemes can generate spurious imbalance, particularly near the grid scale. This might mean that near-grid-scale orography generates more imbalance than it should, and so smoothing reduces this. Flow regimes A, B and C are the most balanced to begin with, which may make the spurious imbalance generation more noticeable in these model runs. This could explain the observations earlier about figures 5.19 and 5.20 – the greater imbalance inherent in flow regime E may mask the effects of numerically-generated imbalance. However, these effects may be masked, to some extent, by the presence of an unbalanced component of the flow, generated by the model initialisation, and more work would be needed to confirm the hypothesis given here.

The most important conclusion to draw from this discussion is that the results do not imply that constrained orographic smoothing is detrimental to the performance of an NWP or climate model. The results show that, in the spectral shallow-water model used in the study, $\varepsilon = \gamma = 0$ is the optimal parameter combination, but they also point to the importance of non-linearity and balance in this problem. NWP and climate models are considerably more non-linear, and atmospheric flows are usually more balanced, than is the case in this experiment. The results suggest that increased non-linearity and a more balanced flow may render the constrained orographic smoothing of more use in this context.

Nevertheless, the results highlight the possible disadvantage of smoothing, namely that the accuracy of the model solution may be adversely affected. As with all aspects of numerical model design, the costs and benefits of different approximations have to be weighed against one another, and the case of orographic smoothing is no different.

Finally, it should be remembered that the constraint of barrier height depends on the generation of a mask of ridge points. The algorithm developed in chapter 4 is effective in selecting some major orographic ridges, but it is not perfect. While it is
Figure 5.36: Sensitivity of $R_2(h^*)$ to $\gamma$, for regime A, and showing the results obtained for different choices of diffusion.

quite easy to argue that the height of a particular mountain range (for instance, the Andes) is important for the nature of the flow, this is a piecemeal approach that may not lead to an objective assessment of the suitability of a particular mask of points. Possible ways of improving this selection procedure are discussed in chapter 6; the important point is that an unsuitable mask may have an adverse effect on the accuracy of the solution, without in any way invalidating the overall approach.

5.4.6 The impact on the results of the choice of diffusion

As mentioned above, the application of diffusive terms to the divergence and height fields is not standard practice, and represents an unusual choice. More commonly, the diffusion is applied to the divergence and vorticity, and so part of the experiment was rerun to evaluate the impact on the conclusions of the choice of diffusion. The regime A reference run, the corresponding control run, and selected experimental runs with different values of $\gamma$ were rerun, applying the diffusion to the divergence and vorticity. The values of $R_2$ were recalculated and the resulting sensitivity to $\gamma$ was compared with the results obtained in the main part of the experiment.

Typical results from this comparison are shown in figures 5.36 and 5.37 (the $R_2$
errors on height and $x$-velocity, respectively). In both cases shown, the $R_2$ error is smaller when the diffusion is applied to $\delta$ and $\zeta$ than when it is applied to $\delta$ and $h^*$. However, the form of the curve is unchanged by the alteration in diffusion, showing the same dependence on $\gamma$ in both cases. There is no reason to suppose that this sensitivity to the diffusion will not be the same for all variables and all flow regimes. Since the conclusions derived from the main experiment depend only on the form of the dependence on $\gamma$ (and later on $\varepsilon$), these additional results show that those conclusions are not affected by changing the variables to which diffusion is applied.

5.5 Summary and conclusions

In this chapter, the variational smoothing technique developed in chapters 3 and 4 has been evaluated thoroughly in a non-linear, global shallow water model. The evaluation was conducted under a wide range of flow regimes, and the effect of the various smoothing constraints on the flow was characterized using two global error measures.

The conclusions drawn from these experiments may be summarized as follows:
• The balance of evidence does not point to a benefit from smoothing in the non-linear, global, spectral shallow water model used here.

• Smoothing has a negative effect on some aspects of the flow.

• The evidence only weakly supports the imposition of barrier-height and flat-sea constraints.

• However, smoothing has a more beneficial, or less detrimental, effect in more non-linear, more geostrophically-balanced flows.

• Moreover, smoothing seems to be most effective at reducing errors in the divergence, already identified as one of the most important fields in relation to the problem in the context of NWP and climate modelling.

The practical implications of these conclusions is that in NWP and climate models, where flow is generally much more non-linear and more balanced, the use of smoothing is likely to be useful. The experiment here does not strongly support the use of extra constraints in orographic smoothing, but that does not preclude the possibility of a benefit from constrained smoothing being seen in a full NWP or climate model. However, an additional investigation would have to be undertaken in that context to resolve the question. Finally, this experiment does highlight the loss in accuracy which could result from smoothing, and which would have to be taken into account when deciding whether to use orographic smoothing for an NWP or climate modelling application.
Chapter 6

Discussion and conclusions

6.1 Thesis summary

The overall aim of this thesis has been to determine the orographic smoothing optimal for use in NWP and climate models. In approaching this, two things were seen as fundamental: the need to study the problem in the context of simplified models, and the desire to alleviate some of the negative aspects of orographic smoothing. While the study has not provided a generalized 'recipe' for optimal smoothing in a full NWP model, it has contributed significantly to the understanding of the nature of the problem and how it may be addressed, as well as furnishing designers of atmospheric models with a powerful new orographic smoothing method.

In chapter 2, the problem of near-grid-scale orography was studied in the context of a very simple model. The one-dimensional, linear shallow water model used had the advantage of being easy to analyse, and facilitated the study of both steady and transient forced components of the flow. The results showed that there are problems with the representation of near-grid-scale forcing in this model, but the evidence presented leads to the conclusion that the near-grid-scale orography should be amplified rather than filtered out. Given that there exists much evidence to support the smoothing of orography in NWP models, the main conclusion drawn from the work was that the model used here was not representative enough of a full NWP model to reproduce the problem under study. Aspects of an NWP model missing from the model used here include higher dimensionality, non-linearity, and
parameterizations of sub-grid-scale processes. The low dimensionality of the model and its lack of non-linearity were postulated as its main inadequacies. Nevertheless, the work presented in chapter 2 supported the findings of Davies and Brown (2001) that scales shorter than approximately six grid-lengths are those which are most significantly misrepresented, and whose removal from the orographic field might thus be beneficial.

Chapter 3 concerned the mathematical formulation of a new, variational smoothing method. The generalization of a simple variational smoother to allow the emulation of a large class of linear filters was shown, and the consequences of discretization were explained in terms of the eigenfunctions of differential operators. In particular, it was shown that the effect of the filter by Raymond (1988), used for orographic smoothing by Webster et al. (2003), can be replicated by the variational method. However, analysis showed that predicting the spectral response of the method on the sphere is impractical, due to the discrepancies between the eigenfunctions of the continuous and discretized $\nabla^2$ operators; it was argued, though, that the discretized eigenfunctions would be more relevant to smoothing in a numerical model than their continuous counterparts. The ease with which this variational scheme can be extended to include other constraints of interest, such as the maintenance of a flat sea or the height of barrier ridges, was also explained; this is what promises to make the method so useful.

Aspects of the practical implementation of the variational method was discussed in the first part of chapter 4. In particular, attention was given to several possible methods of identifying the barrier ridges whose heights we might wish to constrain; for the testing and evaluation of the variational smoother, a recursive ridge search method was chosen. However, it was recognized that the choice of peak mask requires more research, especially to take into account the prevailing flow regimes at different locations on the Earth. In the second part of chapter 4, the predicted spectral properties of the variational smoothing method were successfully verified. In addition, the flat-sea and barrier-height constraints were demonstrated, and it was shown that their use does not significantly negate the effect of the smoothing near the grid-scale.

Finally, chapter 5 detailed the evaluation of the variational smoothing method in a
Chapter 6 Discussion and conclusions

spectral, non-linear, global shallow water model. Care was taken to evaluate a range of parameter combinations under a variety of flow regimes. Dimensionless numbers were used to characterize these regimes, and proved invaluable in the interpretation of the results. The main conclusion from this work was that, in the model used, smoothing does not have a generally beneficial effect on the accuracy of the model flow. However, the smoothing has the most positive, or least negative, effect when the flow is most non-linear (diagnosed by Ro/Fr) and most geostrophically-balanced (diagnosed by the Rossby number). This is a significant result. Non-linearity, both dynamical, and due to the interaction between the dynamics and parameterizations of the model, is thought to be important in the problems with near-grid-scale orography which we set out to solve by orographic smoothing. The results confirm that dynamical non-linearity, as diagnosed by Ro/Fr is important in determining the usefulness of smoothing. NWP and climate models are considerably more non-linear than the model used here, because of having more complex dynamics and interactions between the dynamics and parameterizations. Furthermore, real atmospheric flows are generally more balanced than the flows studied here. Taken together, this suggests that we should expect the benefits of smoothing to be greater in an NWP context than in this simple model. Finally, smoothing was seen to be most effective at reducing errors in the divergence field, which is important, as divergence can be thought of as a proxy for vertical motion, which is instrumental in the formation of the grid-point storms associated with near-grid-scale orography in NWP models. Nevertheless, despite the evidence that smoothing may be beneficial in NWP models, this study highlights the possible negative impacts of smoothing. It also only weakly supports the imposition of extra constraints on the smoothing.

6.2 Applications of this work

The work presented in this thesis is most immediately relevant to the representation of orographic forcing in NWP and climate models, particularly where finite-difference schemes are used, as in the Met Office Unified Model. The insight gained into the nature of the problems encountered with near-grid-scale orography gives confidence that orographic smoothing can be employed beneficially, and the new variational smoothing method presented here significantly augments the range of
smoothing tools available to designers of numerical models. Additionally, the orographic smoothing method is very versatile, and could be applied in any context where a well-defined linear filter is required to operate in the presence of other constraints. For instance, if calculating a precipitation climatology, noisy data might be smoothed beneficially, but with the requirement that the field has no negative values.

6.3 Remaining questions and future work

Several questions remain outstanding from the work presented in this thesis, and could form the basis for further exploration. The most important of these concern the possible use of the new variational smoothing method in a full NWP or climate model. The broad question is whether the method can be used beneficially in such a model, and, on the basis of the present work, and that of Webster et al. (2003), there is strong reason to suppose that it would be. However, before using orography smoothed in this way in an operational weather prediction model, we need to answer these subsidiary questions:

- How can we effectively evaluate the variational method in a way which is most relevant to the workings of an NWP model?

As explained in chapter 1, the evaluation of orographic smoothing in an NWP model is problematic, partly because the establishment of a reference solution (‘truth’) is difficult, but mostly because sub-grid parameterizations will have been tuned to operate best with the pre-existing orography. However, a very simple model, such as the one used in this thesis to evaluate the variational smoother, is not representative enough of an NWP model to be used as the sole piece of evidence supporting the operational use of the new method. Probably the best context in which to undertake the evaluation is therefore one that replicates as near as possible the dynamical properties of the NWP model we are interested in, but with as few sub-grid parameterizations as possible. Dynamical core tests are a well-established means of evaluating numerical methods, and so the evaluation of orographic smoothing in the context of
a well-understood dynamical core would seem to be a consistent approach to the problem. In these tests, particular attention would need to be given to the accuracy of the fields, such as the vertical velocity, which comprise the input to the sub-grid parameterizations of such things as convection, etc. Nevertheless, it would also be instructive to investigate the causes of grid-point storms in more detail. By looking at orographic smoothing in the context of both a dynamical core and the full model, the hypothesis that they depend on interactions between the dynamics and parameterizations could be tested.

- **How should the loss of gravity-wave drag due to the smoothing be compensated for?**

  This topic was addressed by Webster et al. (2003), and their solution seems very sensible. Sub-grid-scale orographic drag parameterizations take as their input the standard deviation of the orography within each grid box. The approach adopted by Webster et al. is to recalculate this standard deviation field relative to the smoothed orography, rather than the pre-existing mean orography, thus incorporating the lost resolved roughness into the parameterised, sub-grid-scale component. Nevertheless, before adoption, this procedure should be evaluated in the dynamical core tests proposed above.

- **In applying the barrier-height constraint to the smoothing, which peak mask should be used?**

  As discussed in chapter 5, the choice of peak mask used in this thesis for the imposition of the barrier-height constraint was made on a reasoned, though rather subjective, basis. While the various methods discussed for the generation of the peak mask were objective in themselves, the choice of the method and parameter combination to be used was made subjectively. This choice was made on the basis that certain mountain ranges are known to be important for global climate, and should therefore perhaps have their height constrained. However, it would be good to have a method for determining the peak mask on the basis of a knowledge of the areas where orographic smoothing has the most negative effect on the representation of the flow. To achieve this, a systematic study should be undertaken in the context of the dynamical core tests already suggested; first, the areas of maximum error could be identified, and
then the importance of each mountain range evaluated in turn by constraining their heights in successive model runs.

- **Is there a risk of introducing systematic errors into climate simulations by the use of constrained smoothing?**

The use of the new variational smoothing method in the context of a weather prediction application may be evaluated in the long term by the methods of forecast verification used presently in the industry. In the field of NWP, the usefulness of the forecast is what counts; this is relatively easy to assess, and may be done on a daily basis. However, when using numerical models for climate prediction, the process of quantitative model verification is both more important and less easy. Of particular concern is the fact that, unlike a linear filter, the constrained smoothing method does not preserve the volume of the orography, and this could introduce a systematic error into the model’s climate. Determining whether this is the case would be quite difficult, due to the computational expense of the long model runs involved, and also the previously discussed problems with evaluating smoothed orography in a full NWP model.

### 6.4 Final thoughts

Considering the evaluation made in chapter 5, and the outstanding questions posed above, it is clear that the use of the variational smoothing method in an NWP context is a real, practical possibility. The remaining questions are answerable using techniques and models already available, and there is no reason why the necessary evaluation of the method should not be undertaken in the near future. However, the use of the method in a climate modelling context needs to be approached with more caution. It may turn out that there are no adverse effects on model climate from its use, but it would be wise to check thoroughly. For, although weather forecasts are important to all sections of society, business and government, our studies of possible future climate look set to be important to the development of human society in the very long term. The work presented in this thesis is only a small contribution to
the range of tools available to climate modellers, but, given the importance of their task, it may, ultimately, be of benefit to us all.
Bibliography


Côté, J., S. Gravel, and A. Staniforth (1995). A generalised family of schemes


But in my spirit will I dwell,
And dream my dream, and hold it true;
For tho’ my lips may breathe adieu
I cannot think the thing farewell.

– Tennyson (In Memoriam)