A Vortex Dynamics Perspective on Stratospheric Sudden Warmings

A dissertation presented

by

Nicolas Joss Matthewman

to

The Department of Mathematics

for the degree of

PhD

in the subject of

Applied Mathematics

UCL

London

September 2009

I, Nicolas Joss Matthewman, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Thesis advisor
J. Gavin Esler

Author
Nicolas Joss Matthewman

A Vortex Dynamics Perspective on Stratospheric Sudden Warmings

Abstract

A vortex dynamics approach is used to study the underlying mechanisms leading to polar vortex breakdown during stratospheric sudden warmings (SSWs). Observational data are used in chapter 2 to construct climatologies of the Arctic polar vortex structure during vortex-splitting and vortex-displacement SSWs occurring between 1958 and 2002. During vortex-splitting SSWs, polar vortex breakdown is shown to be typically independent of height (barotropic), whereas breakdown during vortex-displacement SSWs is shown to be strongly height dependent (baroclinic).

In the remainder of the thesis (chapters 3-7), a hierarchy of models approach is used to investigate a possible resonant excitation mechanism which is responsible for the vortex breakdown seen in our observational study. A single layer topographically forced vortex model is shown to exhibit vortex-splitting behaviour similar to that observed during SSWs. Two analytical reductions, the first a fully nonlinear analytical model of an elliptical vortex in strain and rotation velocity fields, the second a weakly nonlinear asymptotic theory applied to a topographically forced vortex, show that vortex-splitting in the model occurs due to a self-tuning resonance of the vortex with the underlying topography.

Resonant excitation of an idealized polar vortex by topographic forcing is then

investigated in a three-dimensional quasi-geostrophic model, with emphasis on the vertical structure of the vortex during breakdown. It is shown that vortex breakdown similar to that observed during displacement SSWs occurs due to a linear resonance of a baroclinic mode of the vortex, whereas breakdown similar to that observed during splitting SSWs occurs due to a resonance of the barotropic mode. The role of self-tuning in these resonant behaviours is then discussed in relation to the analytic reductions of the single layer model.

Contents

Citations to Previously Published Work

The work of Chapter 2 has appeared in the following paper:

> "A New Look at Stratospheric Sudden Warmings. Part III. Polar Vortex Evolution and Vertical Structure", N. J. Matthewman, J. G. Esler, A. J. Charlton, L.M. Polvani, *J. Clim.* **22**, no. 6, 2009,

written in collaboration with J. G. Esler at University College London, A. J. Charlton-Perez at Reading University , and L. M. Polvani at Columbia University.

Acknowledgments

I would like to thank Dr. Gavin Esler for his support and encouragement throughout the course of my PhD. I would also like to thank Prof. Lorenzo Polvani for making my visit to Columbia enjoyable and worthwhile, and for supporting me during my stay. I'm also grateful to the Natural Environment Research Council, UCL Graduate School, the University of London and UCL Mathematics Department for their financial support over the course of my PhD.

Thanks to Ed Hoyle, Fern Scott and Kevin East for useful discussions about my work, even if it was mainly me talking to myself most of the time. In particular, I'd like to thank Kev for going enormously out of his way to help out towards the end... it was much appreciated.

Finally, I would like to say a huge thank you to Anna, for putting up with the late nights in the office, months away from home, and irrational outbursts without a single complaint. Quite simply, I just wouldn't have got here without you.

Dedicated to Kirby

Chapter 1

Introduction

1.1 The stratospheric polar vortex

Conventionally, the atmosphere is divided up into several layers. The lowest layer is the troposphere, which is characterized by an overall decrease in temperature as height increases. At the tropopause, approximately 10-20 km above the Earth's surface, there is an abrupt change in the vertical gradient of temperature, and above this level temperature increases slowly with height. The main reason for this reversal in the vertical temperature gradient is the presence of large concentrations of ozone (O_3) above this region, with the absorption of solar radiation by the ozone being responsible for the increase in temperature. The layer in which the temperature slowly increases with height is referred to as the stratosphere and is characterized by strong stratification, meaning that it is highly stable to vertical displacements of fluid. The flow in the stratosphere is consequently layerwise two-dimensional to a good approximation. The temperature continues to increase until the stratopause is encountered

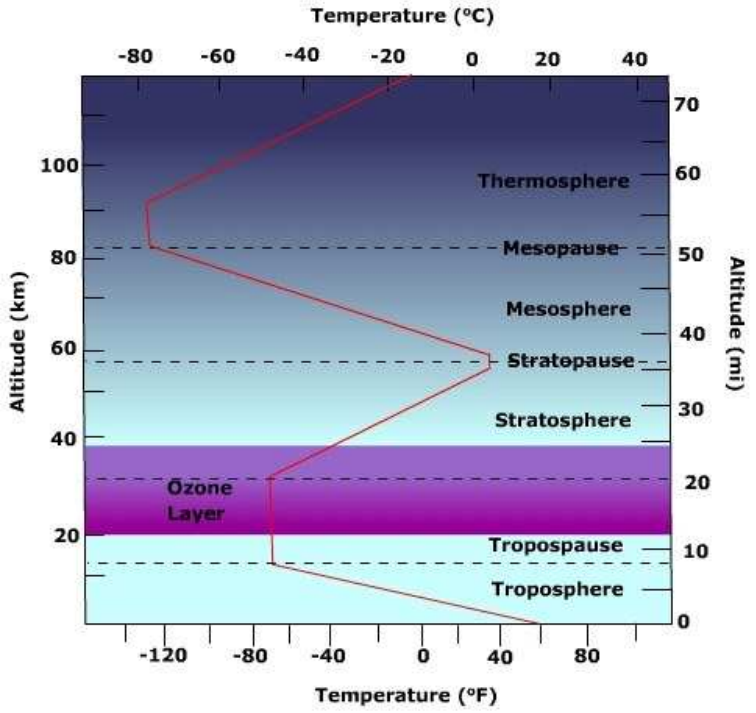

Figure 1.1: Schematic showing the structure of the atmospheric layers, with the mean vertical temperature profile corresponding to radiative equilibrium shown by the red line.

at an altitude of approximately 50 km, at which point the vertical temperature gradient reverses once again, with the region directly above the stratopause being the mesosphere. The large concentrations of carbon dioxide (CO_2), which emit radiation into space, are responsible for the decrease in temperature with height which is observed in the mesosphere. A schematic of this atmospheric structure is shown in Fig. 1.1.

In the winter hemisphere, there is a strong pole-equator temperature gradient throughout the extratropical stratosphere, as shown by the zonal-mean temperature plots of Fig. 1.2 for Northern and Southern Hemisphere winters. From thermal wind balance, this latitudinal temperature gradient is in balance with strong vertical gradi-

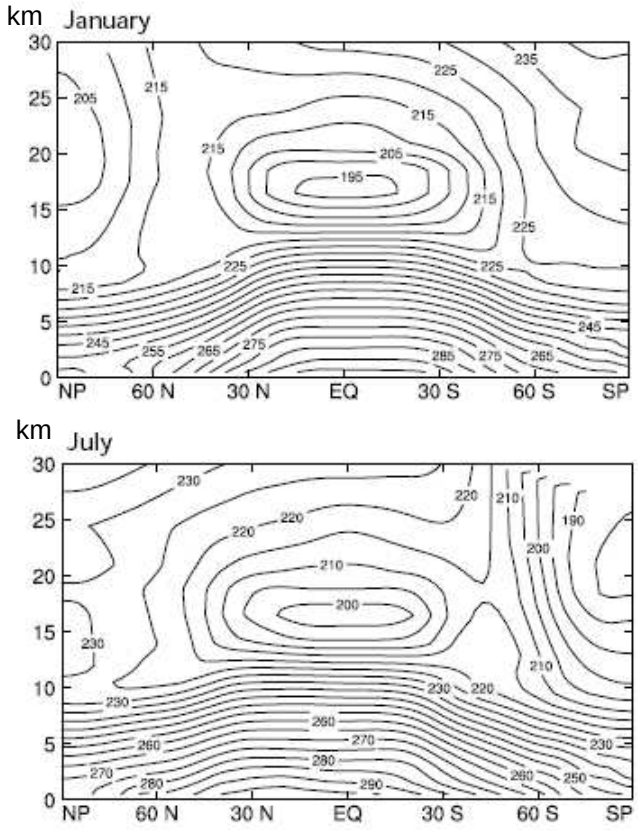

Figure 1.2: Latitude-height cross section of zonal-mean temperature in January (top) and July (bottom). The contour interval is 5 K (figure from *Atmospheric Circulation* course notes by David A. Randall).

ents in the zonal (east-west) wind and hence strong westerly circumpolar jets exist in the winter polar stratosphere. The ultimate cause of the latitudinal temperature gradient, and hence the polar jets, is the differential in net radiative heating between the equator and polar regions. However, it is important to note that the active dynamics of the stratosphere lead to the stratosphere being far from radiative equilibrium, i.e. large scale motion means heating and cooling due to absorbed and emitted radiation respectively need not balance locally.[1] It is this region of strong westerly flow in the winter polar stratosphere that is referred to as the stratospheric polar vortex.

The stratospheric polar vortex forms at the start of the winter in each hemisphere, and breaks down during the following spring, when the net radiative heating differential which drives the westerly flow weakens. The bottom of the polar vortex, or sub-vortex, exists at heights of approximately $15 - 20$ km in the lower stratosphere, with the main vortex structure existing in the region between 20 km and the stratopause at ~ 50 km (Schoeberl and Strobel 1978). The zonal jet associated with the polar vortex is shown in the zonal-mean zonal wind ($\bar{u}$) plots in Fig. 1.3, where an increase in westerly $\bar{u}$ is seen at 60° N in January and 60° S in July above 15 km, with the peak occurring at heights greater than 30 km in both cases. In addition to the circumpolar jet which characterizes the polar vortex, a subtropical jet between 20°-40° at a height of 7-15 km exists in both the winter and summer hemispheres, as seen in Fig. 1.3. The subtropical jet in the winter hemisphere is stronger and located closer to the equator when compared to the summer hemisphere subtropical jet, with

[1] As explained in Andrews et al. (1985, chapter 7), the radiative disequilibrium is maintained by a large scale meridional (latitude-height) circulation, known as the Brewer-Dobson circulation, which is understood to be driven by the active dynamics of the stratosphere.

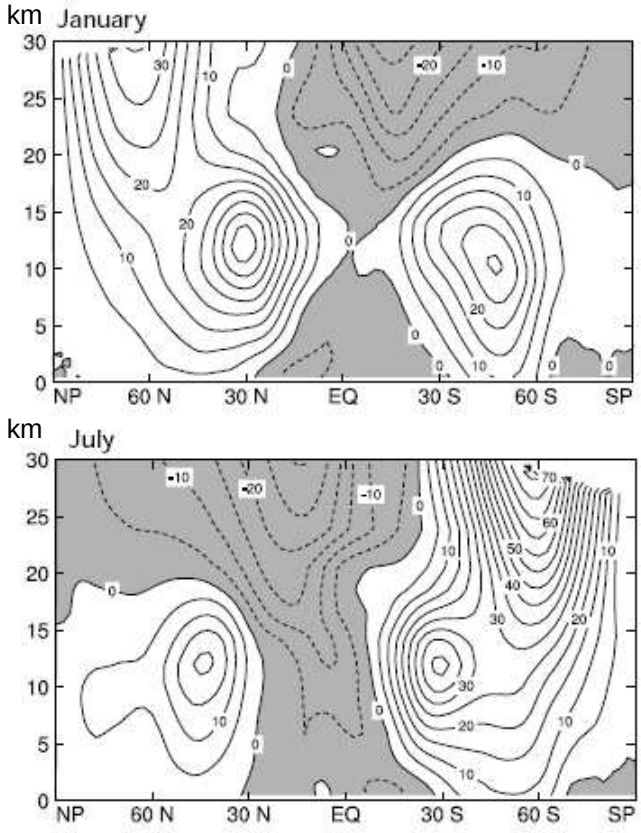

Figure 1.3: Latitude-height cross section of zonal-mean zonal wind in January (top) and July (bottom). The contour interval is 5 m s^{-1}, and easterlies are shaded (figure from *Atmospheric Circulation* course notes by David A. Randall).

the zonal-mean zonal wind at heights above the summer hemisphere jet being weak and easterly.

In both hemispheres, Waugh and Randel (1999) showed that the polar vortex has on average an approximately circular cross section and is centred slightly off the pole. However, it was shown that the location and shape of the vortex differs marginally between the two hemispheres, with more variability observed in the Northern Hemisphere polar vortex compared to its southern counterpart.

In the middle and lower atmosphere, quasi-horizontal layerwise stirring of chemicals occurs due to large scale turbulence. To a good approximation this stirring takes place on isentropic surfaces, or surfaces of constant entropy S, or equivalently constant potential temperature θ which is related to S by $S = c_p \ln \theta$, where c_p is the specific heat of dry air at constant pressure. Note that for an ideal gas, such as dry air, the potential temperature is related to the temperature T and pressure p by $\theta = T(p_0/p)^{R/c_p}$, where R is the ideal gas constant for dry air, and p_0 is a reference pressure usually taken to be 10^3 hPa.

In situ data from the Arctic Airborne Stratospheric Experiment (AASE) and the Antarctic Airborne Ozone Experiment (AAOE) have shown that for the polar vortex in both hemispheres, the chemical composition of the airmass within the vortex and that outside the vortex are distinct (see Loewenstein et al. 1989; Proffitt et al. 1989, and the review issues of *Journal of Geophysical Research* vol. **94** D14 1989, *Geophysical Research Letters* **17** no. 4 1990, *Science*, **261** 1993). In particular, anomalously low ozone concentrations were seen to coincide with the location of the polar vortex, with the Southern Hemisphere exhibiting the largest anomaly (Proffitt

et al. 1989).

A useful diagnostic for stratospheric flows is that of Ertel's potential vorticity (PV hereafter), which is a materially conserved dynamical quantity in the absence of diabatic heating and frictional processes. Due to the relatively short time scales on which advection occurs in the winter polar stratosphere, when compared to the time scales of diabatic processes, it is generally assumed that PV is a materially conserved quantity on timescales of approximately 5 to 10 days.

Observations of the distribution of PV in the stratosphere have shown that the polar vortex is characterized by PV values which are anomalously high when compared to that of the surrounding "surf zone" (McIntyre and Palmer 1983, 1984; Nash et al. 1996). The surf zone exists throughout the stratosphere and on average extends in latitude from the edge of the polar vortex at approximately $\pm 70^{\circ}$ latitude to the subtropics (approximately $\pm 20^{\circ}$). It is so named due to the large scale dynamical stirring of chemical species in this region, in a process which has been compared to waves breaking on a beach (McIntyre and Palmer 1983) and which also results in the PV being well mixed in this region.

The AASE and AAOE missions, have also shown that the transition between surf zone and polar vortex air masses is characterized by steep gradients in trace chemicals, as well as PV and other long lived chemical tracers, for example nitrous oxide (N_2O) (see Fahey et al. 1990; Loewenstein et al. 1990, and the issues dedicated to the AASE and AAOE given above). The polar vortex edge is defined by the region of these steep gradients, and acts as a transport barrier between the cold, chemically isolated airmass within the vortex and the warmer, vigorously mixed air of the surf zone (e.g.

Hartmann et al. 1989). Transport of air out of the polar vortex and into the surf zone still occurs however, via small scale turbulence in the vicinity of the vortex edge, weak transport due to radiative processes, and most significantly, ejection of filaments of vortex air (with typical scales of order 100 km).

For the main part of the polar vortex, filaments ejected in the middle stratosphere (850 K) were shown by Clough et al. (1985), McIntyre and Palmer (1983, 1984) and Waugh et al. (1994) to be quickly mixed with the air of the surrounding surf zone. During the AASE, strong internal mixing was observed in the polar vortex interior through which air at the centre of the vortex was transported to the vortex edge. Combined with PV filament ejection and mixing in the middle stratosphere, this resulted in air from deep within the vortex interior being advected out of the vortex and mixed into the surrounding surf zone. In this way, Waugh et al. observed that during the AASE approximately 50% of the interior air of the Northern Hemisphere polar vortex was advected out of the vortex and mixed into the surf zone, such that the transport barrier of the vortex edge could more accurately be described as a "leaky" transport barrier. Waugh et al. found that air which was ejected from the vortex in this way was replenished by downwelling from the upper stratosphere and other diabatic processes, such that the cross sectional area of the vortex remained on the whole constant despite these ejections.

For the sub-vortex in the lower part of the stratosphere (around 450 K), this process of filamentation is somewhat different. Waugh et al. (1994) found that filaments in this region were of significantly smaller scale than in the main vortex and, once ejected, do not mix with the air of the surf zone but encircle the polar vortex core,

reducing the sharpness of PV gradients at the vortex edge.

In comparison to these extrusions of air from the polar vortex into the surrounding surf zone, Plumb et al. (1994) found that over the same aircraft mission period, *intrusions* of air from the surf zone into polar vortex were quite modest.

1.2 Stratospheric sudden warmings

Stratospheric sudden warmings are phenomena occurring in the winter polar stratosphere in both hemispheres. The World Meteorological Organization (WMO) defines a stratospheric sudden warming as occurring whenever the meridional gradient in 10 hPa zonal-mean temperatures between 85° and 60° is positive for more than 5 days (Andrews et al. 1985).[2] This is marked by a sharp increase in temperature over the polar cap, in some cases as large as 40 K, an example of which is shown in panel A of Fig. 1.4 for the sudden warming which occurred in February 1979. Within this definition, there are four sub-classifications of stratospheric warming (for a thorough review of these warming classifications see Labitzke 1977):

1. A major warming occurs when reversal of the meridional temperature gradient is accompanied by a reversal of the zonal-mean zonal wind at 60° latitude on the 10 hPa pressure level, reversing from the westerly flow of the climatology to easterly flow (Andrews et al. 1985, see also panel B of Fig. 1.4). A distinguishing feature of major stratospheric sudden warmings is that they are accompanied by a dramatic breakdown of the polar vortex, followed by a gradual recovery of

[2]Latitudes here, and in the description of warming classifications, refer to latitudes in the winter hemisphere, for example 85°-60° N in January, 85°-60° S in August.

the polar vortex to its pre-warming state. The polar vortex behaviour during this breakdown will be discussed shortly.

2. Minor warmings occur when the meridional temperature gradient reverses from its climatological state, as described in the WMO definition, but there is no reversal of the zonal-mean zonal wind or breakdown of the polar vortex during minor warmings.

3. Final warmings occur at the end of every winter, as the polar night draws to an end and temperatures in the stratosphere steadily increase. As stratospheric temperatures increase, meridional temperature gradients driving the strong westerlies in the stratosphere disappear, leading inevitably to a reversal of the zonal-mean zonal wind, accompanied by complete destruction of the polar vortex. Unlike major sudden warmings however, the polar vortex does not reform as the westerlies do not return, hence the term "final warming".

4. The final sub-classification is that of Canadian Warmings which are only observed in the Northern Hemisphere. These events are identified using the major and minor warming criteria, but were determined by Labitzke (1977) and Labitzke and Naujokat (2000) to be distinct from major and minor warmings as they are characterized by a displacement of the polar vortex from the pole by the Aleutian anticyclone, typically occurring earlier than other types of warming in November as opposed to December-February.

A comprehensive list of all winters in which major and Canadian warmings have occurred is given in Labitzke and Naujokat (2000), along with other information

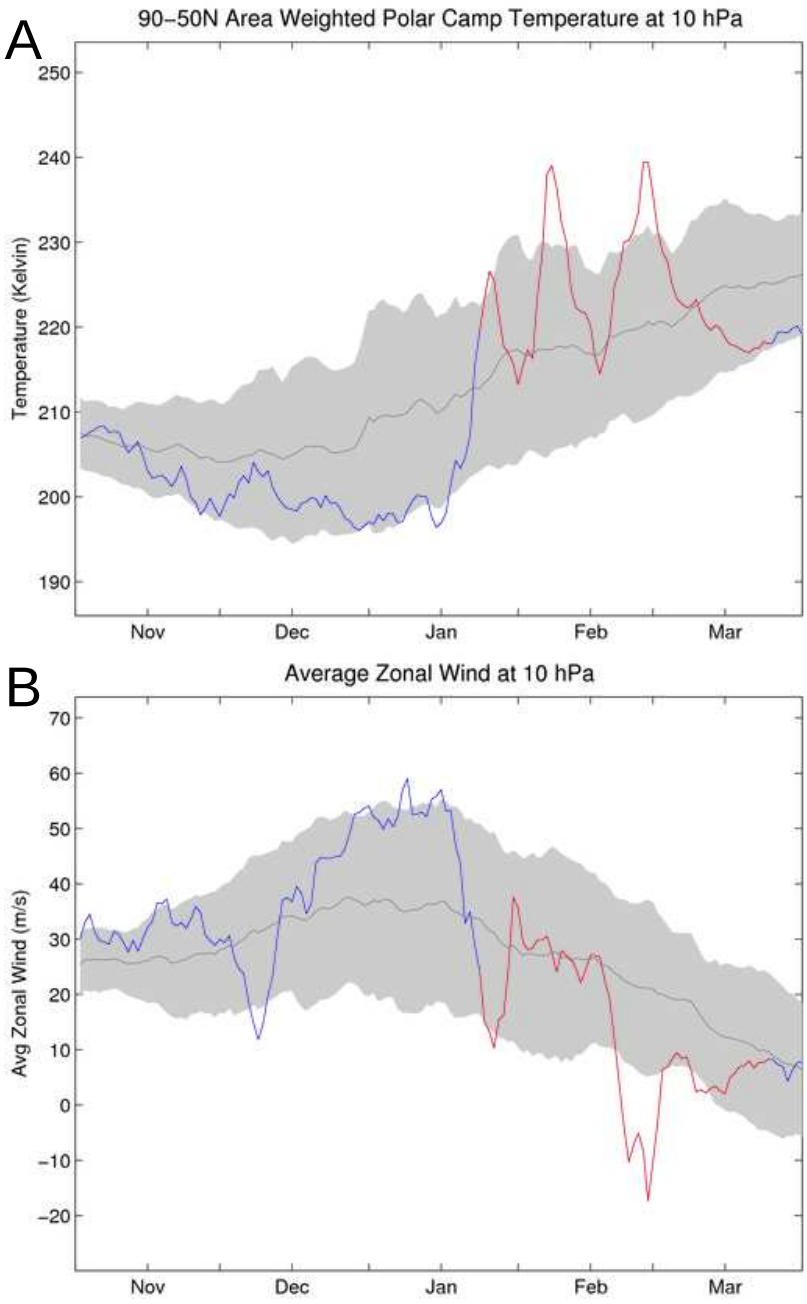

Figure 1.4: Panel A: Area weighted polar cap temperature on the 10 hPa pressure surface during the 1978-1979 Northern Hemisphere winter. Panel B: zonal-mean zonal wind at 60° N on the 10 hPa pressure surface during the 1978-1979 Northern Hemisphere winter. In both panels, the red portion of the curve represents the warming period defined in Charlton and Polvani (2007), and the thin grey line shows the climatological polar cap temperature and zonal-mean zonal wind for all winters in the period 1958-2002. The shaded region shows plus/minus one standard deviation, calculated from a climatology of all winters in the period 1958-2002. Figures produced using ERA-40 re-analysis data and taken from the Stratospheric Sudden Warmings Online Library http://www.appmath.columbia.edu/ssws/index.php.

including whether the final warming for each winter was observed to occur earlier or later than usual.

Recently, Limpasuvan et al. (2004) performed a study in which observational data were used to analyze typical behaviour during stratospheric sudden warmings. Using data from major and minor warming events in the period 1958-2001, Limpasuvan et al. (2004) created composites of zonal-mean zonal wind and zonal-mean temperature, to investigate typical behaviour in the stratosphere and upper troposphere. It was found that sudden warmings exhibit a life cycle spanning approximately 75 days. By setting a central date for each warming, defined as being when conditions similar to those of the WMO warming definition are satisfied, the life cycle was divided into five stages which were *onset* (-37:-23 days), *growth* (-27:-8 days), *mature* (-7:+7 days), *decline* (+8:+22 days) and *decay* (+24:+37), where all days are given relative to the central date. It was found that during the onset, growth and mature stages, a weakening of the zonal-mean zonal flow and warming of the polar cap descended from the middle to lower stratosphere. Anomalies of the zonal-mean zonal wind and zonal-mean temperature from their respective climatologies peaked during the mature stage, before each returned to their climatological value during the decline and decay phases. Similar to the findings of Baldwin and Dunkerton (1999, 2001), it was noted that the largest disturbances occurring in the stratosphere descended into the troposphere as the life cycle developed.

The work presented in this thesis will focus solely on the dynamics responsible for major warmings. Due to their dependence on the radiative annual cycle, the dynamics of final warmings are distinct to those of major warmings. In particular, a

single final warming occurs at the end of every winter in both hemispheres, whereas the occurrence of major warmings is more unpredictable, with winters having at most one or two major warmings or none at all. A study of the dynamical behaviour of final warmings can be found in Black and McDaniel (2007). As we are only considering major warmings, the acronym SSW for stratospheric sudden warmings is taken in the remainder of this work to refer exclusively to major mid-winter warming events.

Traditionally, SSWs have been labelled either "wave-1" or "wave-2" events. This naming convention comes from observations on pressure surfaces of geopotential disturbances, where the geopotential is the potential energy per unit mass relative to sea level, which have either wavenumber-1 or wavenumber-2 as the dominant Fourier mode in the longitudinal structure. The way in which the polar vortex breaks down during these two classifications of SSW is distinctly different. During wave-1 SSWs there is typically an increase in displacement of the polar vortex from the pole, while wave-2 SSWs are characterized by a complete split of the polar vortex into two separate fragments. An example of this difference in behaviour is given by observing shaded contours of vertically weighted PV on the 840 K isentropic surface during the wave-1 SSW occurring in January 1987 and wave-2 SSW occurring in February 1979, as shown in Fig. 1.5.[3] For the disturbed vortex states which typically occur during polar vortex breakdown associated with SSWs, the overall PV distribution is seen to depart significantly from that corresponding to zonal flow around the pole, meaning that classifications relying purely on zonal averages, and waves defined as

[3]The vertically weighted is derived using the data and methodology described in chapter 2. It is noted that the vertically weighted PV distribution on each isentropic surface is identical to that of Ertel's PV.

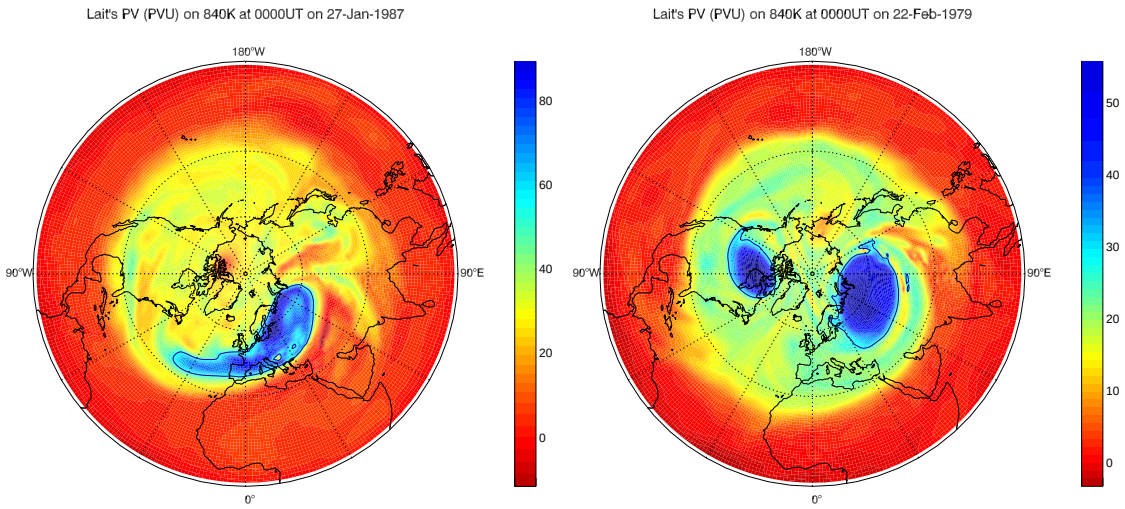

Figure 1.5: Contours of vertically weighted PV on the 840 K isentropic surface during the wave-1 SSW in January 1987 (left panel) and wave-2 SSW in February 1979 (right panel). In both panels, the dark blue contour denotes the vortex edge, which is calculated using the methodology of chapter 2, and blue shading corresponds to PV belonging to the polar vortex interior.

disturbances relative to these, may obscure a realistic and physically based picture of the vortex structure (O'Neill and Pope 1988; Waugh 1997).

Recently, a new algorithm for an objective classification of SSWs was introduced by Charlton and Polvani (2007), in which the observed absolute vorticity distribution of the vortex was used to classify all Northern Hemisphere SSWs in the period 1958-2002 as either *vortex-displacement* or *vortex-splitting* SSWs, closely corresponding to the wave-1 and wave-2 categories above. The algorithm of Charlton and Polvani (2007) identifies an SSW as having occurred when the zonal-mean zonal wind at 60°N reverses on the 10 hPa pressure level. Once an SSW has been identified, the ratio of the two largest distinct regions of absolute vorticity on the 10 hPa pressure level is calculated, with the SSW being classified as a vortex-splitting SSW if the

ratio exceeds a given threshold value. If not classified as a vortex-splitting SSW, the SSW is instead classified as a vortex-displacement event. In this thesis, all references to SSWs and their classifications will refer to SSWs defined using the algorithm of Charlton and Polvani (2007).

The SSWs identified using the algorithm of Charlton and Polvani (2007) occur in the Northern Hemisphere with a frequency of approximately 0.62 per year, equivalent to two SSW events every three years (note that this is slightly higher than the frequency of 0.5 per year given in Andrews et al. 1985). To date, only one major warming has been recorded in the Southern Hemisphere, occurring in September 2002 (see *J. Atmos. Sci.* Special Issue, **62**, no. 3; see also Fig. 1.6). One other strong warming event occurring in the Southern Hemisphere during July 1974 was documented by Barnett (1975). Although this warming was similar in intensity to observed SSWs in the Northern Hemisphere, a reversal of the zonal-mean zonal wind was not observed meaning this event was still in fact a minor, rather than major, warming (see Quiroz 1974). Although not discussed here, minor warmings are observed more frequently than major warmings, and are known to occur in both hemispheres.

One of the main findings of Charlton and Polvani (2007) was that vortex-displacement and vortex-splitting SSWs are dynamically distinct, and should therefore be treated separately when studying stratospheric behaviour during SSWs. As part of their study, Charlton and Polvani (2007) found that vortex-splitting events in the Northern Hemisphere most often occurred in January and February, while vortex-displacement events occurred with almost constant frequency throughout the winter period of November-March.

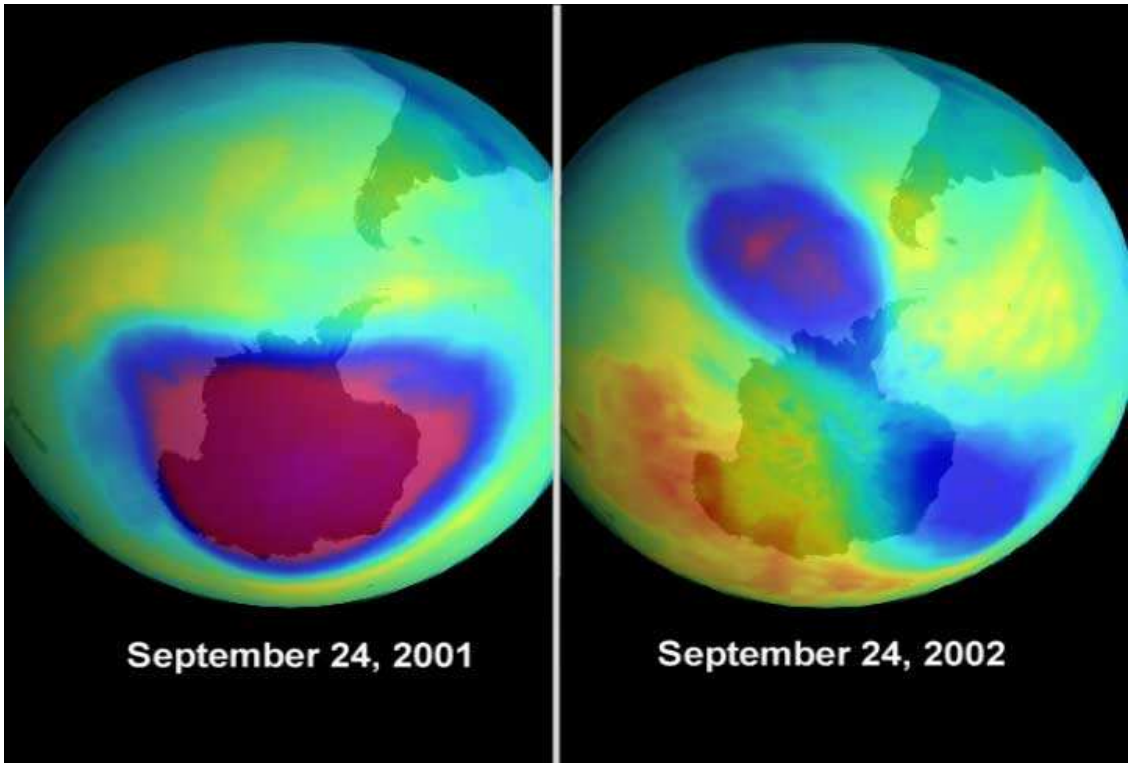

Figure 1.6: Comparison of stratospheric ozone concentrations on 24 September in the Southern Hemisphere between a "typical" winter observed in 2001 and the 2002 winter, in which the vortex-splitting SSW was observed. Data are provided by the Total Ozone Mapping Spectrometer (TOMS) instrument, and the figure has been provided by the NASA Ozone Hole Watch web resource.

The reason for treating vortex-splitting and vortex-displacement SSW separately can be seen by looking at the observed behaviour of the polar vortex during each type of SSW.

A particularly intense vortex-splitting event which has been the focus of many studies is that which occurred in February 1979 in the Northern Hemisphere (Andrews et al. 1985; Jung et al. 2001; Manney et al. 1994, see also Fig. 1.5). During this SSW, an almost columnar polar vortex was seen to split into two smaller daughter vortex columns of comparable size. Furthermore, the split was seen to occur

almost simultaneously at all heights (Manney et al. 1994, see their Fig. 14). Another vortex-splitting SSW which has attracted much attention is that observed during the Southern Hemisphere SSW in September 2002. In this instance, an initially columnar polar vortex was seen to split into two in the middle stratosphere while remaining columnar in the lower stratosphere, such that the overall vertical structure of the polar vortex was reminiscent of the letter "Y" (see *J. Atmos. Sci.* Special Issue, **62**). This Southern Hemisphere split was also observed in the distribution of stratospheric ozone, highlighting the role of the vortex edge as a barrier between air within the polar vortex and that outside (see Fig. 1.6). An example of the polar vortex behaviour observed during vortex-displacement events is provided by the SSW occurring in December 1998, discussed in Manney et al. (1999) and Naujokat et al. (2002). In Manney et al. (1999) (their Fig. 12), three-dimensional representations of the polar vortex show that the evolution is quite different to that of the February 1979 vortex-splitting SSW, with the polar vortex exhibiting a strong tilt with height while developing a cross sectional "comma" shape, with the "comma" becoming more elongated the higher up the vortex one looks (see Fig. 1.5 for an example of the January 1987 vortex-displacement SSW).

1.3 Dynamical theories of stratospheric sudden warmings

The underlying mechanisms responsible for the onset and development of SSWs are not entirely understood, in that there is no deterministic theory that fully explains

why SSWs occur in some years and not in others. However, there is a general consensus that the large scale disturbances observed during SSWs, along with the short timescales on which these disturbances develop, indicate that the phenomenon is of a dynamical rather than radiative nature. That is, rather than chemical or radiative heating effects, it is the isentropic transport of warm air masses into the polar regions during the polar vortex breakdown that is responsible for the temperature increases observed during SSWs.

There is a long history of investigation into the dynamical mechanisms which can lead to the breakdown of the polar vortex observed during SSWs. Early work (see Charney and Stern 1962; Fleagle 1958; Lindzen 1966; McIntyre 1972; Murray 1960) aimed to clarify the role of instabilities of the polar vortex as a possible mechanism for the breakdown of the polar vortex during SSWs (see also McIntyre 1982). However, it was found that even if the polar vortex were unstable to Rossby wave-like disturbances, the scale and behaviour resulting from such instabilities would not resemble that observed during the polar vortex breakdown during SSWs (McIntyre 1972). As a result, other ideas were investigated which looked to external factors such as tropospheric forcing or interactions between the polar vortex and other vortex structures (McIntyre 1982). Three of the most prominent theoretical ideas, upward propagating Rossby wave theory, resonant excitation theory and vortex-vortex interaction theory, will be introduced below.

Before looking at each of these theoretical ideas, it is useful to introduce two important concepts, each of which plays an important role in the study of SSWs. The first concerns conditions under which Rossby waves generated in the troposphere are

permitted to propagate vertically into the stratosphere. Rossby waves (also known as Planetary waves) are large scale wave-like disturbances which propagate on latitudinal gradients of PV in the atmosphere, similar to the way in which gravity driven water waves propagate on the density gradient between water and air (Andrews et al. 1985). The steep gradients in PV which are observed at the polar vortex edge therefore act as a wave-guide for the propagation of latitudinally trapped Rossby waves.

The ability of Rossby waves generated in the lower stratosphere to propagate to higher altitudes was investigated by Charney and Drazin (1961). In their seminal paper, Charney and Drazin (1961) found that for Rossby waves to propagate vertically into the stratosphere, the relationship

$$0 \quad < \bar{u} - c \quad < \bar{u}_{crit}, \tag{1.1}$$

must be satisfied, where $\bar{u}$ is the zonal-mean zonal wind of the stratosphere, c is the phase-speed of the Rossby wave in the zonal direction, and $\bar{u}_{crit}$ is a critical value dependent on parameters including the zonal wavenumber of the Rossby wave. For disturbances which are quasi-stationary with respect to the Earth's surface, that is Rossby waves with c close to zero, this condition dictates that vertical propagation of Rossby waves will only take place if the zonal-mean zonal winds in the stratosphere are moderate and westerly. That is, should upward propagating Rossby waves reach a height at which the zonal-mean flow is easterly, or strongly westerly, the waves will be unable to propagate any further in the vertical and are reflected or absorbed. In other words, if easterly zonal-mean zonal winds occur at some height in the stratosphere, any Rossby waves generated in the lower stratosphere will be trapped between the lower stratosphere and the height at which easterlies occur. The easterly zonal flow

which characterizes the stratosphere in the summer hemisphere (see Fig. 1.3), in addition to the absence of a persistent wave-guide such as the polar vortex edge, ensures that vertical propagation of Rossby waves into the stratosphere in summer is much less than that in winter.

The parameter $\bar{u}_{crit}$ exhibits an inverse dependence on the zonal wavenumber of the Rossby wave, such that $\bar{u}_{crit}$ is close to zero for disturbances with large zonal wavenumber. As a result, in the winter stratosphere vertical propagation is only generally observed for Rossby waves with zonal wavenumber 1, 2 or 3, with vertical propagation of higher wavenumber disturbances being suppressed.

The second concept is that of the Eliassen-Palm (EP) flux $\mathbf{S}$, and its associated wave-activity A, which were first introduced by Eliassen and Palm (1961) for internal gravity waves, and generalized in Andrews and McIntyre (1976, 1978) for Rossby waves. The EP-flux $\mathbf{S}$ is defined such that a conservation relation exists for the wave-activity A in the form of the "generalized Eliassen-Palm relation" (Andrews and McIntyre 1976, 1978)

$$\frac{\partial A}{\partial t} + \nabla \cdot \mathbf{S} \;=\; D,$$

where D is due to frictional and diabatic effects, such that in frictionless adiabatic flow, increases in wave-activity A arise directly from a divergence of the EP-flux $\mathbf{S}$. Use of the EP-flux as a diagnostic for geophysical flows was first introduced by Edmon et al. (1980), who noted that studying the EP-flux $\mathbf{S}$ in a meridional height cross section gave insight into the behaviour of eddy disturbances to a zonal-mean flow. In particular, by representing the EP-flux vector $\mathbf{S}$ with arrows, the direction of the arrows represents the direction of wave propagation, the vertical

component the northward heat flux due to the eddies, and the horizontal (meridional) component the northward momentum flux due to the eddies.The divergence of $\mathbf{S}$ has a further interpretation as it equals the magnitude of the northward potential vorticity flux due to the eddies. Importantly, under conditions of a monochromatic plane wave propagating in a slower varying background flow, the EP-flux also satisfies a group velocity property $\mathbf{S} = \mathbf{c}_g A$, where $\mathbf{c}_g$ is the Rossby wave group velocity. Consequently, the EP-flux is often interpreted as being indicative of a Rossby wave packet propagating with its group velocity.

We now give an overview of the three theoretical ideas regarding the dynamical mechanisms of SSWs, as mentioned earlier in this section. Although presented separately, it is pointed out that these three ideas are not necessarily mutually exclusive, and that the true dynamical mechanisms responsible for the onset of SSW events may well include properties of some, if not all, of these different ideas.

1.3.1 Upward propagating Rossby waves

A well established view of the dynamical mechanism responsible for SSWs is that they are caused by upward propagating Rossby waves generated in the troposphere. An outline of this follows. One of the important conceptual features of this theory is that the troposphere is thought of as acting independently of the stratosphere over the period in which warmings occur. Unusually high amplitude disturbances in the troposphere are manifest in the lower stratosphere as an anomalous forcing which excites Rossby wave disturbances (Nishii and Nakamura 2004; Peters et al. 2007). Due to the moderate westerly zonal-mean zonal winds which characterize the stratosphere

during winter, these Rossby waves are able to propagate to higher altitudes in accordance with the theory of Charney and Drazin (1961). The vertical propagation can be diagnosed in the meridional cross sections of the EP-flux as an upward vector **S**. Due to the exponential decrease in density with height, the amplitude of these upward propagating Rossby waves increases until wave saturation occurs in the upper stratosphere, and the Rossby waves break. This wave breaking initiates strong mixing in the upper stratosphere, which reverses the meridional temperature gradient which in turn reverses the zonal-mean zonal flow to easterlies. Following Charney and Drazin (1961), this easterly mean flow acts as a barrier to further upward propagation of Rossby waves, such that subsequent Rossby waves must deposit their energy below this height, initiating mixing and reversal of the meridional temperature gradient at lower altitudes. This process continues, with the height at which easterlies are first encountered descending through the stratosphere as the SSW develops.

This mechanism was first proposed in Matsuno (1971), who used a novel dynamical model of the stratosphere, perturbed by lower boundary forcing simulating forcing at the tropopause, to produce the first semi-realistic model SSWs. As the model flow evolved, SSW-like behaviour was observed in the model for lower boundary forcing with both wavenumber-1 and wavenumber-2 zonal structure. In particular, it was seen that reversal of the zonal-mean wind to easterly flow descended steadily through the stratosphere as SSWs developed. A more realistic form of this model was developed by Holton (1976), whose simulations agreed with those of Matsuno (1971) with the exception that the reversal of the zonal-mean zonal wind to easterly flow was seen to occur almost simultaneously throughout the stratosphere, rather than the steady

descent observed by Matsuno. This model was also used by Butchart et al. (1982) to simulate the SSW which was observed in February 1979.

One important question which must be asked with respect to the upward propagating Rossby wave theory outlined above is the interpretation of the upward pointing EP-flux in the vortex edge region. In idealized dynamical model studies, Scott et al. (2004) and Scott and Polvani (2004) showed that when considering stratospheric disturbances due to lower boundary forcing, the upward EP-flux at the lower boundary is not only dependent on the forcing amplitude, but on the state of the stratosphere itself. That is, large EP-flux at the lower boundary is partly due to the excitation of Rossby waves by the lower boundary forcing, and partly due to the state of the stratosphere "inviting" the excited Rossby waves to propagate up into the stratosphere (see also Gray et al. 2003). The strong variability of upward EP-fluxes seen in modelling studies remains a difficulty for the upward propagating Rossby wave theory, in which the EP-flux at the lower boundary is conceptually attributed to the constant forcing at the lower boundary, at least in idealized models. One apparent shortcoming of the theory of upward propagating Rossby waves is that it fails to explain why SSW-like behaviour is not always observed for large amplitude forcing in the lower stratosphere.

1.3.2 The resonant excitation theory

An alternative theory for the large and sudden wave growth of Rossby waves on the polar vortex edge during SSWs is that a resonant excitation of a free Rossby wave mode internal to the stratosphere occurs, with the excitation due to remote forcing (e.g. forcing from the troposphere).

Resonance of atmospheric flows subject to topographic forcing was studied by Charney and DeVore (1979) as a possible mechanism for blocking phenomena in the troposphere. Charney and DeVore (also see Tung and Lindzen 1979a, b) found that as disturbances in the flow became stationary with respect to the forcing, resonance between the two resulted in a large and sudden increase in the amplitude of the disturbance. This concept of resonant wave growth had previously been applied to the study of SSWs by Clark (1974), who postulated that resonant excitation of stratospheric disturbances with forcing from the troposphere, in addition to an increase in amplitude of the forcing, was required for SSW-like behaviour to be observed. This would appear to account for the fact that an increase in forcing amplitude alone does not generate SSW-like behaviour in all instances. The linear theories presented in Clark (1974) and Tung and Lindzen (1979a, b) were further developed by Plumb (1981), who quantified the effects of nonlinearity when determining the evolution of the stratosphere during near resonant wave growth. It was found that for stratospheric flows close to resonance, nonlinear effects due to wave growth can act to "self-tune" disturbances closer to resonance. Plumb also pointed out that any growth of disturbances in an initially exactly resonant system would act to de-tune the system from resonance, causing disturbance amplitudes to decay over time. The process of self-tuning resonance in a more realistic model of the stratosphere was then studied by Smith (1989), who found that a self-tuning resonance of disturbances with topographic forcing produced zonal-mean diagnostics which were similar to those observed during real SSWs.

One limitation of the Tung and Lindzen and Plumb studies is the use of the

β-plane channel model. The PV distribution of this model, which does not closely correspond to that observed in the real polar vortex, in conjunction with its channel geometry, means that the β-plane model is unable to predict the realistic evolution of the polar vortex during vortex-splitting and vortex-displacement SSWs.

By using a more suitable three-dimensional f-plane model (see chapters 3 and 6), with an explicit representation of the polar vortex edge as a sharp jump in PV, Esler and Scott (2005) showed that a resonant excitation of Rossby waves on the vortex edge by topographic forcing leads to vortex-splitting behaviour similar to that observed during SSWs. The same model was then used by Esler et al. (2006) in which more realistic polar vortex configurations were used to construct a conceptual model of the September 2002 vortex-splitting SSW in the Southern Hemisphere. In both of these studies, numerical experiments indicated that the effects of nonlinearity in highly disturbed vortex states acted to either self-tune the vortex towards resonance, or to de-tune the vortex away from resonance, depending on conditions in the stratosphere during the experiment. These findings reinforced those of Tung and Lindzen (1979a) and Plumb (1981) in a somewhat more realistic model framework.

Quantifying the effects of nonlinearity on the resonant excitation of Rossby waves by topography, leading to vortex behaviour similar to that observed, during SSWs, will be of central importance to the work in this thesis.

1.3.3 The vortex-vortex interaction theory

As mentioned previously, the fact that the vortex is often displaced from the pole means that partitioning the stratospheric flow into a zonal-mean component and

small scale disturbance component may lead to an unrealistic picture of the vortex behaviour during SSWs. To avoid partitioning the flow in this way, another theoretical mechanism for the onset and development of SSWs instead focuses on interaction of the cyclonic polar vortex with large deep anticyclonic vortex structures which form during the winter in the stratosphere.

Large and deep anticyclonic vortex structures have been observed in the stratosphere by O'Neill et al. (1994) and Lahoz et al. (1996), and were seen to form from the vortex merger of smaller anticyclonic vortices. For the minor warming occurring in 1992, O'Neill et al. found that a large anticyclonic structure was formed from the merger of the "Aleutian high", a relatively stationary stratospheric anticyclonic vortex structure located over the Aleutian islands at approximately 10 hPa during the winter, and another transient anticyclonic vortex, with their merger contributing to the onset of the warming. The merger of travelling anticyclonic vortices in the Southern Hemisphere was observed by Lahoz et al., who noted that the large anticyclonic vortex resulting from the merger interacted with a weakened polar vortex to contribute to the onset of the final warming.

A simple dynamical model of vortex-vortex interaction was used by Scott and Dritschel (2006) in an attempt to understand the nature of such interactions on polar vortex deformation. It was found that for a relatively small anticyclonic vortex structure in the lower stratosphere, interaction with a large and deep cyclonic polar vortex leads to significant deformation of the polar vortex, with very little deformation of the anticyclonic vortex.

An important question regarding this mechanism is the formation of the anti-

cyclonic vortex playing a role in such vortex-vortex interactions. Using idealized model experiments as well as observational data, Gray et al. (2004) found that in the Northern Hemisphere, the development of the anticyclonic vortex associated with the Aleutian high was dependent on both the phase of the Quasi-Biennial Oscillation (QBO) and the solar cycle. It was found that when the QBO was in its east phase, corresponding to easterly zonal-mean zonal winds in the equatorial lower stratosphere, and the solar cycle was at its maximum, easterly anomalies were observed in the zonal-mean zonal winds in the tropical upper stratosphere. These easterly anomalies were seen to enhance the anticyclonic vortex associated with the Aleutian high, which in turn speeded up the onset of SSWs. In contrast, when the QBO was in its west phase, corresponding to westerly zonal-mean zonal flow in the equatorial lower stratosphere, and the solar cycle was at its minimum, westerly anomalies were observed in the tropical upper stratosphere, suppressing formation of the Aleutian high anticyclone and delaying, or in some cases prohibiting, the onset of SSWs. These results are consistent with the Holton-Tan mechanism (Holton and Tan 1980), which suggests that Rossby wave refraction makes SSWs more probable during the QBO east phase, although the Holton-Tan mechanism places more emphasis on wave refraction in the lower stratosphere while Gray et al. (2004) highlighted the possible role of the background winds in the upper stratosphere.

1.4 Summary of thesis

Stratospheric sudden warmings (SSWs) are studied from a vortex dynamics perspective with the aim of identifying the underlying mechanism that causes such events.

In chapter 2, observational data are used to investigate the behaviour of the Arctic polar vortex during vortex-splitting and vortex-displacement SSWs. Utilizing moments of the potential vorticity distribution in the stratosphere, an objective diagnostic is introduced and used to obtain climatologies of the polar vortex behaviour during vortex-splitting and vortex-displacement SSWs.

In the remainder of the thesis, a hierarchy of models approach is used to develop a resonant excitation theory as a possible mechanism leading to polar vortex breakdown similar to that observed during SSWs, as detailed in chapter 2. In chapter 3, a single layer quasi-geostrophic shallow-water model and a three dimensional quasi-geostrophic compressible atmosphere model are introduced and the numerical algorithms used to perform fully nonlinear experiments for each type of model are discussed.

Chapters 4 and 5 focus on developing nonlinear analytical theories to explain the behaviour of the numerical models, which are shown to reproduce vortex-splitting behaviour reminiscent of that observed during SSWs. A simple model of elliptical vortices in strain flow (the Kida vortex) is shown to be a good approximation to the quasi-geostrophic shallow water model in which vortex-splitting behaviour is observed.

A three-dimensional idealized columnar vortex model is used in chapter 6 to investigate the relevance of resonant excitation theory in predicting vortex breakdown similar to that observed during vortex-displacement SSWs. Finally, in chapter 7, the three-dimensional model is used to investigate resonant behaviour in more realistic vortex structures.

In the modelling aspects of this thesis, when investigating the underlying mech-

anisms leading to polar vortex breakdown during SSWs, the main focus will be on the role of a resonant excitation mechanism, rather than upward propagating Rossby waves or vortex-vortex interactions.

Chapter 2

Observations of Stratospheric Sudden Warmings: Polar Vortex Evolution and Vertical Structure

2.1 Introduction

This study is the third in a series investigating the observed dynamical behaviour of major mid-winter stratospheric sudden warmings (SSWs). Observational benchmarks derived from the studies in Charlton and Polvani (2007), Charlton et al. (2007) and this chapter will serve to characterize typical dynamical behaviour in the polar stratosphere during both vortex-displacement and vortex-splitting SSW events. Part I of the series (Charlton and Polvani 2007, CP07 hereafter) was primarily concerned with the identification and classification of all SSWs in the period 1957-2002, and on the development of diagnostic benchmarks. In particular, in CP07, the evolution

of temperature, geopotential height, zonal wind and eddy flux fields during SSW events of each type were documented using a composite method. In part II (Charlton et al. 2007), the diagnostic benchmarks developed in CP07 were used to evaluate the performance of several stratosphere-resolving models in simulating SSWs. Here, the aim is to extend the study further by analyzing SSWs from the perspective of three-dimensional vortex dynamics.

It is well-established (e.g. McIntyre and Palmer 1983; Nash et al. 1996) that the Arctic polar vortex can be identified with an airmass of elevated Ertel's potential vorticity (PV hereafter), and that the vortex itself is strongly disrupted during an SSW (e.g. O'Neill 2003, and references therein). The aim of this chapter is therefore to investigate the structure of the Arctic polar vortex during SSW events of each type, by a detailed analysis of the PV distribution in the Arctic polar region in the altitude range 15-45 km. Following Limpasuvan et al. (2004) and CP07, time-lag composite diagnostics are used to characterize the development of the vortex leading up to and following each type of SSW. The aim is to answer the following questions:

- What is the vertical structure of the Arctic polar vortex during the vortex-displacement and vortex-splitting SSWs identified in CP07? Does each type of SSW event have a characteristic vertical structure, or does the vertical structure vary between individual events.

- What is the location and orientation of the Arctic polar vortex with respect to the Earth's surface during the development of each type of SSW event? To what extent are SSW events 'phase-locked' to stationary tropospheric planetary waves generated by surface topography and land-sea contrast?

- How does the shape of the Arctic polar vortex develop prior to, during, and subsequent to the onset time of SSWs? Is there a characteristic life cycle associated with the development of the polar vortex during each type of SSW?

It is anticipated that our results, and the answers to the above questions in particular, will be of assistance to modellers attempting to simulate SSWs in stratosphere-resolving general circulation models. Additionally, the results should be of use to researchers seeking to validate different dynamical theories of SSWs, as will be discussed in detail below.

A number of authors have analyzed the evolving vertical structure of the vortex during specific SSW events in both the northern and southern hemisphere (e.g. Manney et al. 1994, 2005b, c). However, we are not aware of a study attempting a comprehensive analysis of the three-dimensional structure of the polar vortex during all recorded SSWs. One relevant, complementary investigation is that of Limpasuvan et al. (2004), who compiled composite diagnostics such as zonal mean winds and temperatures, and Eliassen-Palm flux fields during the life cycles of 48 SSW and minor warming events. Limpasuvan et al. (2004) did not, however, distinguish between vortex-displacement and splitting events, which from the perspective of vortex dynamics are fundamentally different phenomena. Nor did they consider the evolution of the vertical structure of the polar vortex itself during the SSW events. As will be seen below, the shape of the polar vortex undergoes robust and distinctive life cycles during each type of SSW, details of which emerge when the above aspects are considered. Another relevant study is Liberato et al. (2007), who used NCEP/NCAR data to investigate the vertical structure of planetary wave energetics in relation to

the variability of the northern hemisphere polar vortex during SSW events. In their work, composite analysis of displacement and splitting SSWs showed the dynamics of the two SSW types to be distinct. Displacement SSWs were forced by planetary wave energy anomalies with strong vertical dependence, whereas splitting SSWs were forced by planetary wave energy anomalies with predominantly barotropic structure in the vertical.

A detailed study of the climatological behaviour of the Arctic and Antarctic polar vortices has been carried out by Waugh (1997), who introduced the methodology of 'elliptical diagnostics' in order to produce time series describing the main aspects of the evolution of the Arctic and Antarctic polar vortex structure during their respective winter seasons. These time series were subjected to statistical analysis, resulting in reference climatologies for the position and the variability of the polar vortex centroid, aspect ratio and orientation. Three key results of Waugh and Dritschel (1999, see in particular their Fig. 4) that are specifically relevant to the work of this chapter, relate to the climatology of the Arctic polar vortex during the winter period (late November-March), when all of the SSWs identified in CP07 occur. Firstly, the centroid of the vortex is found to be displaced off the pole; at all levels throughout the winter it is located near latitudes 80°N (December) to 75°N (March). Secondly, the vortex is found to have a significant westward tilt with height; the longitude of the vortex centroid is located around 20-30°E at low levels (450 K), near the 0° meridian in the mid-stratosphere (850 K), and around 30-40°W in the upper stratosphere (1300 K). Finally, the aspect ratio of the vortex is generally in the range 1.5-1.9, with the largest values occurring in February in the lower stratosphere. Results will be presented below

with reference to this climatology.

Nevertheless, the focus in the Waugh (1997) and Waugh and Dritschel (1999) studies was not the three-dimensional evolution of the polar vortex during SSWs themselves. Here, the elliptic diagnostic approach, which was adapted from techniques developed in the fluid dynamics literature to investigate two-dimensional vortex dynamics (Dritschel 1993; Legras and Dritschel 1993; Melander et al. 1986), is further adapted to study the structure of the Arctic polar vortex during SSW events of each type. One important novelty in this work is that we have removed the need to start by choosing the vortex edge, such as a specific isopleth of N_2O mixing ratio or of PV, before applying the elliptical diagnostics methodology. This modification allows independent results from the elliptical diagnostics method to be compared with those from a separate vortex edge-selecting algorithm. Results from the two techniques will therefore act to cross-validate each other and help confirm the robustness of our results.

The remainder of the chapter is structured as follows. Section 2.2 describes the datasets used to identify SSWs and the methodology employed to objectively define the boundary of the Arctic polar vortex, and to generate the composite fields used to characterize SSWs of each type. The elliptical diagnostics are then introduced, the new aspects described, and their utility demonstrated. In section 2.3 the vertical structure, orientation and elongation of the Arctic polar vortex are described for both vortex-splitting SSWs and vortex-displacement SSWs. Particular attention is given to the temporal evolution of these quantities relative to the SSW onset time. Both composites and representative individual events are compared and contrasted. In

section 2.4 a list of proposed modelling benchmarks are presented and section 2.5 contains our conclusions.

2.2 Datasets and methodology

2.2.1 Dataset, SSW definition and categorization

The methodology used to identify SSWs, and then classify them into vortex-displacement and vortex-splitting events, is described in detail in CP07. CP07 use data from both the NCEP/NCAR re-analysis (Kistler et al. 2001), and the ECMWF ERA-40 re-analysis dataset (Uppala et al. 2005), to establish a climatology of SSW events between the winter seasons 1957/58 and 2001/2002. The results from the two re-analysis datasets are very similar, as expected given that they are created using a largely common observational dataset. In this chapter, the vertical structure of SSWs is investigated using Ertel's potential vorticity (PV) from the ERA-40 re-analysis only, because PV is directly available at a higher vertical and temporal resolution in ERA-40 than in the NCEP/NCAR re-analysis.

Following CP07, our study is confined to major mid-winter SSWs which occur during the extended boreal winter season, November to March. SSWs are defined to occur when the zonal mean zonal wind at 10 hPa and 60°N changes direction from westerly to easterly. An additional criterion, that the zonal mean zonal winds at this location recover to westerlies for at least 10 or more consecutive days following the SSW, is used to remove events which are final, as opposed to midwinter, warmings.[1]

[1]The CP07 definition differs slightly from those in the existing literature in that it includes warming events which occur in March, and can be robustly ascertained not to be final warmings. In

The dynamics of Northern hemisphere final warmings have recently been discussed in detail by Black and McDaniel (2007). During the 1957-2002 period of investigation, 29 SSWs that satisfy these criteria are identified by CP07. These 29 events are then classified into vortex-displacement and vortex-splitting events, using an algorithm based on first identifying the vortex edge from the absolute vorticity field, and then comparing the size and amplitude of cyclonic vortices. CP07 divide the 29 SSWs into 15 vortex-displacement events and 14 vortex-splitting events using this algorithm.

For the purposes of the current study ECMWF ERA-40 Ertel's potential vorticity and temperature fields were obtained from the British Atmospheric Data Centre (BADC, http://badc.nerc.ac.uk/home/). Global fields were obtained at $1^\circ \times 1^\circ$ horizontal resolution on 16 standard pressure levels between 400 hPa and 1 hPa with a vertical distance of approximately 2.6 km between each level. The temperature fields were used to interpolate the PV onto 36 isentropic levels in the range 380-2850 K.

Note that Manney et al. (2005b), Manney et al. (2005a) and Simmons et al. (2005) have reported an erroneous vertical oscillation in temperature in the polar regions in ERA-40. We have found that this oscillation, which is of lower amplitude in the Arctic compared to the Antarctic and is only significant in the last few years of the period of interest, is not of sufficient amplitude to impact on our results. A direct comparison between ERA-40 and ERA operational analysis data for the February 2002 event (not shown) reveals no significant differences in any of our diagnostics.

addition, some researchers may be uncomfortable with use of the term 'midwinter warming' when discussing SSW events occurring in November, preferring to use 'Canadian warming' for these events. For consistency with CP07, we have chosen to adhere to the dynamically based SSW definition given there, and have therefore included any November SSWs in CP07 in our analysis.

2.2.2 Definition of the arctic polar vortex boundary

In order to investigate the Arctic polar vortex during each SSW period, an objective definition of the vortex airmass is required. A natural way of defining the vortex airmass is to select a surface in three-dimensional space to define its outer boundary. The edges of the polar vortices throughout the stratosphere and lower mesosphere are identified with strong isentropic gradients in trace gas concentrations and in PV; hence any surface defining the outer boundary of the vortex should coincide with these strong gradients where they exist. For the purposes of this study, the vortex is assumed to extend between the 400 K and 1600 K isentropic surfaces (approximately 14 and 44 km) in the vertical. For each SSW identified in CP07, the vortex outer boundary within this range is selected to coincide with the region of strong PV gradient using the following method.

Firstly, the Ertel PV q, on each isentropic surface θ, is expressed as a function of equivalent latitude ϕ_e (Butchart and Remsberg 1986) during the 20 day period surrounding each SSW. In order to obtain a quantitative snapshot of the vortex before the development of each SSW, $q(\phi_e, \theta)$ is considered 9 days before the SSW onset date identified in Table 1 of CP07. A time lag of 9 days was found to be sufficient for the vortex to be in a relatively undisturbed state prior to the SSW. To proceed, the equivalent latitude $\phi_m(\theta)$ corresponding to the maximum latitudinal gradient in PV, i.e. $\max\{|\partial q/\partial \phi_e|\}$, is identified on each isentropic level θ. At nearly all times and levels, the 'vortex edge', defined by $\phi_m(\theta)$, lies in the equivalent latitude range $45^\circ - 80^\circ$ N.

In theory, it might be desirable to proceed using the function $\phi_m(\theta)$ to identify

the vortex edge during each SSW. In practice, however, there are numerical problems associated with discontinuities and noise in $\phi_m(\theta)$, due for example to (occasional) spurious maxima in the latitudinal gradient of PV. Additionally, such an approach requires a significant amount of information to be stored in conjunction with each SSW. Here, an alternative method, exploiting a transformation of the PV, is employed. Lait (1994) pointed out that the exponentially increasing component of PV q can be removed without affecting its conservation properties, by a transformation of the form

$$Q = q \left(\frac{\theta}{\theta_0} \right)^{\alpha}, \tag{2.1}$$

where θ_0 and α are constants. Müller and Günther (2003) point out that variation in the vertical temperature structure of the polar stratosphere is sufficiently large to warrant different values of α being used for different events. Hence, we aim to choose α so that a single isosurface of the vertically weighted PV, $Q = Q^*$, gives a good approximation to the vortex edge as defined by the equivalent latitude $\phi = \phi_m(\theta)$ on all isentropic surfaces. Without loss of generality, we can take $\theta_0 = 475$ K for all events. For each SSW, a particular choice of α and Q^* that leads to a good fit to the vortex edge is obtained by setting

$$Q^* = q(\phi_m(\theta_1), \theta_1) \left(\frac{\theta_1}{\theta_0} \right)^{\alpha} = q(\phi_m(\theta_2), \theta_2) \left(\frac{\theta_2}{\theta_0} \right)^{\alpha}, \tag{2.2}$$

i.e. Q^* is set equal to the vertically weighted PV at the vortex edge (defined by ϕ_m) at two isentropic levels θ_1 and θ_2. The system (2.2) constitutes two equations in two unknowns, Q^* and α, and can be straightforwardly solved numerically to obtain a unique solution. We take $\theta_1 = 600$ K and $\theta_2 = 1400$ K, representative of the vortex in the lower and upper stratosphere respectively. For each SSW identified in

CP07, except for the March 1971 event, the vortex edge is found to be well-defined at both of these levels, and the surface $Q = Q^*$ lies close to the equivalent latitude ϕ_m throughout the potential temperature range of interest.[2]

The three-dimensional isosurface $Q = Q^*$, bounded below and above by the 400 K and 1600 K isentropic surfaces respectively, serves to define the outer boundary of the Arctic polar vortex throughout the period of each SSW. In addition to being a good fit to the location of the maximum latitudinal gradient of PV, the $Q = Q^*$ isosurface is straightforward to calculate and plot, requiring only the values of α and Q^* for each SSW. Following the above procedure for each event, the vortex edge was found to be within the range $Q^* = 38 \pm 7$ PVU (1 PVU $\equiv 10^{-6}$ K m^2 s^{-1} kg^{-1}), and the vertical scaling parameter $\alpha = -4.25 \pm 0.35$. The vertically weighted PV Q thus generated can also be used to generate composite SSWs, as will be described next.

It is worth noting that this method of deriving an isosurface of PV $Q = Q^*$ representing the vortex edge at any time is not restricted to SSW events. By generating a value of Q^* corresponding to the vortex edge, whether on a daily basis or by using a single value over a given time period, the three-dimensional evolution of the vortex can be easily represented at any time. However, the fact that PV is not conserved over long time periods in the stratosphere implies that at large times before or after the time at which Q^* is calculated, the isosurface $Q = Q^*$ may no longer correspond closely to the vortex edge.

[2]For the March 1971 event the vortex was found to have an unusual structure in the upper stratosphere, and the above method could not be applied. This is most likely due to the March 1971 event occurring late in the winter, and to the polar vortex being relatively weak at upper levels. Although the algorithm of CP07 identifies the event as a midwinter SSW, it in fact shares many dynamical features with final warmings. The March 1971 event is therefore excluded from our analysis below, leaving 15 displacement events and 13 splitting events.

2.2.3 Construction of composite fields

The evolution of the Arctic polar vortex during 'typical' vortex-displacement and vortex-splitting events can be investigated using time-lag composites of the vertically weighted PV Q. The advantage of using Q to create the composites, rather than for example PV q, is that the component of the PV field that increases exponentially with height has been removed from Q by the transformation (2.1) on a case-by-case basis. This ensures that the effects of the seasonal cycle and interannual variability on that part of the untransformed PV that grows exponentially with height are removed, and thus avoids biasing the resulting composites towards a few individual events.

In order to generate the time-lag composites an onset time ($\tau = 0$) must be selected for each event. For the vortex-displacement events, $\tau = 0$ is taken to be the event onset date given in CP07. For vortex-splitting events, the dominant feature of the SSW is the time at which the vortex splits from a single vortex into two smaller daughter vortices. This time is not necessarily the onset time as given in CP07. More physically meaningful composites are generated if $\tau = 0$ is taken as the first recorded time at which the vortex, defined as the region $Q > Q^*$, splits into two pieces of comparable area on the 850 K isentropic surface, rather than using the onset date given in CP07. In most cases this time differs by less than 2 days compared to the onset time identified by CP07, with 6 vortex split SSWs having an observed onset before that given in CP07 and 7 vortex split SSWs having an observed onset after that given in CP07.

Time-lag composites of Q can be used to examine the typical evolution of the vortex, relative to the SSW onset time, for events of each type. The composite PV

field $\bar{Q}(\tau)$ is constructed by averaging the vertically weighted PV fields at time lags τ, relative to the onset time. Here, we focus on the time-lag range $\tau = -10$ to $\tau = 10$ days. A simple average of all relevant events is taken

$$\bar{Q} = \frac{\sum_{i=1}^{N} Q_i}{N}, \tag{2.3}$$

where N is the number of individual events of each type. For vortex-displacement events $N = 15$, corresponding to the events identified by CP07, and for vortex-splitting events $N = 13$, corresponding to the 14 events identified by CP07 minus the March 1971 event. For each event, the composite vortex edge is defined as $\bar{Q} = \bar{Q}^*$, where $\bar{Q}^*$ is the average of Q^* over all the events.

2.2.4 Vortex moment diagnostics and the equivalent ellipse

Next, an alternative methodology is described that can be used to generate objective measures of the vortex centroid, orientation and ellipticity. The method consists of a novel adaptation of the elliptical diagnostics methodology of Waugh (1997). The basic idea is to exploit the approach of vortex moments (e.g. Melander et al. 1986), in order to define an 'equivalent ellipse'; an elliptical vortex of uniform PV that has identical moment diagnostics to the polar vortex on a given isentropic level. The main modification to the technique as applied by Waugh (1997) is that we avoid altogether the definition of a vortex edge prior to calculating the vortex moments. A significant advantage of this modification to the technique is that the new diagnostics can serve to verify that our results are independent of the choice of vortex edge described above. Note, however, that Waugh (1997) in fact found that the results of his climatological study were not very sensitive to the choice of vortex edge.

Specifically, the equivalent ellipse at a given time t and on the isentropic surface with potential temperature θ, is calculated as follows. The Ertel's PV $q(\lambda, \phi, \theta, t)$ is considered initially as a function of longitude λ and latitude ϕ. As we are interested in diagnosing the location, strength and orientation of the polar vortex only, q is first modified to remove any anomalously low values of PV near the pole. Air with low PV is likely to have originated in the tropics and will have its own distinct dynamical significance, which could be investigated separately if required. A modified PV field $\hat{q}$ is defined by

$$\hat{q} = \begin{cases} q & \text{where } q \geq q_b, \\ q_b & \text{where } q < q_b. \end{cases} \tag{2.4}$$

Here $q_b(\theta, t)$ is a representative background value of PV calculated by averaging q polewards of 45°N. The moment diagnostics are then applied to the modified field $\hat{q}$, after first transforming to Cartesian coordinates (x, y), to facilitate the direct exploitation of the results of Melander et al. (1986). Although it is possible to generalize the moments approach to spherical polar coordinates (Dritschel 1993), Waugh (1997) found that the transformation to Cartesian coordinates greatly simplified the moment calculations, and has minimal impact upon the results obtained. Hence, Lambert's azimuthal equal-area projection is employed giving $x = R\cos\lambda$, $y = R\sin\lambda$ and $R = a\sqrt{2(1-\sin\phi)}$, with a the Earth's radius. Note that this choice of R ensures that the mappings from spherical to Cartesian coordinates and back again are area-preserving.

It is useful to introduce the absolute vortex moment $M_{kl}(\theta, t)$, given by

$$M_{kl} = \int\int (\hat{q}(x, y) - q_b) x^k y^l \, dx \, dy, \tag{2.5}$$

where the integral is taken out to the 'equator' at $R = \sqrt{2}a$. Note that because $\hat{q} = q_b$ nearly everywhere where $R > a/\sqrt{2}$, the region between $R = a/\sqrt{2}$ and $R = \sqrt{2}a$ seldom contributes to the moment integrals. The equivalent ellipse is uniquely defined by its centroid $(\bar{x}, \bar{y})$, its aspect ratio r, its angle of orientation ψ and its area A. These are determined as follows. Firstly the centroid is obtained

$$(\bar{x}, \bar{y}) = \frac{1}{M_{00}} (M_{10}, M_{01}) , \tag{2.6}$$

allowing the relative vortex moments

$$J_{kl} = \int\int (\hat{q}(x, y) - q_b)(x - \bar{x})^k (y - \bar{y})^l \, dx \, dy, \tag{2.7}$$

to be defined. Then ψ, defined as the angle between the x-axis and the major axis of the ellipse, and the aspect ratio r are obtained as

$$\psi = \frac{1}{2} \arctan \left(\frac{2J_{11}}{J_{20} - J_{02}} \right) , \tag{2.8}$$

$$r = \left| \frac{(J_{20} + J_{02}) + \sqrt{4J_{11}^2 + (J_{20} - J_{02})^2}}{(J_{20} + J_{02}) - \sqrt{4J_{11}^2 + (J_{20} - J_{02})^2}} \right|^{1/2} . \tag{2.9}$$

Importantly, the value of the aspect ratio r is invariant under both translations and rotations of the modified PV field $\hat{q}$.

Finally, the equivalent area of the ellipse is defined to be

$$A = \frac{M_{00}}{q_b}. \tag{2.10}$$

The definition (2.10) implies that the equivalent area A is an objective measure of not only the actual area but also the intensity of the polar vortex. The area can therefore be regarded as an objective measure of the total strength, and hence dynamical significance, of the vortex on a given isentropic level. This is a substantial difference

from Waugh (1997), where the area represents the area within a closed PV contour selected by a separate algorithm. Note that here the PV anomaly associated with the vortex is normalized by the background value of PV q_b, so that if the PV in the polar vortex is significantly less than $2q_b$, then the equivalent ellipse will appear smaller than the original vortex. Conversely, the equivalent ellipse will be larger if PV in the polar vortex significantly exceeds $2q_b$. Once calculated, the values of $\bar{x}$, $\bar{y}$, r, ψ and A allow the coordinates of the boundary of the unique equivalent ellipse to be determined parametrically. The Cartesian coordinates of this boundary are then mapped back to longitude-latitude coordinates.

The result of the above procedure applied to the February 1979 vortex on the 850 K isentropic surface, 2 days before the warming onset time at 0600UTC on February 21 1979, is shown in Fig. 2.1 a, plotted together with the actual PV distribution at that time. It is clear that the equivalent ellipse corresponds closely to the actual polar vortex (note that the shaded region corresponds to the vortex as defined by the edge-finding method described above). The equivalent ellipse method therefore serves to reduce the essential elements of a two-dimensional field to a few key parameters.

The above method is found to work well whenever there is a single, relatively well-defined polar vortex. In order to investigate vortex-splitting SSWs, however, we also need to investigate situations where the vortex splits into two. The above procedure can be adapted when a vortex split has taken place, i.e. when $\tau > 0$ during the vortex split SSWs, as follows. Firstly, the vortex centroid $(\bar{x}, \bar{y})$ and the orientation of the modified PV field $\hat{q}$ are determined as above. Next the Cartesian domain is divided into two regions $\mathcal{R}_1$ and $\mathcal{R}_2$ by the straight line $(y - \bar{y}) = -\cot\psi(x - \bar{x})$, i.e.

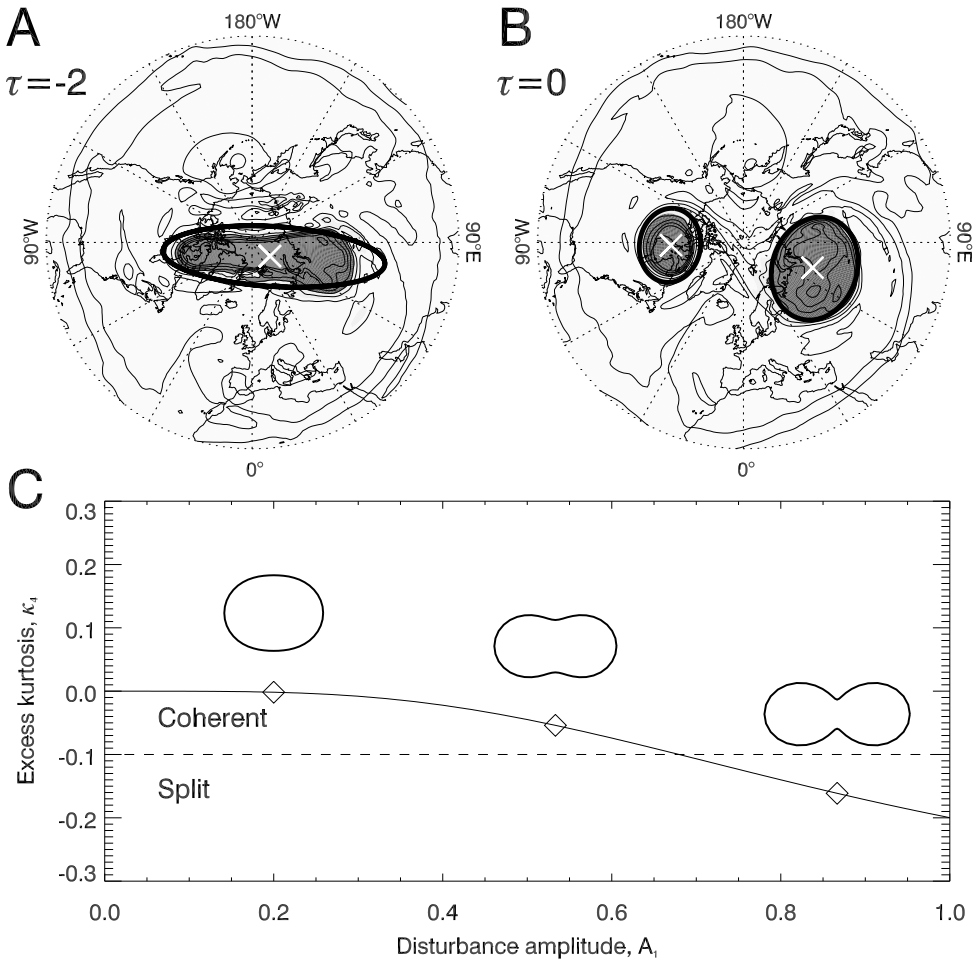

Figure 2.1: Top panels: Ertel's potential vorticity q on the 850 K isentropic surface on February 19 1979 at 0600UTC (A: $\tau = -2$ days) and February 21 1979 at 0600UTC (B: $\tau = 0$ days). Shaded regions correspond to $q > q^*$ where $q^* = 370$ PVU (1 PVU $\equiv 10^{-6}$ K m^2 s^{-1} kg^{-1}) is the vortex edge value (for the vertically weighted PV $Q^* = q^*(\theta/\theta_0)^\alpha = 27.5$ PVU, where $\alpha = -4.467$ for the February 1979 event). The black ellipses show the 'equivalent ellipses' calculated as described in section 2.2. The location of the ellipse centroids are marked by white crosses. Note that in (B) a vortex split has occurred and therefore two ellipses are found following the algorithm described in the text. Panel (C) illustrates the dependence of the excess kurtosis parameter κ_4 on the parameter A_1 (in equation 2.13) that controls the extent to which an idealized vortex of uniform vorticity differs from an ellipse.

the straight line through the centroid that is perpendicular to the major axis of the $\hat{q}$ distribution. Two new modified PV distributions are then defined as

$$\hat{q}^{(1)} = \begin{cases} \hat{q} & \text{in } \mathcal{R}_1, \\ q_b & \text{in } \mathcal{R}_2, \end{cases} \qquad \hat{q}^{(2)} = \begin{cases} \hat{q} & \text{in } \mathcal{R}_2, \\ q_b & \text{in } \mathcal{R}_1. \end{cases} \tag{2.11}$$

The method described above is then applied to $\hat{q}^{(1)}$ and $\hat{q}^{(2)}$ in turn to derive two equivalent ellipses. The results of the application of this method to the polar vortex on 850 K on February 21 1979, i.e. at the SSW onset date ($\tau = 0$) two days later than Fig. 2.1 a, is shown in Fig. 2.1 b. The method clearly captures the location, shape, size and orientation of each of the two vortices following the split.

It is perhaps undesirable to be reliant on a vortex split definition based on the behaviour of a particular vortex edge contour (e.g. the vortex edge contour on the 850 K surface as described above). Instead, can the vortex moments themselves be used to define when a vortex split takes place? To answer this question, higher order moments of the $\hat{q}$ field must be considered. In statistics, the bipolarity of a given distribution with respect to a reference distribution is measured by the excess kurtosis $\kappa_4 = J_4/J_2^2 - F$, where J_4 and J_2 are the 4th and 2nd order moments relative to the centroid, and F is determined by the reference distribution. Ideally, we would like an analogous quantity which measures the bipolarity of the PV distribution, while satisfying the following properties:

- It is invariant to translations of $\hat{q}$.
- It is invariant to rotation of $\hat{q}$.
- It is independent of the size of the vortex (or equivalently, the horizontal scale of the $\hat{q}$ distribution).

- It is equal to zero for a vortex of uniform PV that is exactly elliptical, whilst it takes negative values if the vortex becomes 'pinched' in a manner that might lead to a split, and takes positive values (for example) for a diamond-shaped vortex or a vortex undergoing strong filamentation.

After some working, it can be shown that the following quantity satisfies the above five properties

$$\kappa_4 = M_{00}\frac{J_{40} + 2J_{22} + J_{04}}{(J_{20} + J_{02})^2} - \frac{2}{3}\left(\frac{3r^4 + 2r^2 + 3}{(r^2+1)^2}\right), \tag{2.12}$$

where r is the aspect ratio (2.9). The second term in (2.12) serves to ensure that $\kappa_4 = 0$ for an ellipse of uniform vorticity. We refer to κ_4 as the 'excess kurtosis' parameter, following the terminology for the fourth order moments in statistics, and it serves as a measure of the bipolarity of the PV distribution $\hat{q}$. Fig. 2.1 c shows κ_4 as a function of A_1 for the idealized vortex patch of uniform vorticity with boundary in polar coordinates given by

$$r = 1 + A_1 \cos 2\theta. \tag{2.13}$$

As A_1 is increased (see Fig. 2.1 c), the patch vortex becomes progressively more pinched, and the corresponding value of κ_4 falls uniformly. This allows us to define a practical criterion $\kappa_4 < -0.1$ that can serve as an indicator of whether or not a split has taken place. The evolution of κ_4 during actual vortex split SSWs will be discussed in the following section.

2.3 Results

2.3.1 Vortex structure prior to stratospheric sudden warmings

Fig. 2.2 shows the vertically weighted PV field $\bar{Q}$, as a function of equivalent latitude and potential temperature, 9 days before the SSW onset. The aim is to illustrate the initial structure of the polar vortex prior to the SSW taking place, both in a mean sense, as illustrated by the splitting and displacement composites in Fig. 2.2 a and 2.2 b, and also for the January 1987 and February 1979 individual events in Figs. 2.2 c and 2.2 d. The advantage of choosing equivalent latitude and potential temperature, is that when viewed in these coordinates the PV distribution is relatively insensitive to particular synoptic conditions, and the results are consequently largely independent of displacements or distortions of the polar vortex. The January 1987 and February 1979 events are chosen as case studies because they have been previously identified as archetypal SSWs of the displacement and splitting type respectively. The February 1979 event in particular has been much studied (Andrews et al. 1985; Dunkerton and Delisi 1986; Jung et al. 2001; Labitzke 1981; Manney et al. 1994; Palmer 1981), although not primarily from a vortex dynamics perspective as here. Aspects of the January 1987 SSW are discussed by Manney et al. (2005b).

In each panel the polar vortex is clearly visible as a region of elevated $\bar{Q}$ at high latitudes. The thick black contour shows the vortex edge $\bar{Q} = \bar{Q}^*$ defined above, with the black crosses marked on Figs. 2.2 c and 2.2 d showing the equivalent latitudes $\phi_m(\theta)$ of the local maxima in PV gradient. In each case the vortex edge isosurface

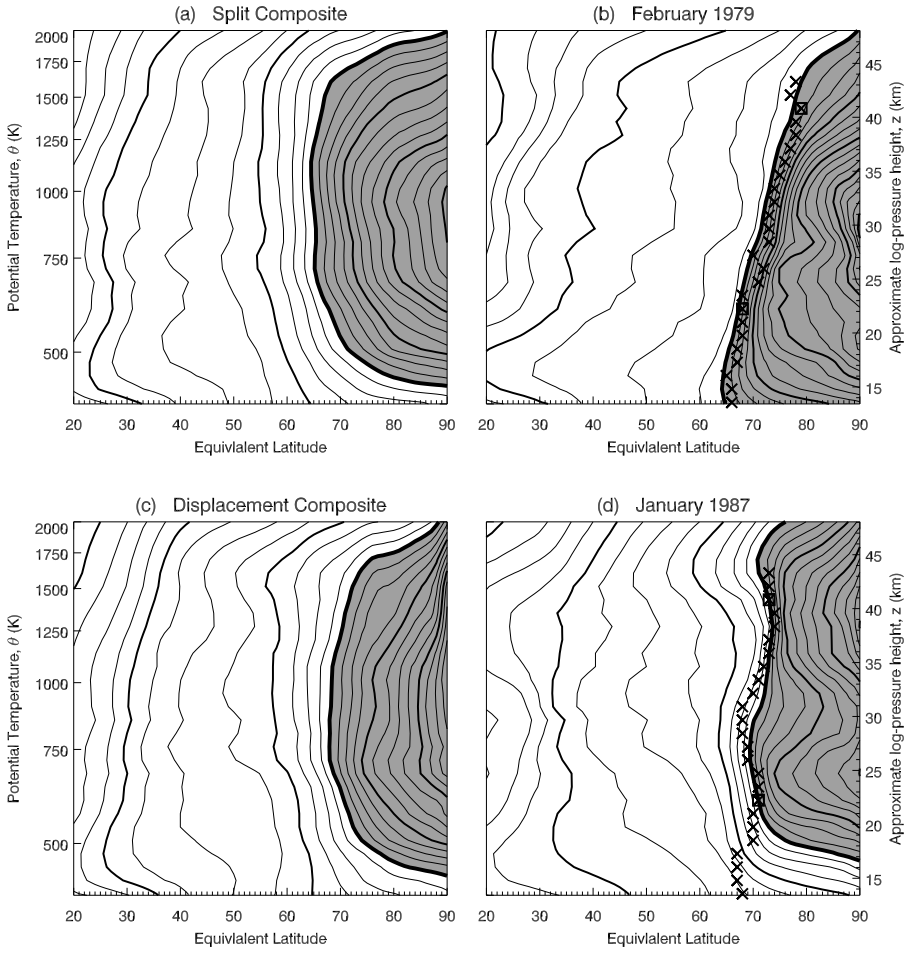

Figure 2.2: (a) Composite vertically weighted potential vorticity $\bar{Q}$ at time lag $\tau = -9$ days, as a function of equivalent latitude and potential temperature, for the vortex split events. The contour interval is 2.93 PVU (1 PVU $\equiv 10^{-6}$ K m^2 s^{-1} kg^{-1}), and every fourth contour is heavy. The thick black contour dividing the shaded (vortex) region from the white (background) plots the vortex edge value $\bar{Q} = \bar{Q}^* = 38$ PVU. (b) As (a) but for the vertically weighted PV Q on February 12 1979 0600UTC ($\tau = -9$ days before the calculated onset date of the February 1979 event). Contour interval, 3.19 PVU, $\bar{Q}^* = 27.5$ PVU. The black crosses show the location $\phi_m(\theta)$ of the maximum of the gradient $|\partial Q/\partial \phi_e|$, used in determining the vortex edge. (c) As (a) but for $\bar{Q}$ at $\tau = -9$ days for the vortex-displacement events. Contour interval, 2.78 PVU, $\bar{Q}^* = 38.0$ PVU. (d) as (b), but for Q on 0000UTC January 23 1987. Contour interval, 4.51 PVU, $\bar{Q}^* = 54.5$ PVU.

$\bar{Q} = \bar{Q}^*$ is a good fit to the equivalent latitudes of the PV gradient maxima across a large altitude range. The region of elevated PV associated with the vortex is located largely between the 400 K and 1600 K isentropic surfaces, justifying our choice of these levels as the upper and lower boundaries of our three-dimensional isosurface plots.

Does the vortex, viewed in this framework, have a characteristic structure that is distinct to the type of SSW that is about to take place? To answer this question comprehensively requires a detailed statistical study that is beyond the scope of the current work. However, Fig. 2.2 a and 2.2 c reveal that in a mean sense the vortex is somewhat stronger prior to splitting SSWs compared to displacement SSWs.[3] Nevertheless, there is substantial variability between individual events in the pre-warming vortex structure, particularly in the slope of the vortex edge and in the vortex strength as a function of height.

The February 1979 and January 1987 vortices have some distinguishing features. Fig. 2.2 b shows that in February 1979 the polar vortex edge slopes significantly poleward with height, i.e. the vortex has a distinctive conical shape. The conical shape may be a consequence of the dynamical activity 'preconditioning' the vortex, as described by Labitzke (1981), and interestingly the Antarctic vortex had a similar conical shape prior to the September 2002 SSW (e.g. Esler et al. 2006). The February 1979 vortex also remains well-defined at altitudes as low as 15 km (400 K). By contrast, the January 1987 vortex shown in Fig. 2.2 d is poorly defined below 20 km, i.e.

[3]Note that the equivalent latitude transformation negates the possibility that the difference in vortex strengths is due to differences in the displacement or distortion of the vortex, as might be the case if a straightforward zonal average were taken.

there is relatively little contrast below 20 km between the vortex and surf-zone PV. The question of whether or not the prior vortex structure is important in determining the subsequent evolution of the polar vortex during the SSW will be the subject of further investigation.

2.3.2 Vortex-splitting stratospheric sudden warmings

The three-dimensional evolution of the Arctic polar vortex, defined here as the airmass enclosed by an isosurface of the vertically weighted PV between the 400 K and 1600 K isentropes, during vortex-splitting SSWs is shown in Fig. 2.3. The left panels show the evolution of the vortex during the composite SSW, illustrated by the isosurface $\bar{Q} = \bar{Q}^*$ defined above, whereas the right panels show the corresponding pictures of the isosurface $Q = Q^*$ for the February 1979 event. Snapshots of the isosurfaces at time lags $\tau = -9, -2, 0$, and $+5$ days are shown to illustrate the vortex behaviour prior to and during the SSW. For comparison, Fig. 2.4 shows the equivalent ellipses on the 505, 715, 1010 and 1425 K isentropic surfaces at the same time lags. Note that for visualization purposes the equivalent ellipses are not shown at their full size (see Fig. 2.4 caption for scaling details).

Figs. 2.3-2.4 reveal that the composite SSW and February 1979 SSW share many key features. At time lag $\tau = -9$ days, i.e. 9 days before the onset time when the vortex edge contour on the 850 K isentrope divides, the composite vortex has a relatively circular cross section at all levels. The composite vortex centroid location ranges between (75°E, 84°N; 450 K) and (40°W, 80°N; 1425 K). This location is somewhat polewards and eastwards of the climatological values for most of the winter,

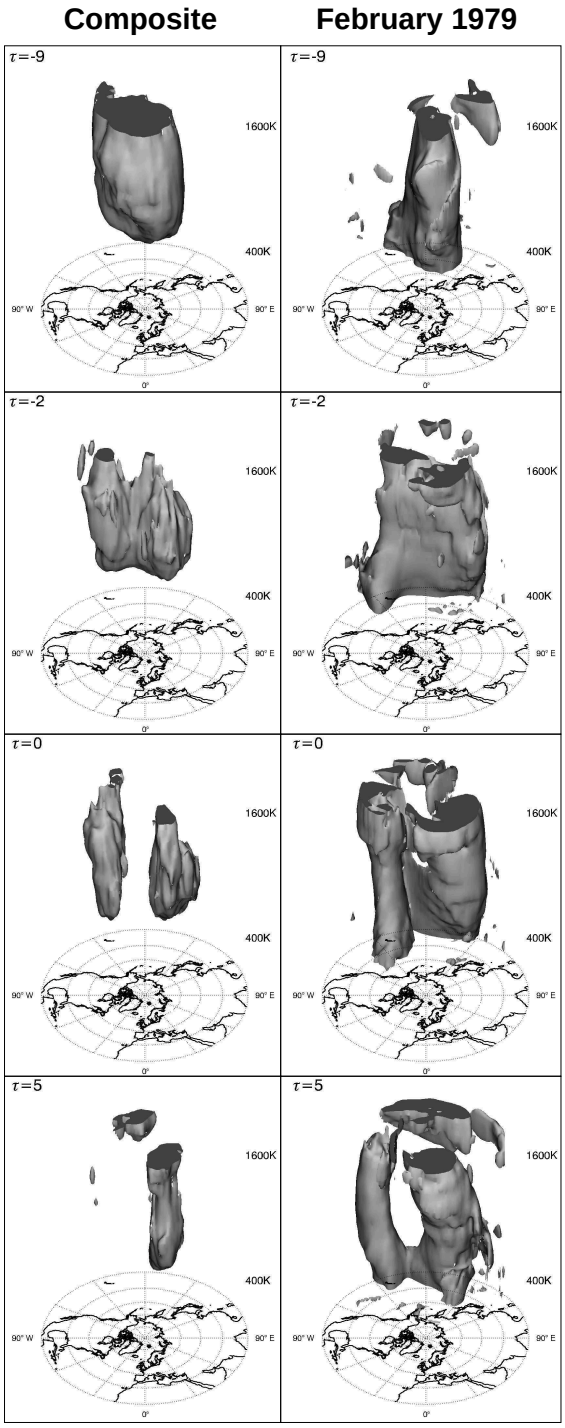

Figure 2.3: Left panels: A three-dimensional isosurface of the composite vertically weighted PV $\bar{Q} = \bar{Q}^* = 38$ PVU, shown between isentropic surfaces 400 K and 1600 K at time lags $\tau = -9, -2, 0$ and 5 days, for the composite vortex split SSW. Right panels: The isosurface $Q = Q^* = 27.5$ PVU for the February 1979 SSW at the same time lags, this time relative to the SSW onset defined to be at February 21 1979 at 0600UTC.

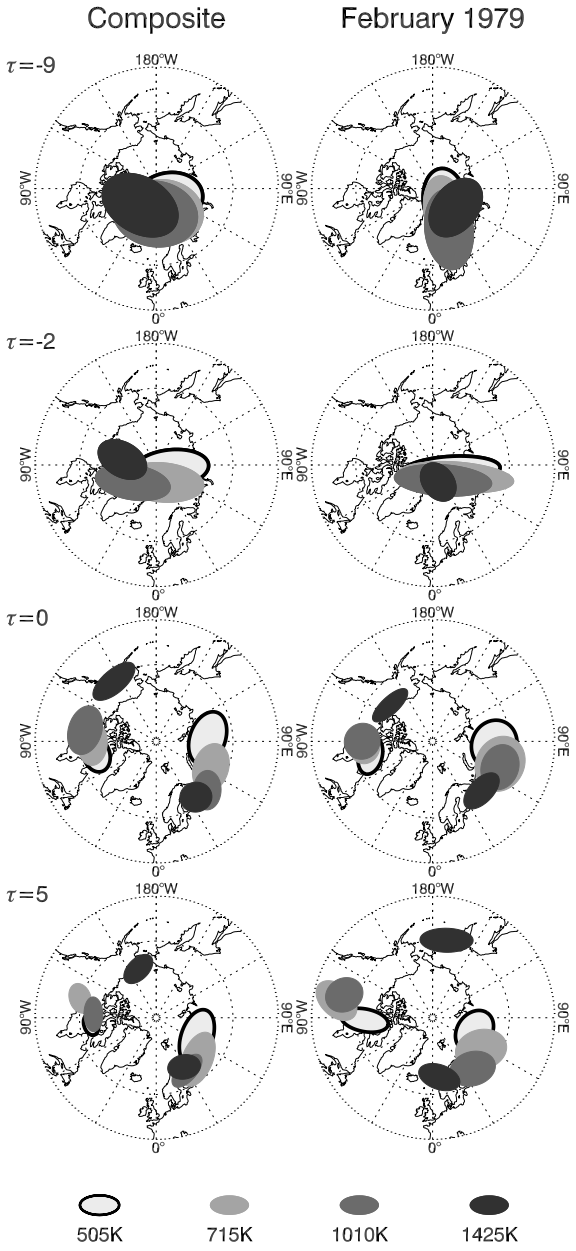

Figure 2.4: The two-dimensional 'equivalent ellipses' on isentropic levels 505 K, 715 K, 1010 K and 1425 K prior to and during the composite SSW (left panels) and February 1979 SSW (right panels), at time lags $\tau = -9, -2, 0$ and 5 days measured relative to the SSW onset time defined in the text in each case. For visualization purposes the equivalent ellipses are plotted at 0.6 times (left panels) and 0.3 times (right panels) their actual area, as defined in equation (2.10).

as seen in Waugh and Dritschel (1999, see their Fig. 4d,f).

As the SSW onset time approaches, the vortex begins to slowly elongate and, for both the composite and February 1979 events, the major axis of the vortex rotates cyclonically. The February 1979 vortex rotates through approximately 60° between $\tau = -9$ days and -2 days. By $\tau = -2$ days the cyclonic rotation slows and the vortex elongates rapidly along an axis approximately parallel to the 80°E-100°W longitude circle. Both the composite and February 1979 vortices have by this stage become 'pinched', and begin to roll-up into two distinct centres located on either side of the original centroid.

By $\tau = 0$, the polar vortex has split completely and two distinct 'daughter' vortices have formed. At τ=0 the centroids of the daughter vortices at 850 K are located approximately 4500 km apart, and are approximately located at 90°W 65°N (Hudson Bay, NE Canada) and 60°E 65°N (Siberia) respectively. The two daughter vortices initially have a relatively 'barotropic' vertical structure, although each has a slight westward tilt with height. As the SSWs develop further, the two vortices begin to rotate in retrograde motion around their (weighted) mean position. The rotation rate varies strongly with altitude, with stronger rotation in the upper stratosphere, as can be seen in Fig. 2.4 at $\tau = 5$ days, when the vortices at 1425 K are at 10°E and 170°E respectively.

The differential rotation inevitably leads to the breakdown of at least one of the vortices. For both the composite and February 1979 event, it is clear that it is the Siberian vortex that is the dominant one. By $\tau = 5$ days the Canadian vortex is no longer detectable in Fig. 2.3 except at high levels. In contrast to at earlier times,

however, the details at this stage vary significantly between individual events. Hence the composite at later times is no longer a meaningful guide to 'typical' behaviour.

The most striking aspect of Figs. 2.3 and 2.4 are the similarities between the composite and February 1979 SSWs. Compositing or otherwise averaging observational data can easily obscure important information. However Figs. 2.3 and 2.4 reveal that:

- The composite vortex undergoes a vortex split resembling the individual events without any filtering of the data other than the PV transformation described above. This is because the vortex-splitting SSWs generally occur at the same orientation in relation to the Earth's surface. At a given time-lag τ the vortex (or vortices) has a fixed characteristic orientation relative to the Earth's surface, and this orientation evolves systematically during the life cycle of the SSW.

- The composite vortex has a 'life cycle' of elongation and cyclonic rotation, followed by splitting and then retrograde rotation of the two resulting vortices about their mean position.

- During the composite SSW, the vortex elongation and split occur near-simultaneously at all altitudes, and both vortices remain coherent for several days.

- For both the composite and February 1979 SSWs it is the Siberian 'daughter' vortex that dominates. A few days after the SSW onset, the Canadian vortex has become sheared out at upper levels (> 1010 K).

Despite the similarities detailed above, there remain some notable differences between the composite and the February 1979 SSW. Firstly, the pair of vortices that result from the February 1979 vortex split have a significantly larger cross sectional

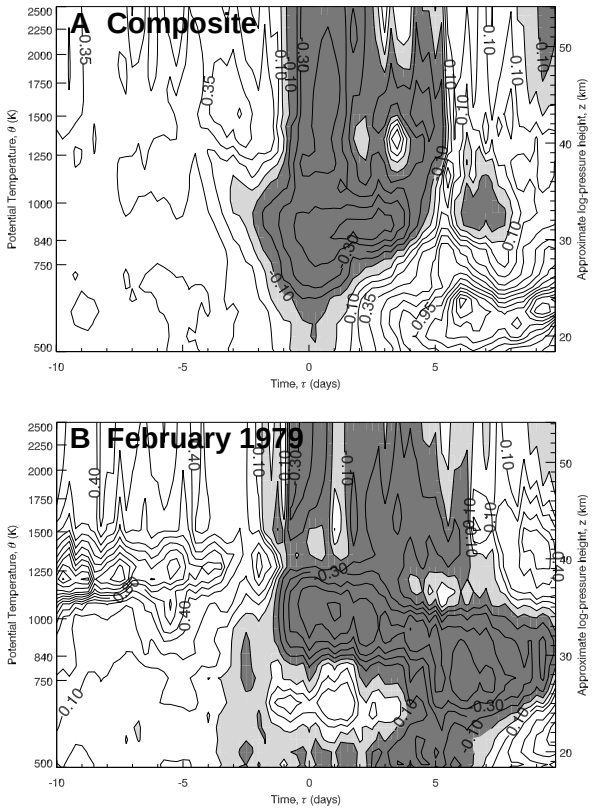

Figure 2.5: Evolution in time and altitude / potential temperature of the 'excess kurtosis' parameter κ_4, defined in equation (2.12). Upper panel shows the composite SSW and lower panel the February 1979 event. The contour interval is 0.1 in each panel. Values below zero are lightly shaded and values below -0.1 are heavily shaded.

area than those of the composite. The explanation is that, as will be seen below, the composite vortices are unrepresentatively small as they are an average over an ensemble of vortices with slightly different locations. Secondly, the February 1979 vortex extends much lower in the stratosphere, as discussed in connection with Fig. 2.2 b. In fact there is considerable variability in the vortex structure at low levels between SSWs. Finally, the vortex is substantially more elongated prior to the February 1979 SSW, and the vortex major axis rotates cyclonically much more rapidly prior to the warming (through 60° as opposed to 10° for the composite).

The near-simultaneous splitting of the vortex at all heights can be highlighted further by examining how the vortex-splitting parameter, or 'excess kurtosis', κ_4 evolves during the composite and February 1979 SSWs. Fig. 2.5 shows a contour plot of κ_4 as a function of altitude and time lag τ in each case. Dark shaded regions indicate where $\kappa_4 < -0.1$, i.e. where a vortex split is diagnosed to have taken place. In each case, it is clear from Fig. 2.5 that the split occurs near-simultaneously over a wide-altitude range, with the split criterion $\kappa_4 < -0.1$ being widely satisfied by $\tau = -1$ day. One unexpected feature of the February 1979 event is the relatively high values of κ_4 in the altitude range 600-750 K throughout the period of the SSW. These high values can be explained by the large difference in the magnitude of the 'daughter vortices' at these levels, with the Siberian vortex being around 3 times greater in magnitude than the Canadian vortex. If the asymmetry between the vortices is sufficiently large, the excess kurtosis parameter does not indicate that a split has taken place. In summary, the κ_4 parameter serves as a useful, objective definition of the onset of a vortex split, although some care is necessary in its interpretation.

To give an impression of the inter-event variability, Fig. 2.6 shows the equivalent ellipses for all 13 splitting events, plus the composite (filled black ellipse) on the 600 K isentropic surface. Note that, as with Fig. 2.4, in order to emphasize differences between events the equivalent ellipses are scaled to be smaller than the actual vortex size. To facilitate comparison with Fig. 2.4 the February 1979 event is also highlighted in Fig. 2.6 as a filled blue ellipse. Fig. 2.6 reveals why the composite method has been successful in describing a 'typical' splitting SSW: 9 out of the 13 events follow quite closely the sequence of rotation, elongation and splitting described for the composite /

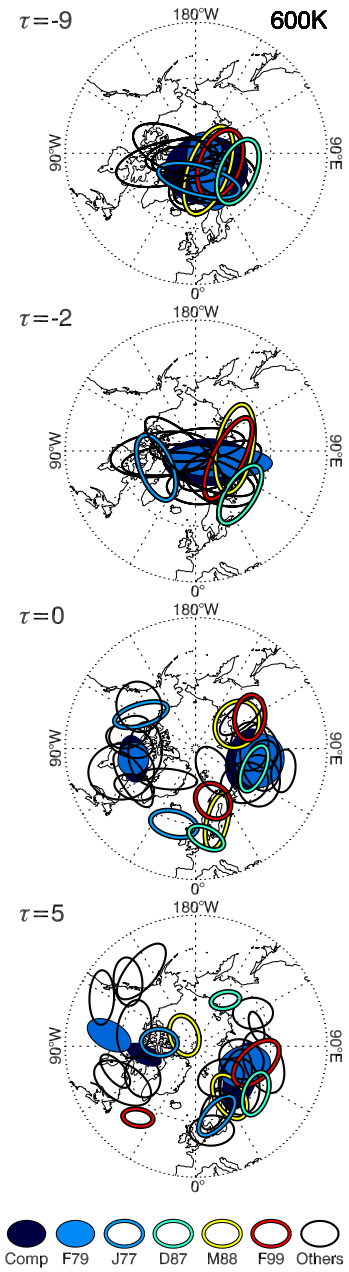

Figure 2.6: The equivalent ellipses for all 13 vortex-splitting SSW events at time lags $\tau = -9, -2, 0$ and 5 days, on the 600 K isentropic surface. The composite SSW is shown as a filled black ellipse, and the February 1979 SSW as a filled blue ellipse. Four further events discussed in the text are highlighted as unfilled coloured ellipses, January 1977 (blue), December 1987 (cyan), March 1988 (yellow), February 1999 (red), and the remaining events are all solid black ellipses. The equivalent ellipses are plotted at 0.3 times their actual area (as in Fig. 2.4).

February 1979 events above.

Four events can be singled out as exhibiting somewhat different behaviour. These are the SSWs of January 1977 (thick blue ellipse), December 1987 (cyan), March 1988 (yellow) and February 1999 (red). The January 1977 event stands out amongst the four events, because the vortex split occurs in a direction parallel to the 160°W-20°E great circle. Notably, it is not selected as an SSW by the CP07 algorithm when NCEP re-analysis data are used in place of ERA-40. For the remaining three events the split occurs approximately parallel to the 20°W-160°E great circle. Further examination of these latter three 'outlier' events reveals significant vertical structure of the daughter vortices compared to the composite 'barotropic' splitting SSW. The European vortex in particular has a baroclinic structure and location that is reminiscent of the vortex structure during displacement SSWs, which will be described next.

Although not pursued here, further investigation of these hybrid characteristics may go some way to determining which SSW behaviour is dominant during warming events which exhibit conditions favouring both types of SSW.

2.3.3 Vortex-displacement stratospheric sudden warmings

The evolution of the Arctic polar vortex during both the composite vortex-displacement and the January 1987 SSW is shown in Figs. 2.7 and 2.8. The time lags relative to the SSW onset time (taken to be 0000UTC January 23 1987 in our case study, following CP07) shown in Figs. 2.7 and 2.8 are $\tau = -10, -1, 1$ and 4 days. As above, Fig. 2.7 shows the three-dimensional isosurface defined by $Q = Q^*$ (or $\bar{Q} = \bar{Q}^*$ for the composite), whilst Fig. 2.8 shows the equivalent ellipses of the polar vortex on a number

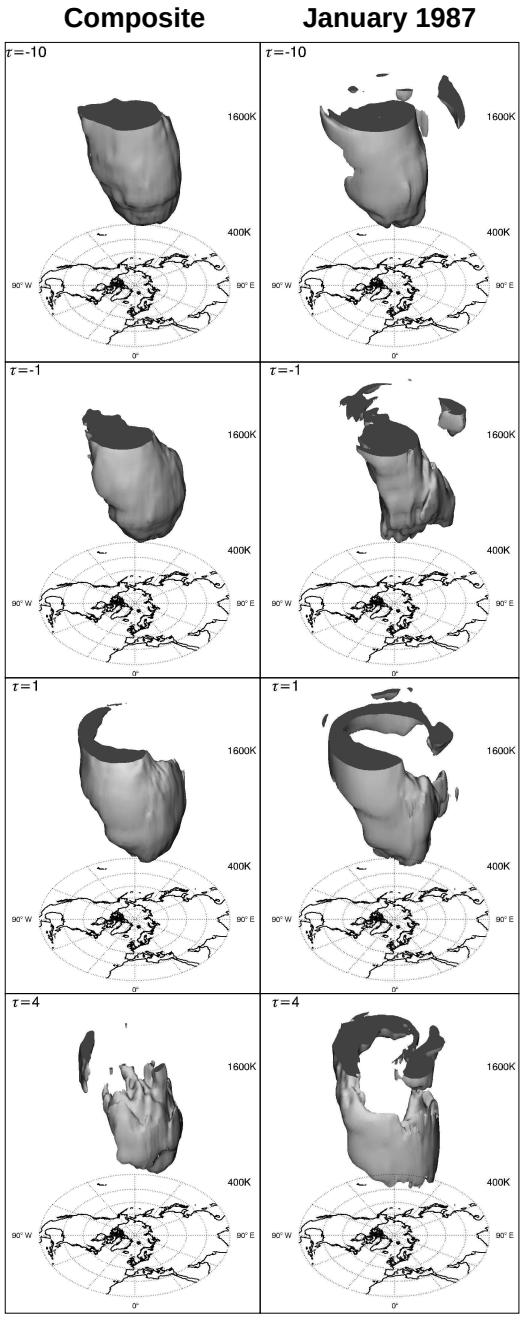

Figure 2.7: Left panels: A three-dimensional isosurface of the composite vertically weighted PV $\bar{Q} = \bar{Q}^* = 38$ PVU, shown between isentropic surfaces 400 K and 1600 K at time lags $\tau = -10, -1, 1$ and 4 days, for the composite vortex split SSW. Right panels: The isosurface $Q = Q^* = 54.5$ PVU for the January 1987 SSW at the same time lags, this time relative to the SSW onset defined to be at 0000UTC January 23 1987.

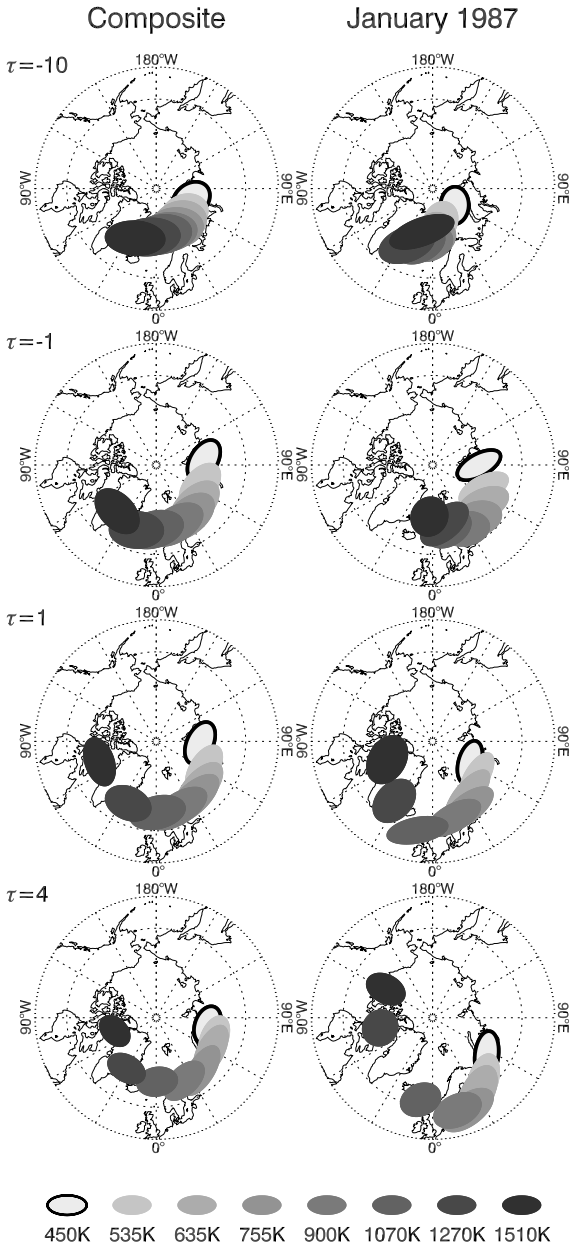

Figure 2.8: The two-dimensional 'equivalent ellipses' on 8 isentropic levels between 450 K and 1510 K prior to and during the composite vortex-displacement SSW (left panels) and January 1987 SSW (right panels). Time lags of $\tau = -10, -1, 1$ and 4 days are relative to the SSW onset time defined in the text in each case. For visualization purposes the equivalent ellipses are plotted at 0.2 times (left panels) and 0.1 times (right panels) their actual area, as defined in equation (2.10).

of isentropic surfaces (450-1510 K). As above, the composite and January 1987 SSWs share many key features. At τ=-10 days, i.e. 10 days before the SSW onset date, the composite and January 1987 vortices each have a near circular cross section in the lower stratosphere. The composite vortex centroid tilts westward and equatorward with altitude, ranging between (70°E, 77.5°N; 450 K) and (15°E 72°N; 1425 K). The vortex is therefore significantly further equatorward, and has a substantially greater tilt with altitude prior to a displacement SSW than a splitting SSW. Also, the vortex is displaced further from the pole than the climatological displacement of 10-14° (Waugh and Dritschel 1999, see their Fig. 4,f), indicating that the vortex may be somewhat 'pre-conditioned' prior to a displacement SSW (see e.g. Labitzke 1981). Both the composite and January 1987 vortices are seen to be more elliptical at higher levels. The major axis of the ellipse is roughly parallel to the direction of the westward tilt along the 70-80°N latitude band in each case.

As the onset time for the SSW draws closer, the westward tilt of the vortex with altitude increases, and the entire vortex simultaneously moves equatorwards, so that by $\tau = -1$ day the composite vortex centroid lies within the 67-74°N latitude band. The lower vortex moves slowly eastwards whilst the upper vortex moves rapidly westwards so that at $\tau = -1$ day the composite vortex centroid at 450 K is separated from that at 1425 K by 135° longitude. The process continues through the SSW onset itself, and by τ=1 day the composite vortex centroid position ranges between (95°E, 75°N; 450 K) and (53°W, 68°N; 1425 K), i.e. the vortex spans nearly half the globe in longitude. During this period the upper vortex is strained by the large-scale flow, and by τ=1 day the vortex consequently has a crescent-like cross section.

During the next few days the deformation of the upper vortex continues until it is thinned and stirred into the background. By τ=4 days the averaging process used to create the composite acts to diffuse the upper level composite vortex over a wide region, hence the upper vortex is entirely absent from the τ=4 days three-dimensional representation. In the January 1987 event the upper and lower parts of the vortex have become largely disconnected. At this stage the life cycle of the displacement SSW is essentially complete. The remnants of the vortex in the lower stratosphere are located equatorward and eastward compared to the vortex at the beginning of the life cycle (at $\tau = -10$ days). This typical life cycle displayed by the composite vortex is in agreement with individual SSWs as discussed in Manney et al. (1999), Manney et al. (2005b), and references therein.

Apart from minor details of vortex orientation and timing, remarkably few differences are apparent between the January 1987 SSW and the composite in Figs. 2.7 and 2.8. The January 1987 event is revealed to be somewhat stronger than typical (Manney et al. 2005b), as the vortex at middle levels (850 K) is displaced as far equatorwards as 55°N, compared to 67°N for the composite.

An impression of the inter-event variability for all 15 displacement SSWs is given in Fig. 2.9. The vortex centroids at 450 K, 850 K and 1425 K are plotted for each event at $\tau = -10, -1, 1$ and 4 days. The lines joining these three points intersect the vortex centroids at the intermediate levels. A remarkably consistent picture is revealed for all of the SSWs, with the centroids quite tightly clustered at all levels at both early times and as the SSW develops. Differences in timing and vertical structure cause the largest variability in centroid position to occur for the middle vortex level (850 K).

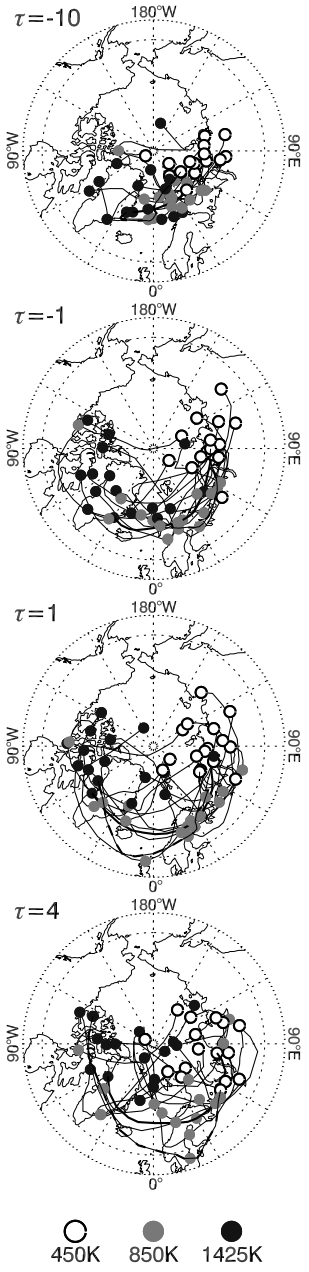

Figure 2.9: The location of the vortex centroid (circles) at different isentropic levels (white circle 450 K, grey circle 850 K, black circle 1425 K) for all 15 vortex-displacement SSWs, at $\tau = -10, -1, 1$ and 4 days. The lines joining the points trace out the centroid location at intermediate levels.

For some events the 850 K centroid remains close to the lower vortex (450 K) centroid during the SSW onset, whereas for other events the 850 K centroid follows the upper vortex (1425 K) as it moves rapidly westwards and equatorwards. By $\tau = 4$ days the upper vortex (1425 K) centroids also exhibit a large scatter. This is consistent with the lack of coherence of the upper vortex at this time, due to the effects of the straining flow thinning out and mixing the vortex into the background at upper levels. Overall, however, Fig. 2.9 reveals a remarkable degree of similarity between all of the displacement SSW events.

2.4 Proposed benchmarks

The results above suggest that a new set of modelling benchmarks may now be defined, against which the dynamical performance of GCMs may evaluated. These new benchmarks, in contrast to those proposed in section 7 of CP07, are primarily qualitative in character and relate to the evolving shape of the polar vortex during SSWs.

Specifically, after analyzing model output for vortex-splitting events, one might answer the following questions:

S-1. Does the vortex split near-simultaneously, and with an approximate barotropic structure, over the 20-40 km altitude range?

S-2. Is the mean vortex orientation at the onset time of the vortex split such that the daughter vortices are approximately located over Canada and Siberia?

S-3. Is the SSW followed by a retrograde rotation of the daughter vortices about a

common centroid, and is this rotation more pronounced in the upper stratosphere?

S-4. Is the Siberian daughter vortex typically larger than its Canadian counterpart?

For the displacement SSWs different questions must be asked.

D-1. Is the vortex centroid located in the 90°E-0°E sector prior to the warming, with a distinct westward tilt with height?

D-2. Is the vortex evolution primarily baroclinic, with the westward tilt with height increasing rapidly as the onset time approaches and an accompanying equatorward displacement of the vortex centroid at all heights?

D-3. At the end of the SSW, does the longitude of the vortex centroid vary strongly with altitude, with up to 180° separating the centroid position in the lower stratosphere (450 K) from that in the upper stratosphere (1425 K)?

D-4. At the end of the SSW is the polar vortex almost entirely contained within the Atlantic hemisphere (90°E-90°W)?

It is proposed that straightforward application of the methodology of section 2.2 to absolute vorticity GCM output fields (or PV output fields if available) should be sufficient to reveal the bulk behavioural properties of the vortex during a SSW, which can then be compared with the benchmarks above.

2.5 Conclusions

The current investigation has focused on the evolution of the Arctic polar vortex during major midwinter stratospheric sudden warmings (SSWs). Time-lag composites and vortex moment diagnostics have been used to investigate the distinctive behaviour of the polar vortex during both displacement and splitting SSW events. The main results, answering, as far as possible, the questions posed in the introduction, can be summarized as follows:

- The Arctic polar vortex has a highly characteristic vertical structure during the two different types of SSW. During vortex-splitting events, the vortex remains nearly barotropic, with the vortex split occurring near-simultaneously over a large altitude range (at least 20-40 km, see Figs. 2.3 and 2.4). In contrast, during vortex-displacement events, the vortex displacement off the pole increases with altitude above 30 km, as does the accompanying elongation of the vortex during the SSW, as can be seen in Figs. 2.7 and 2.8)..

- For both types of SSW, there is relatively little variation between individual events in the orientation of the developing vortex relative to the underlying topography, i.e. SSWs of each type are to a large extent fixed in relation to the Earth's surface. For displacement events the vortex is located in the 90°E-0°E sector prior to the warming, and by the end of the warming spans the entire Atlantic hemisphere from 90°E (lower vortex) to 90°W (upper vortex). There is somewhat more variability in the case of vortex-splitting events (see Fig. 2.6), but the dominant direction for the splits is found to be parallel to the

60°E-120°W great circle.

- Each type of SSW exhibits a distinct life cycle, in the sense that the centroid position, orientation and elongation of the Arctic polar vortex evolves in a similar fashion during the onset and occurrence of individual splitting events, with a separate behaviour during displacement events. Splitting events are characterized by a rapid increase of the aspect ratio of the vortex a few days prior to the SSW ($\tau = -4$ to -2 days), followed by the roll-up of the elongated vortex into two distinct 'daughter' vortices ($\tau = -2$ to 0 days). The daughter vortices propagate rapidly apart to a distance of up to 5000 km by $\tau = 1$ day, after which they experience a retrograde rotation around their common centroid, usually leading to the destruction of the weaker Canadian vortex and reformation of the main vortex around the stronger Siberian vortex. For vortex-displacement events, the separation in longitude between the lower vortex (450 K) and the upper vortex (1425 K) increases uniformly during the SSW from 85° longitude (at τ =-10 days) to 150° longitude (by τ =1 day). By the end of the displacement SSW life cycle (at τ =4 days) the upper part of the vortex is destroyed by deformation, and the vortex subsequently reforms around its remnants in the lower stratosphere.

In addition to providing the modelling benchmarks listed above, it is intended that the results of this study will be of use to researchers seeking to differentiate between alternative dynamical theories of SSWs. The vertical structure of the vortex is necessarily a key diagnostic of the vertical propagation and nonlinear behaviour of Rossby waves in the polar stratosphere. Following Matsuno (1971), SSWs of either

type are often described as resulting from the vertical propagation and breaking of stationary planetary waves generated in the troposphere. Alternatively, it has been suggested (O'Neill and Pope 1988; Scott and Dritschel 2006) that the interaction between the polar vortex and Aleutian anticyclone may be important in the development of SSWs. Further, Tung and Lindzen (1979a) and Plumb (1981) have suggested that SSWs might occur following the resonant excitation of a normal mode of the stratosphere, an idea recently developed by Esler and Scott (2005) and Esler et al. (2006) who have suggested that the 'barotropic' mode of the polar vortex is the relevant mode in the case of vortex-splitting events. Note that if the resonant excitation of the vortex normal modes in the above theories is due to stationary waves, then the resulting SSWs will be fixed in relation to the Earth's surface as observed in this study. The 'vortex dynamics' view of SSWs may be of particular relevance to splitting events, as the vortex split occurs rapidly and in a manner strongly reminiscent of Love's instability of Kirchhoff's elliptical vortex in two-dimensional hydrodynamics (Dritschel 1986; Love 1893), an aspect that will be further investigated. The climatology presented here, of the vertical structure of the Arctic polar vortex during observed SSWs, should be of considerable assistance in quantitatively distinguishing between the above theories.

Chapter 3

A Hierarchy of Models for the Winter Polar Stratosphere

3.1 Introduction

In this chapter, a hierarchy of models is introduced with which we will investigate polar vortex breakdown during SSWs in the following chapters of this thesis. Using this approach, vortex behaviour in simple models is used to predict behaviour in increasingly complex models of the stratosphere. By gaining an understanding of mechanisms leading to SSW-like vortex breakdown in the simplest systems, we can therefore investigate the role of similar mechanisms when studying vortex breakdown in more realistic models.

When introducing the hierarchy of models, we will start with the most complex models and will work backwards towards more idealized models. At first, this may seem at odds with the concept of simple models predicting behaviour in more complex

models. However, by introducing the most complex models first, it is easier to justify the conditions and assumptions leading to the next, more idealized, member of the hierarchy.

The structure of the chapter can be broadly split into two sections. The first section will start with a brief description of the primitive equations governing fluid motion in an idealized three-dimensional model of the stratosphere, giving the assumptions under which they are valid. From the primitive equations, we will then derive the three-dimensional quasi-geostrophic equations (QG hereafter), which will be the most complex member of the hierarchy of models used in this thesis. This derivation includes a discussion on appropriate boundary conditions for the system, including the representation of forcing at the lower boundary. As part of this derivation, the Lamb vertical modes of the model are introduced. It is then shown that fluid motion in the atmosphere can be approximated by a single layer of fluid which satisfies the two-dimensional shallow water QG equations.

In the second section, starting from the most idealized member of our hierarchy of models, numerical algorithms are introduced which will enable us to perform fully nonlinear simulations of vortex behaviour in each of our models.

3.2 Equations of motion in the atmosphere

Fluid motion of the air in the Earth's atmosphere is governed by the equations of motion

$$\frac{D\mathbf{u}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{u} \;=\; -\frac{1}{\rho}\nabla p - \nabla\Phi_e + \mathbf{F}, \tag{3.1}$$

subject to the mass conservation relation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) \;=\; 0. \tag{3.2}$$

Here, $\mathbf{F}$ represents nonconservative forces (e.g. frictional), $\mathbf{u}$ is the fluid velocity, $\mathbf{\Omega}$ is the rotation vector due of the Earth's rotation, p is pressure, ρ is the fluid density, $\Phi_e = \Phi - \nabla(\mathbf{\Omega} \times (\mathbf{\Omega} \times \mathbf{r}))$ is the effective gravitational potential, Φ is the gravitational geopotential and $\mathbf{r}$ is a position vector in spherical coordinates. The operator D/Dt is defined as

$$\frac{D}{Dt} \;\equiv\; \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla,$$

and the gravitational geopotential is defined as the work required to raise a unit mass from mean sea level to height z^* above the Earth's surface

$$\Phi \;=\; \int_0^{z^*} g dz.$$

In addition to these equations of motion, we also have the ideal gas law

$$p \;=\; \rho R T, \tag{3.3}$$

where R is the ideal gas constant for dry air, and the first law of thermodynamics applied to isentropic flow

$$\frac{D\theta}{Dt} \;=\; Q, \tag{3.4}$$

where Q is a diabatic heating term which is zero for adiabatic flow. The potential temperature θ in this context is defined as being "the temperature that a parcel of dry air at pressure p and temperature T would acquire if it were expanded or compressed

adiabatically to the reference pressure $p_s = 1000$ hPa". In equation form, θ is given by

$$\theta \equiv = T\left(\frac{p_s}{p}\right)^{\kappa}, \tag{3.5}$$

where $\kappa \equiv R/c_p \approx 2/7$ and c_p is the specific heat at constant pressure.

3.2.1 Primitive equations

In most cases, these equations are far more complicated than is necessary when considering flows which are observed in the stratosphere. A comparatively more simple framework governing fluid flow in the atmosphere is that of the primitive equations. In order to derive the primitive equations, we first note several assumptions regarding the behaviour and geometry of the Earth's atmosphere. These assumptions are (e.g. Vallis 2006):

1. *Shallow fluid approximation*: If H is a scale height for the atmosphere and a the radius of the approximately spherical Earth, it is assumed that $H/a \ll 1$.

2. *Traditional approximation*: The horizontal component of the rotation vector $\mathbf{\Omega}$ is neglected.

3. *Spherical geopotential approximation*: If the Earth is approximately spherical, surfaces of constant geopotential are taken as spherical shells around the Earth.

4. *Anelastic approximation*: All variables are expressed as deviations from a vertically varying reference profile, e.g. $p(x, y, z, t) = p_0(z) + p'(x, y, t)$, where $p'/p_0 \ll 1$.

5. *Hydrostatic approximation*: The vertical pressure gradient and forces due to gravity balance one another.

6. *f-plane approximation*: The coordinate system is defined in terms of Cartesian coordinates (x, y, z) which lie in the tangent plane to a given point on the Earth's surface. The height coordinate in this Cartesian coordinate system is taken to be the log-pressure height, given by $z = H \ln(p_s/p)$.

It can be seen from observational data that the winter polar stratosphere is close to being isothermal, with very little variation in temperature with height (Andrews et al. 1985). Therefore, it is specified that the reference profile for the temperature is constant at all heights such that $T_0 = T_s$, where $T_s \approx 210$ K from observations.

Once these assumptions have been made, simplification of the equations of motion (3.1) and mass conservation (3.2), in conjunction with the first law of thermodynamics (3.4), gives

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_H + w\frac{\partial}{\partial z}\right)\mathbf{u} + 2\Omega\mathbf{k} \times \mathbf{u} = -\nabla_H \Phi, \tag{3.6}$$

$$\frac{\partial \Phi}{\partial z} = \frac{RT}{H}, \tag{3.7}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\theta + w\frac{d\theta_0}{dz} = 0, \tag{3.8}$$

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{1}{\rho_0}\frac{\partial(\rho_0 w)}{\partial z} = 0. \tag{3.9}$$

Here, $\nabla_H = (\partial_x, \partial_y)$ is the horizontal gradient operator, $\mathbf{u} = (u, v)$ is the horizontal velocity, w is the vertical velocity, and diabatic heating terms and nonconservative

forces have been neglected such that $Q = 0$ and $\mathbf{F} = 0$. In equation (3.7), the vertical momentum equation has been replaced by the hydrostatic approximation, and the geopotential has replaced pressure using

$$\frac{\partial \Phi}{\partial(\ln p)} = -RT,$$

where $H = RT_s/g$.

Using $\Phi_z = RT/H$ and $\theta = T_s(p_s/p)^\kappa = T_s \exp\{\kappa z/H\}$, it is convenient to make another change variables in (3.8) to rewrite the thermodynamic equation in the form

$$\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\theta + w\frac{d\theta_0}{dz} = \left[\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)T + w\left(\frac{dT_0}{dz} + \frac{\kappa}{H}T_0\right)\right]e^{\kappa z/H},$$

$$= \frac{H}{R}\left[\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\frac{\partial \Phi}{\partial z} + wN^2\right]e^{\kappa z/H},$$

where the constant buoyancy frequency N in (3.10) is the Brunt-Väisälä frequency given by

$$N^2 = \frac{\kappa R T_s}{H^2}.$$

For an isothermal stratosphere with $T_s = 210$ K, the buoyancy frequency is $N \approx 0.02$ s^{-1}. Therefore, an alternative to the thermodynamic equation in (3.8) is

$$\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\frac{\partial \Phi}{\partial z} + wN^2 = 0. \tag{3.10}$$

The lower boundary of the domain is represented by setting the geopotential on the log-pressure height $z = z_s(x, y, t)$ to be

$$\Phi(x, y, z_s, t) = gh(x, y), \tag{3.11}$$

where g is the acceleration due to gravity and $h(x, y)$ is the height of the surface $z = z_s$ in log-pressure coordinates. Taking the material derivative of (3.11) gives

$$\frac{D}{Dt}\left(\Phi - gh\right) \;=\; 0 \quad \text{on} \quad z = z_s,$$

which is rewritten as

$$\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\left(\Phi - gh\right) + w\frac{d\Phi_0}{dz} \;=\; 0. \tag{3.12}$$

Rearranging the thermodynamic equation (3.10) gives an expression for w

$$w \;=\; -\frac{1}{N^2}\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\frac{\partial \Phi}{\partial z},$$

which when substituted into equation (3.12) yields

$$\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\left\{\Phi - gh - \frac{1}{N^2}\frac{d\Phi_0}{dz}\frac{\partial \Phi}{\partial z}\right\} \;=\; 0 \quad \text{on} \quad z = z_s, \tag{3.13}$$

where $d\Phi_0/dz$ can be taken inside the curly bracket as the reference profile Φ_0 is a function of z alone. Finally, substitution of $d\Phi_0/dz = g$ into (3.13) gives

$$\left(\frac{\partial}{\partial t} + \mathbf{u}.\nabla_H\right)\left\{\frac{\partial \Phi}{\partial z} - \frac{\kappa}{H}\Phi + N^2 h\right\} \;=\; 0 \quad \text{on} \quad z = z_s, \tag{3.14}$$

where $N^2/g = \kappa/H$ has been used for the middle term inside the bracket.

Equations (3.6)-(3.9) are the primitive equations in Cartesian coordinates. Note that these equations describe a system which is semi-infinite in the vertical log-pressure height coordinate, as no upper boundary has been specified in the model.

3.2.2 The quasi-geostrophic approximation in an isothermal atmosphere

The primitive equations, while being a simplified set of equations when compared to the full equations of motion for the atmosphere, are still in themselves relatively

complex. In fact, for the stratospheric flows which we will be considering, further simplifications may be made without significantly sacrificing the realism of the model. In this section we will derive the QG equations, describing flow in a compressible atmosphere in the presence of stratification and rotation. These equations are derived by making a series of physical assumptions about stratospheric fluid flow in the primitive equations.

We start by nondimensionalizing the primitive equations (3.6)-(3.10) by setting $\mathbf{x} = L\tilde{\mathbf{x}}$, $\mathbf{u} = U\tilde{\mathbf{u}}$ and $z = H\tilde{z}$, and take the Coriolis parameter $f = f_0$ to be constant and evaluated at the pole, giving

$$\left(\frac{\partial}{\partial \tilde{t}} + \tilde{\mathbf{u}}.\tilde{\nabla}_H\right)\tilde{\mathbf{u}} + \frac{1}{\epsilon}\mathbf{k}\times\tilde{\mathbf{u}} = -\frac{1}{\epsilon}\tilde{\nabla}_H\tilde{\psi}, \tag{3.15}$$

$$\frac{\partial\tilde{\psi}}{\partial\tilde{z}} = \frac{RT}{f_0UL}, \tag{3.16}$$

$$\left(\frac{\partial}{\partial \tilde{t}} + \tilde{\mathbf{u}}.\tilde{\nabla}_H\right)\frac{\partial\tilde{\psi}}{\partial\tilde{z}} + \tilde{w}\frac{\epsilon}{\mathrm{Fr}^2} = 0, \tag{3.17}$$

$$\frac{\partial\tilde{u}}{\partial\tilde{x}} + \frac{\partial\tilde{v}}{\partial\tilde{y}} + \frac{1}{\rho_0}\frac{\partial(\rho_0\tilde{w})}{\partial\tilde{z}} = 0, \tag{3.18}$$

where scaling of the vertical velocity $w = (UH/L)\tilde{w}$ and time $t = \tau\tilde{t} = (L/U)\tilde{t}$ are found as part of the scaling process. Other quantities in equations (3.15)-(3.18) are $\psi = \Phi/f_0$ with $\psi = (UL)\tilde{\psi}$, $\epsilon = \mathrm{Ro}$ with $\mathrm{Ro} = U/f_0L \ll 1$, and $\mathrm{Fr} = U/NH = \mathrm{Bu}^{-1/2}\epsilon$. Nondimensional parameters are Ro, which is the Rossby number, Fr which is the Froude number, and $\mathrm{Bu} = (NH/f_0L)^2$ which is the Burger number, with $\mathrm{Bu} \sim 1$ indicating that stratification and rotation play an equal role in the dynamics of the

flow. By comparison with the Rossby number, it is seen that the time-scale has been chosen to satisfy $\tau \ll f_0$. This time scaling filters out effects due to high frequency acoustic waves. The conditions $\mathrm{Ro} \ll 1$ and $\tau \ll f_0$ are equivalent to restricting our attention to "large-scale, low-frequency motions" (see Andrews et al. 1985, section 3.2.3).

For convenience, all tildes on dimensionless quantities are dropped, and it is assumed that all variables are in dimensionless form unless specified otherwise. The velocity field is now expressed as $\mathbf{u} = \mathbf{u}_g + \epsilon \mathbf{u}_a$[1], where the geostrophic velocity $\mathbf{u}_g = (u_g, v_g, 0)$ is defined as the solution of $\mathbf{k} \times \mathbf{u} = -\nabla_H \psi$, that is

$$u_g = -\frac{\partial \psi}{\partial y}, \qquad v_g = \frac{\partial \psi}{\partial x}, \tag{3.19}$$

and $\mathbf{u}_a$ are ageostrophic corrections to the velocity field. To leading order in ϵ, application of the conservation of mass (3.18) gives

$$\frac{\partial u_g}{\partial x} + \frac{\partial v_g}{\partial y} = 0. \tag{3.20}$$

Using this quasi-geostrophic assumption for the horizontal velocity $\mathbf{u}$, the conservation of planar horizontal momentum (3.15) becomes

$$\left(\frac{\partial}{\partial t} + (\mathbf{u}_g + \epsilon \mathbf{u}_a) . \nabla_H\right)(u_g + \epsilon u_a) - \frac{1}{\epsilon}(v_g + \epsilon v_a) = -\frac{1}{\epsilon}\frac{\partial \psi}{\partial x}, \tag{3.21}$$

$$\left(\frac{\partial}{\partial t} + (\mathbf{u}_g + \epsilon \mathbf{u}_a) . \nabla_H\right)(v_g + \epsilon v_a) + \frac{1}{\epsilon}(u_g + \epsilon u_a) = -\frac{1}{\epsilon}\frac{\partial \psi}{\partial y}. \tag{3.22}$$

Taking $\partial_x(3.22) - \partial_y(3.21)$ gives

$$\left(\frac{\partial}{\partial t} + (\mathbf{u}_g + \epsilon \mathbf{u}_a) . \nabla_H\right)\left(\frac{\partial v_g}{\partial x} - \frac{\partial u_g}{\partial y} + \epsilon\frac{\partial v_a}{\partial x} - \epsilon\frac{\partial u_a}{\partial y}\right)$$

[1]Formally, a series expansion approach $\mathbf{u} = \sum_{i=0}^{\infty} \epsilon^i \mathbf{u}_i$ should be used here, similar to that seen in section 3.2.3. However, in the interests of brevity, a less rigorous alternative is given here.

$$+\frac{1}{\epsilon}\left(\frac{\partial u_g}{\partial x}+\frac{\partial v_g}{\partial y}+\epsilon\frac{\partial u_a}{\partial x}+\epsilon\frac{\partial v_a}{\partial y}\right) \;=\; 0. \qquad (3.23)$$

To leading order $O(1/\epsilon)$, (3.23) gives

$$\frac{\partial u_g}{\partial x}+\frac{\partial v_g}{\partial y} \;=\; 0,$$

which agrees with the non-divergence of the geostrophic velocities given in (3.20). Investigation of $O(1)$ terms in (3.23) gives

$$\frac{D_g}{Dt}\left(\frac{\partial v_g}{\partial x}-\frac{\partial u_g}{\partial y}\right)+\left(\frac{\partial u_a}{\partial x}+\frac{\partial v_a}{\partial y}\right) \;=\; 0,$$

which when combined with the conservation of mass for the ageostrophic velocity correction

$$\frac{\partial u_a}{\partial x}+\frac{\partial v_a}{\partial y}+\frac{1}{\rho_0}\frac{\partial(\rho_0 w_a)}{\partial z} \;=\; 0,$$

gives

$$\frac{D_g}{Dt}\left(\frac{\partial v_g}{\partial x}-\frac{\partial u_g}{\partial y}\right)-\frac{1}{\rho_0}\frac{\partial\left(\rho_0 w_a\right)}{\partial z} \;=\; 0, \qquad (3.24)$$

where $D_g/Dt = \partial_t + \mathbf{u}_g.\nabla_H$ is the material derivative following the geostrophic velocity. Recalling that $\mathrm{Fr} = \mathrm{Bu}^{-1/2}\epsilon$ and $w = \epsilon w_a$, at order $O(1)$ the thermodynamic equation gives

$$w_a = -\mathrm{Bu}^{-1}\frac{D_g}{Dt}\left(\frac{\partial\psi}{\partial z}\right).$$

Substituting this expression for w_a into (3.24) and simplifying gives

$$\frac{D_g}{Dt}\left\{\frac{\partial^2\psi}{\partial x^2}+\frac{\partial^2\psi}{\partial y^2}+\frac{\mathrm{Bu}^{-1}}{\rho_0(z)}\frac{\partial}{\partial z}\left[\rho_0(z)\frac{\partial\psi}{\partial z}\right]\right\} \;=\; 0. \qquad (3.25)$$

Nondimensionalizing the lower boundary condition (3.14) in the same way as the primitive equations, and considering small forcing heights in the nondimensional system, $h(x,y) = \epsilon\eta(x,y)$, gives the nondimensional boundary condition for the isothermal QG system as

$$\frac{D_g}{Dt}\left\{\frac{\partial\psi}{\partial z} - \kappa\psi + \text{Bu }\eta\right\} = 0 \quad \text{on} \quad z = 0. \tag{3.26}$$

Finally, redimensionalizing all quantities yields

$$\frac{D_g}{Dt}\left\{\frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{1}{\rho_0(z)}\frac{\partial}{\partial z}\left[\frac{f_0^2}{N^2}\rho_0(z)\frac{\partial\psi}{\partial z}\right]\right\} = 0, \tag{3.27}$$

along with the the lower boundary condition

$$\frac{D_g}{Dt}\left\{\frac{\partial\psi}{\partial z} - \frac{\kappa}{H}\psi + \frac{N^2}{f_0}h\right\} = 0 \quad \text{on} \quad z = 0. \tag{3.28}$$

Equation (3.28) is an advection equation for the potential temperature θ on the lower boundary of the model. If initially we set

$$\frac{\partial\psi}{\partial z} - \frac{\kappa}{H}\psi + \frac{N^2}{f_0}h = 0 \quad \text{on} \quad z = 0$$

then for times thereafter, the lower boundary condition is given by

$$\frac{\partial\psi}{\partial z} - \frac{\kappa}{H}\psi = -\frac{N^2}{f_0}h \quad \text{on} \quad z = 0. \tag{3.29}$$

Equation (3.27) can be interpreted as conservation of the QG potential vorticity

$$\frac{D_g}{Dt}(q - f_0) = 0,$$

where

$$q(\mathbf{r}) = f_0 + \nabla_H^2\psi + \frac{1}{\rho_0(z)}\frac{\partial}{\partial z}\left[\frac{f_0^2}{N^2}\rho_0(z)\frac{\partial\psi}{\partial z}\right]. \tag{3.30}$$

The $\nabla_H^2\psi$ term on the right hand side of (3.30) represents the relative vorticity on a given isentropic surface, and the final term on the right hand side represents potential vorticity contributions due to "vortex stretching".

The form of the lower boundary condition

One of the important aspects of the QG f-plane model is the form of the lower boundary condition arises from the fact that potential temperature θ is conserved on the bottom boundary, as seen from equation (3.8). An important question to ask, is how do approximations to this conservation of potential temperature on the lower boundary affect the resulting dynamics of the system?

One common approximation is conservation of the normal temperature (T) on the lower boundary, rather than potential temperature (for example Dritschel and Saravanan 1994; Esler and Scott 2005; Waugh and Dritschel 1999). This requires that $D_g T/Dt = 0$ on the lower boundary, which is equivalent to taking the value of κ appearing in (3.14) to be zero. The corresponding "Neumann" lower boundary condition for the streamfunction, termed the "ersatz" boundary condition by Esler and Scott (2005), is

$$\frac{\partial \psi}{\partial z} = -\frac{N(z)^2}{f_0} h \quad \text{on} \quad z = 0. \tag{3.31}$$

Another common approximation to the lower boundary condition is one in which the geopotential is specified on the lower boundary, giving

$$\psi = h \quad \text{on} \quad z = 0. \tag{3.32}$$

The choice of the lower boundary condition is important when calculating the vertical normal modes of the model, as shall be seen in the following section.

Vertical normal modes in a quasi-geostrophic compressible atmosphere

Vertical normal modes in a semi-infinite stratosphere

The vertical normal modes, which are of fundamental importance when considering disturbances in the semi-infinite three-dimensional QG model, are derived by considering fluid flow governed by (3.30) when $q - f_0 = 0$. It shall be seen in chapter 6 that the vertical normal modes provide an orthogonal basis which can be used to perform integral transforms in the vertical coordinate z. When $q - f_0 = 0$, the QG potential vorticity equation (3.30) becomes

$$\nabla_H^2 \psi + \frac{1}{\rho(z)} \frac{\partial}{\partial z} \left[\frac{f_0^2}{N^2} \rho(z) \frac{\partial \psi}{\partial z} \right] = 0,$$

subject to the lower boundary condition (3.29) in the absence of forcing

$$\frac{\partial \psi}{\partial z} - \frac{\kappa}{H} \psi = 0 \quad \text{on} \quad z = 0.$$

Radial symmetry implies that ψ is independent of the azimuthal coordinate ϕ, so a separation of variables may be applied to the streamfunction $\psi(r, z) = \tilde{\psi}(r)\chi(z)$ giving

$$\chi(z) \nabla_H^2 \tilde{\psi}(r) + \tilde{\psi}(r) \frac{1}{\rho(z)} \frac{\partial}{\partial z} \left[\frac{f_0^2}{N^2} \rho(z) \frac{\partial \chi(z)}{\partial z} \right] = 0,$$

$$\frac{\partial \chi(z)}{\partial z} - \frac{\kappa}{H} \chi(z) = 0 \quad \text{on} \quad z = 0. \tag{3.33}$$

Using the separation of variables method, the first expression in (3.33) can be rearranged to give

$$\frac{1}{\rho(z)} \frac{\partial}{\partial z} \left[\frac{f_0^2}{N^2} \rho(z) \frac{\partial \chi(z)}{\partial z} \right] + \gamma^2 \chi(z) = 0, \tag{3.34}$$

$$\nabla_H^2 \tilde{\psi}(r) - \gamma^2 \tilde{\psi}(r) \quad = \quad 0,$$

where $\gamma \in \mathbb{R}$ is a real valued constant. Setting $\chi(z) = \rho(z)^{-1/2}\bar{\chi}(z)$ in equation (3.34) yields an ODE for $\bar{\chi}(z)$

$$\frac{d^2\bar{\chi}}{dz^2} + \left(\frac{N^2\gamma^2}{f_0^2} - \frac{1}{4H^2}\right)\bar{\chi} \quad = \quad 0, \tag{3.35}$$

subject to the condition on the lower boundary

$$\frac{d\bar{\chi}}{dz} + \frac{1}{2H}\left(1 - 2\kappa\right)\bar{\chi} \quad = \quad 0, \quad \text{on} \quad z = 0. \tag{3.36}$$

Solution of (3.35) subject to (3.36) gives a continuous spectrum of vertical wavenumbers $0 < m < \infty$ corresponding to a spectrum of orthonormal eigenfunctions

$$\chi(z;m) = \sqrt{\frac{2}{\pi}} e^{z/2H} \cos[mz + \epsilon(m)], \quad \text{where} \quad \epsilon(m) = \arctan\left[\frac{1-2\kappa}{2mH}\right], \tag{3.37}$$

with $\chi(z;m)$ satisfying the orthonormality condition

$$\int_0^\infty \chi(z;m)\chi(z;m')\rho(z)dz \quad = \quad \delta(m - m'). \tag{3.38}$$

For each wavenumber m, the eigenvalue associated with the eigenfunction $\chi(z;m)$ is

$$\gamma(m) = \frac{f_0}{N}\sqrt{m^2 + \frac{1}{4H^2}}. \tag{3.39}$$

In addition to the spectrum of wave-like eigenfunctions $\chi(z;m)$ satisfying (3.34) there is one other distinct solution of exponential form $\chi_0(z) = \hat{\chi}_0 \exp\{\alpha z\}$, where $\hat{\chi}_0$ is a constant yet to be determined. Substitution of $\chi_0(z)$ into the lower boundary condition (3.33) gives

$$\alpha \quad = \quad \frac{\kappa}{H},$$

which when substituted back into (3.34) gives the corresponding eigenvalue as

$$\gamma_0 = \frac{f_0}{NH}\sqrt{\kappa\,(\kappa-1)}. \tag{3.40}$$

If $\chi_0(z)$ belongs to the same orthonormal basis as $\chi(z;m)$ it must satisfy the orthonormality conditions

$$\int_0^\infty \chi_0(z)\chi(z;m)\rho(z)dz = 0, \tag{3.41}$$

$$\int_0^\infty \chi_0(z)\chi_0(z)\rho(z)dz = 1. \tag{3.42}$$

Straightforward integration using the forms derived for $\chi_0(z)$ and $\chi(z;m)$ shows that condition (3.41) is true for all m. Using the derived form of $\chi_0(z)$ to evaluate the integral in the second condition (3.42) sets the constant $\hat{\chi}_0$ necessary for orthonormality, giving the final eigenfunction of the orthonormal basis $\chi(z) = \{\chi_0(z), \chi(z;m)\}$ as

$$\chi_0(z) = \sqrt{\frac{1-2\kappa}{H}}e^{\frac{\kappa z}{H}}. \tag{3.43}$$

When referring to the vertical normal modes, the $\chi_0(z)$ mode shall be referred to as the *zero-order Lamb mode* (or *external* mode), and the spectrum of $\chi(z;m)$ as the *baroclinic Lamb modes*. We note that when the ersatz boundary condition (3.31) is used to derive the vertical normal modes, the zero-order Lamb mode becomes "barotropic", that is independent of height z. In this instance, the eigenfunction $\chi_0(z)$ and its associated eigenvalue become

$$\chi_0 = \sqrt{\frac{1}{H}}, \tag{3.44}$$

$$\gamma_0 = 0, \tag{3.45}$$

by taking $\kappa = 0$ in the expressions for $\chi_0(z)$ and γ_0 when using the correct lower boundary condition. The effect of using the ersatz lower boundary condition (3.31) when calculating the baroclinic modes is also seen by taking $\kappa = 0$ in equations (3.37) and (3.39). In this case, it is seen that the wave-like structure of the $\chi(z;m)$ and the expressions for $\gamma(m)$ remain unchanged, with the only difference appearing in the phase shift of $\chi(z;m)$ which becomes

$$\epsilon(m) \quad = \quad \arctan\left[\frac{1}{2mH}\right]. \tag{3.46}$$

Vertical normal modes in a bounded stratosphere

Many QG models of the stratosphere (e.g. Dritschel and Saravanan 1994; Waugh and Dritschel 1999) bound the domain in the vertical by introducing an upper boundary to the model. This is a necessary measure when dealing with numerical models of the stratosphere, in which a model lid must be imposed. While this lid is somewhat unrealistic, by imposing the lid at a sufficiently large height, a good approximation to the unbounded atmosphere can still be achieved.

In the previous section it was shown that changing the lower boundary condition in the QG model necessarily changed the expressions for the vertical normal modes $\chi(z)$. This was crucial for the external mode $\chi_0(z)$, which was seen to lose any dependence on height should the ersatz condition (3.31), rather than the correct condition (3.29), be used at the lower boundary. It is therefore to be expected that introducing an upper boundary, with an associated upper boundary condition, will also have an impact on the vertical normal modes $\chi(z)$.

Following Dritschel and Saravanan (1994), a rigid upper boundary is introduced

in the QG model at $z = D$, with corresponding boundary condition

$$\frac{\partial \psi}{\partial z} - \frac{\kappa}{H}\psi \;\; = \;\; 0, \quad \text{on} \quad z = D. \tag{3.47}$$

The most important implication of the upper boundary is that the solution of equation (3.34) becomes that of a Sturm-Liouville problem. Solving the resulting Sturm-Liouville equation along with the boundary conditions for the bounded stratosphere

$$\frac{\partial \chi(z)}{\partial z} - \frac{\kappa}{H}\chi(z) \;\; = \;\; 0 \quad \text{on} \quad z = 0, D, \tag{3.48}$$

gives a countable yet infinite set of discrete eigenfunctions $\chi_n(z)$. Imposing an orthonormality condition for these discrete eigenfunctions in the bounded domain

$$\int_0^D \chi_n(z)\chi_{n'}(z)\rho(z)dz \;\; = \;\; \delta_{nn'}, \tag{3.49}$$

the orthonormal discrete eigenfunctions are

$$\chi_0(z) = \sqrt{\frac{2\kappa - 1}{H\left(e^{\frac{D(2\kappa-1)}{H}} - 1\right)}} e^{\frac{\kappa z}{H}}, \tag{3.50}$$

$$\chi_n(z) = \sqrt{\frac{2}{D}} \cos\left[\frac{n\pi z}{D} + \epsilon_n\right], \qquad \epsilon_n = \arctan\left[\frac{D(1-2\kappa)}{2n\pi H}\right], \tag{3.51}$$

where $n \in \mathbb{N}/0$ for the baroclinic modes $\chi_n(z)$ in equation (3.51), and $\delta_{nn'}$ is the Kronecker delta. The eigenvalue associated with the discrete external mode (3.50) is the same as that given by (3.40) for the continuous case. The eigenvalue associated with each of the discrete baroclinic modes in (3.51) is retrieved by substituting $m = n\pi/D$ into equation (3.39) giving

$$\gamma_n \;\; = \;\; \frac{f_0}{N}\sqrt{\left(\frac{n\pi}{D}\right)^2 + \frac{1}{4H^2}}, \quad \text{with} \quad n \in \mathbb{N}/0. \tag{3.52}$$

The orthonormal basis $\{\chi_0(z), \chi_n(z)\}$, derived above when the correct boundary condition is used on the upper and lower boundaries, is seen to agree with the results given in Waugh and Dritschel (1999) for the ersatz boundary condition by taking $\kappa = 0$ in (3.50), (3.51), (3.40) and (3.52).

Green's functions

One of the most powerful features of the QG PV equation is that by inverting the operator for ψ in (3.27), the streamfunction, and therefore the velocity in the fluid, can be determined wholly by the PV distribution. Inverting the operator in (3.27) gives the streamfunction at point $\mathbf{x} = (x, y)$ as

$$\psi(\mathbf{x}, t) = \iiint \rho_0(z') G(\mathbf{x}; \mathbf{x}') q(\mathbf{x}', t) dx' dy' dz', \tag{3.53}$$

where $G(\mathbf{x}; \mathbf{x}')$ is the Green's function satisfying

$$\nabla_H^2 G + \frac{1}{\rho(z)} \frac{\partial}{\partial z} \left[\frac{f_0^2}{N^2} \rho(z) \frac{\partial G}{\partial z} \right] = \delta(\mathbf{x} - \mathbf{x}'), \tag{3.54}$$

$$\frac{\partial G}{\partial z} - \frac{\kappa}{H} G = 0 \quad \text{on} \quad z = 0. \tag{3.55}$$

The form of the Green's function depends on the vertical modes of the system, implying that a change in the form of the lower boundary condition will necessarily distort the Green's function (see Scott and Dritschel 2005). This in turn will have an effect on calculation of the streamfunction via (3.53).

3.2.3 Single layer quasi-geostrophic model of the stratosphere

Starting from the equations of motion and mass conservation which govern fluid flow in a three-dimensional atmosphere, as given by equations (3.1) and (3.2), we now discuss the intermediate member of our hierarchy of models, that of a single layer shallow water QG model of the stratosphere.

Assuming fluid in the single layer model is homogeneous, that is the density is the same throughout the fluid, by making the hydrostatic assumption, equations (3.1) and (3.2) become

$$\frac{D\mathbf{u}}{Dt} + f_0\mathbf{k}\times\mathbf{u} = -g\nabla_H h, \tag{3.56}$$

$$\frac{Dh_D}{Dt} + h_D(\nabla_H\cdot\mathbf{u}) = 0. \tag{3.57}$$

In equation (3.57) the depth of the fluid is $h_D(\mathbf{x},t) = h(\mathbf{x},t) - h_T(\mathbf{x})$, where $h_T(\mathbf{x})$ is topography at the lower boundary of the model, $h(\mathbf{x},t)$ is the height of the free surface of the fluid above $z = 0$, and $\mathbf{x}$ is a two-dimensional position vector in the plane. In equation (3.56), the Coriolis parameter $f = 2\Omega$ is taken to be constant over the domain, with $f = f_0$ calculated at the origin.

We now nondimensionalize equations (3.56) and (3.57) using $\mathbf{x} = L\tilde{\mathbf{x}}$, $\mathbf{u} = U\tilde{\mathbf{u}}$ and $t = (L/U)\tilde{t}$ (see Vallis 2006).

By expressing the height of the fluid surface in terms of perturbations around a constant reference depth H

$$h(x,t) = H + b\eta(x,t),$$

where b is a parameter of the system (see Fig. 3.1), it is therefore possible to nondi-

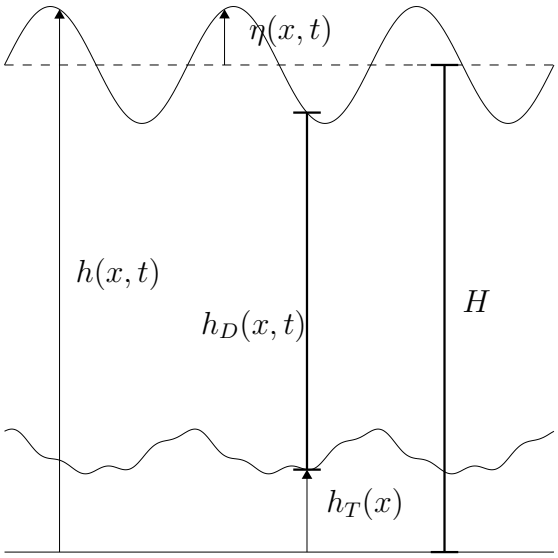

Figure 3.1: Schematic of the shallow water model, with H a reference depth, $h_T(x)$ the topographic forcing height, $h(x,t)$ the height of the fluid surface, $\eta(x,t)$ the height of the disturbance above H, and $h_D(x,t) = h(x,t) - h_T(x)$ the depth of the fluid.

mensionalize the fluid depth using $h_D(\mathbf{x},t) = H\tilde{h}_D(\tilde{\mathbf{x}},\tilde{t})$ where

$$\tilde{h}_D \;=\; 1 + b\tilde{\eta}(\tilde{x},\tilde{t}) - b\tilde{h}_T(\tilde{x}),$$

with $\eta = H\tilde{\eta}$ and $h_T = H\tilde{h}_T$. Here, the nondimensional disturbance amplitude $\tilde{\eta}$ and lower boundary topography $\tilde{h}_T$ are chosen to be of the same order magnitude, hence the factor of b in front of both terms. This yields the nondimensionalized shallow water equations

$$\epsilon\left(\frac{\partial}{\partial\tilde{t}} + \tilde{\mathbf{u}}\cdot\tilde{\nabla}_H\right)\tilde{\mathbf{u}} + \mathbf{k}\times\tilde{\mathbf{u}} \;=\; -\frac{gHb}{f_0UL}\tilde{\nabla}_H\tilde{\eta}, \quad (3.58)$$

$$\left(\frac{\partial}{\partial\tilde{t}} + \tilde{\mathbf{u}}\cdot\tilde{\nabla}_H\right)\left(b\tilde{\eta} - \tilde{h_T}\right) + \left(1 + b\tilde{\eta} - \tilde{h}_T\right)(\tilde{\nabla}_H\cdot\tilde{\mathbf{u}}) \;=\; 0, \quad (3.59)$$

where $\epsilon = U/f_0L = \mathrm{Ro}$ is the Rossby number. For convenience, all tildes are dropped from here on, with all variables being nondimensional unless stated otherwise. Taking small Rossby number $\epsilon \ll 1$, we now make the assumption that, at leading order in ϵ, the Coriolis terms in (3.58) are balanced by the terms on the right hand side of the equation giving $b = f_0UL/gH = \epsilon\mathcal{B}^2$, where $\mathcal{B} = f_0L/\sqrt{gH}$ is the inverse of the nondimensional Rossby radius of deformation.

This gives the geostrophically scaled shallow water equations as

$$\epsilon\left(\frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla_H\right)\mathbf{u} + \mathbf{k}\times\mathbf{u} \;=\; -\nabla_H\eta, \quad (3.60)$$

$$\epsilon\mathcal{B}^2\frac{D\eta}{Dt} - (\mathbf{u}\cdot\nabla_H)\,h_T + \left(1 + \epsilon\mathcal{B}^2\eta - \epsilon\mathcal{B}^2h_T\right)(\nabla_H\cdot\mathbf{u}) \;=\; 0, \quad (3.61)$$

We now make the quasi-geostrophic approximation and express the velocity as an

expansion in ϵ

$$\mathbf{u} = \sum_{i=0}^{\infty} \epsilon^i \mathbf{u}_i,$$

representing small ageostrophic perturbations from the geostrophic velocity $\mathbf{u}_0$. Similarly, the disturbance to the fluid surface and bottom topography are also expressed as expansions in ϵ

$$\eta = \sum_{i=0}^{\infty} \epsilon^i \eta_i,$$

$$h_T = \sum_{i=0}^{\infty} \epsilon^i h_{T_i}.$$

Substituting these expansions into equations (3.60) and (3.61), if the velocity components are given by $\mathbf{u} = (u, v)$ we get to $O(1)$

$$(u_0, v_0) = \left(-\frac{\partial \eta_0}{\partial y}, \frac{\partial \eta_0}{\partial x} \right),$$

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

At $O(\epsilon)$ we get

$$\frac{D_g u_0}{Dt} - v_1 = -\frac{\partial \eta_1}{\partial x}, \tag{3.62}$$

$$\frac{D_g v_0}{Dt} + u_1 = -\frac{\partial \eta_1}{\partial y}, \tag{3.63}$$

and

$$\mathcal{B}^2 \left(\frac{D_g \eta_0}{Dt} - \frac{D_g h_{T_0}}{Dt} \right) + \left(\frac{\partial u_1}{\partial x} + \frac{\partial v_1}{\partial y} \right) = 0, \tag{3.64}$$

where the material derivative following the geostrophic velocity is

$$\frac{D_g}{Dt} \equiv \frac{\partial}{\partial t} + u_0\frac{\partial}{\partial x} + v_0\frac{\partial}{\partial y}.$$

Taking $\partial/\partial x(3.63) - \partial/\partial y(3.62)$, advection of the relative vorticity ζ_0 where

$$\zeta_0 = \frac{\partial v_0}{\partial x} - \frac{\partial u_0}{\partial y} = \nabla_H^2\eta_0,$$

is given by

$$\frac{D_g\zeta_0}{Dt} = -\left(\frac{\partial u_1}{\partial x} + \frac{\partial v_1}{\partial y}\right). \tag{3.65}$$

Finally, substitution of (3.65) into (3.64) gives the conservation of QG potential vorticity

$$\frac{D_g}{Dt}\left(\nabla_H\eta_0 - \mathcal{B}^2\eta_0 + \mathcal{B}^2 h_{T_0}\right) = 0.$$

Redimensionalizing gives

$$\frac{D_g}{Dt}\left(q - f_0\right) = 0,$$

$$q(\mathbf{x}) = f_0 + \nabla_H^2\psi - L_R^{-2}\psi + \frac{f_0}{H}h_T, \tag{3.66}$$

where the dimensional Rossby radius of deformation is $L_R = \sqrt{gH}/f_0$, such that $\mathcal{B} = L/L_R$, and the dimensional streamfunction in terms of the surface disturbance is $\psi = g\eta/f_0$.

3.2.4 Barotropic vorticity equation

The simplest member of our hierarchy of models arises from taking the the QG shallow water formulation in the limit $L_R \to \infty$. Under these circumstances, the

PV equation governing fluid flow is the purely two-dimensional barotropic vorticity equation

$$q(\mathbf{x},t) \;=\; \nabla_H^2\psi + \frac{f_0}{H}h_T(\mathbf{x}). \tag{3.67}$$

In the absence of topographic forcing, this reduces to give the PV form of the Euler equations

$$q(\mathbf{x},t) \;=\; \nabla_H^2\psi. \tag{3.68}$$

3.3 Numerical models of polar vortex evolution in a compressible quasi-geostrophic atmosphere

In order to study the behaviour of the polar vortex during SSW events, it is useful to perform numerical experiments which simulate the vortex behaviour in our hierarchy of models.

For the single layer models described in sections 3.2.3 and 3.2.4, the QG Contour Dynamics with Contour Surgery (QGCS-2D) algorithm is used (see Dritschel 1988). When simulating flow in the QG cylindrical f-plane model described in section 3.2.2 there are two algorithms used in existing studies which will be discussed in this section. The first of these is a three-dimensional QG Contour Surgery model (QGCS) used by Dritschel and Saravanan (1994), Waugh and Dritschel (1999) and Wang and Fyfe (2000). The second is the Contour Advective Semi-Lagrangian (CASL) algorithm of Dritschel and Ambaum (1997) (given in Macaskill et al. (2003) for the cylindrical domain) used by Dritschel et al. (1999), Scott et al. (2004) and Esler and Scott (2005).

3.3.1 Quasi-Geostrophic Contour Dynamics with Surgery (QGCS): 2-D

The QG contour dynamics with surgery (QGCS) algorithm is a member of the contour dynamics family of numerical models (see Dritschel 1988, 1997). The prominent feature of these models is that the evolution of the flow is determined by following the motion of piecewise constant contours of potential vorticity.

Consider a vortex patch in a QG f-plane model in which the QG PV of the patch and the streamfunction are related by (3.66). In this system, the Green's function satisfies

$$\nabla_H^2 G - L_R^{-2} G \;=\; \delta(\mathbf{x}-\mathbf{x}'),$$

which has solution

$$G(\mathbf{x};\mathbf{x}') \;=\; -\frac{1}{2\pi} K_0\left(\frac{|\mathbf{x}-\mathbf{x}'|}{L_R}\right), \tag{3.69}$$

where K_0 is the modified Bessel function of the second kind. The streamfunction for the flow is then given by

$$\psi(\mathbf{x},t) \;=\; \iint q(\mathbf{x}',t) G(\mathbf{x};\mathbf{x}') dx' dy', \tag{3.70}$$

where q is the PV distribution, which consists of a finite number of vortex patches. Using Green's theorem in the plane, the expression for ψ yields the horizontal velocity $\mathbf{u}$

$$\mathbf{u} \;=\; -\frac{1}{2\pi} \sum_{k=1}^{N} \oint_{C^k} \Delta^k K_0\left(\frac{|\mathbf{x}-\mathbf{x}'|}{L_R}\right)(dx', dy'), \tag{3.71}$$

where C^k is the boundary of the kth patch, N is the total number of patches, and Δ^k is the vorticity jump at the boundary C^k. Therefore, by representing each patch of

vorticity by a series of nodes around its boundary, where the density of the nodes is given by a parameter μ, the velocity at each node can be calculated relatively easily by performing the contour integrals in (3.71) numerically.

To complete the dynamical advection in the system, the velocity field $\mathbf{u}(\mathbf{x}, t) = d\mathbf{x}/dt$ is calculated at each PV node location and a 4th order Runge-Kutta integration performed for small time-step dt to calculate the new location of each node.

The reduction of the area integral in (3.70) to an integral around vortex boundaries on which there is a discontinuity in PV is one of the most attractive features of the QGCS algorithm. That is, as the dynamics can be determined by the location and strength of PV boundary nodes alone, there is no need perform numerically expensive grid discretizations in the horizontal domain.

Another feature of the contour dynamics integration method is that if any two adjacent nodes happen to move too far apart during the advection of a bounding contour, extra nodes are inserted onto the contour to maintain the required spacing between nodes resolution. In a similar way, should nodes become too clustered on a bounding contour, nodes can be removed, decreasing computational expense without any significant sacrifice in accuracy. This is in contrast to algorithms using a discretized grid method demanding that the number of grid points, and hence resolution of the model, remain constant with time. In this way, the discretized grid method forces the model to use the optimum resolution for all calculations which greatly increases the expense of the calculation.

One other desirable feature of the QGCS algorithm is that the motion is unbounded in the horizontal. Therefore, no artificial horizontal boundaries must be

introduced into the model when simulating vortex evolution (for an example of such artificial boundaries see equations (3.83) and (3.84), and corresponding discussion in the text, for the CASL algorithm).

Finally, in addition to the contour advection algorithm, the model incorporates a contour surgery process which acts to remove filaments of PV on scales finer than a specified "cut-off" scale. This is performed in the interests of computational efficiency, as the advection of such fine PV filaments is computationally expensive while being seen to have relatively little impact on the overall dynamics of the system. This PV removal process, which violates the conservation of PV, may at first seem worrisome. However, the quantities of PV which are removed from the system in this way are usually small, such that no noticeable effect is usually observed in the dynamics of the flow.

3.3.2 Quasi-Geostrophic Contour Dynamics with Surgery (QGCS): 3-D

In the three-dimensional QGCS algorithm, a vertical coordinate is introduced such that the domain is bounded in the vertical by rigid boundaries at $z = 0$ and $z = D$. Similar to the two-dimensional QGCS algorithm, the three-dimensional model is unbounded in the horizontal.

The concepts of the model are essentially the same as for the two-dimensional QGCS model, and a brief overview of the algorithm follows: The model is discretized in the vertical into L layers of uniform thickness, each corresponding to a particular log-pressure height. On each of these layers, the polar vortex is represented by n_l

continuous contours of piecewise constant PV q_{l_k} (PV of the kth contour, $1 < k < n_l$) taken to be the PV of the vortex interior. Each contour is discretized by a series of nodes, each node carrying a PV of q_{l_k}. In this discretized form, the Green's function for the vertically bounded cylindrical domain is (see section 3.2.2)

$$G_{jl}(\mathbf{x};\mathbf{x}') \quad = \quad -\frac{1}{2\pi}\sum_{n=0}^{L-1}\chi_{jn}\chi_{ln}(z)K_0(\gamma_n r), \tag{3.72}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are planar coordinates and $\chi_{jn} = \chi_n(z_j)$. Substituting this discretized Green's function into equation (3.53) gives the streamfunction on layer j as

$$\psi_j(\mathbf{x}) \quad = \quad -\frac{1}{2\pi}\iint\sum_{l=1}^{L}\sum_{n=0}^{N_z}\rho_l\chi_{ln}\chi_{jn}K_0(\gamma_n|\mathbf{x}-\mathbf{x}'|)q_l(\mathbf{x}')dx'dy', \tag{3.73}$$

where N_z is the highest wavenumber permitted in the truncation of the infinite sum over vertical normal mode wavenumbers n. As the integral in (3.73) has now been reduced to an integral over the horizontal plane on each layer, standard contour dynamics techniques can now be applied.

We now specify that the PV distribution on each layer, q_l, comprises of n_l finite area vortex patches, each with a jump in PV of Δ_l^k at the vortex boundary. Then by differentiating (3.73) and noting that $\mathbf{u} = (-\partial_y\psi, \partial_x\psi)$, utilizing Green's Theorem in the plane to convert the area integral into an integral around the boundary of each vortex patch gives

$$\mathbf{u}_j(\mathbf{x}) \quad = \quad -\frac{1}{2\pi}\sum_{l=1}^{L}\sum_{k=1}^{n_l}\oint_{C_l^k}\sum_{n=0}^{N_z}\rho_l\chi_{ln}\chi_{jn}\Delta_l^k K_0(\gamma_n|\mathbf{x}-\mathbf{x}'|)\,(dx', dy')\,, \tag{3.74}$$

at any point $\mathbf{x}$ on layer $z = z_j$. In this expression, C_l^k is the boundary of the kth contour on layer l, across which the jump in PV is Δ_l^k. Advection of the nodes on the vortex boundaries, and the removal of fine scale vortex filaments, is then identical to that of the two-dimensional QGCS algorithm.

Upper and lower boundaries in the QGCS model act as flat, rigid surfaces. Therefore, topographic forcing at these boundaries cannot be represented by varying the elevation of the surface. To get around this issue, Dritschel and Saravanan (1994, section 2b) showed that boundary forcing in the model can be represented by placing a non-advectable sheet PV distribution in the lower and upper model layers, inducing a velocity field equivalent to physical forcing.

3.3.3 Contour Advective Semi-Lagrangian (CASL) algorithm

The Lagrangian treatment of fluid advection relies on following individual elements of the fluid, using the calculated velocity fields to determine the advection of each element at every time step of the model. The QGCS model described above applies the Lagrangian approach, whereby each node is followed throughout the motion with node locations being the only consideration when determining the overall dynamics. The CASL algorithm (Dritschel and Ambaum 1997), on the other hand, is a hybrid of the Lagrangian approach seen in the QGCS model and a grid discretization spectral method. At first glance, the CASL algorithm is much the same as the QGCS algorithm outlined above in that:

- The model is discretized into L layers of uniform thickness.
- On layer l, the kth region of vorticity is represented by a series of nodes with piecewise constant potential vorticity q_l^k. Each node carries a discontinuity in PV of Δ_l^k between the PV inside the vortex boundary and that outside the vortex boundary.

- Forcing at the lower and upper boundaries is represented by additional non-advectable sheet PV contributions in the lowermost and uppermost layers respectively.

- At each time step, the velocity field at each node is calculated and a 4th order Runge-Kutta integration is performed to calculate the updated node positions.

- Contour surgery is performed to remove fine scale filaments of PV.

However, in contrast to the QGCS algorithm the velocities at node locations are not calculated by performing an integral around each of the vortex bounding contours, as seen in (3.74). Instead, velocities are calculated by interpolating the node positions and their respective PV values onto a nonuniform cylindrical (r, ϕ, z) grid. Once this interpolation has been performed, finite difference spectral methods are used to calculate velocities at the grid points, which are then interpolated back to give the velocities at the node positions. This approach requires the horizontal domain to be bounded, in contrast to the unbounded horizontal domain of the QGCS.

A brief outline of the spectral aspects of the CASL algorithm and the implication of the finite horizontal domain are now given following Macaskill et al. (2003). In cylindrical polar coordinates (r, ϕ, z), subtracting f from both sides of (3.30) and multiplying by r gives

$$\frac{\partial}{\partial r}\left(r\frac{\partial \psi}{\partial r}\right) + \frac{1}{r}\frac{\partial^2 \psi}{\partial \phi^2} + \frac{r}{\rho(z)}\frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\rho(z)\frac{\partial \psi}{\partial z}\right) = \Big(q(r,\phi,z) - f\Big)r. \quad (3.75)$$

Taking discrete transforms of ψ and $(q-f)r$ with respect to z using the orthonormal basis $\{\chi_n\}$ gives

$$\psi(r,\phi,z,t) = \sum_{n=0}^{N_z} \psi_n(r,\phi,t)\chi_n(z), \quad (3.76)$$

$$(q-f)r = \sum_{n=0}^{N_z} Q_n(r,\phi,t)\chi_n(z), \tag{3.77}$$

which gives the transform of (3.75) as

$$\sum_{n=0}^{N_z}\left[r\frac{\partial\psi_n}{\partial r^2}+\frac{\partial\psi_n}{\partial r}+\frac{1}{r^2}\frac{\partial\psi_n}{\partial\phi^2}-r\frac{f_0^2}{N^2}\gamma_n^2\psi_n\right]\chi_n(z) = \sum_{n=0}^{N_z} Q_n\chi_n(z). \tag{3.78}$$

By imposing an azimuthal grid structure and defining the discrete Fourier transform of ψ_n and Q_n in the azimuthal direction to be

$$\psi_n(r,\phi,t) = \sum_{k=-N_\phi}^{N_\phi} \hat{\psi}_{n,k}(r,t)\exp[ik\phi], \tag{3.79}$$

$$Q_n(r,\phi,t) = \sum_{k=-N_\phi}^{N_\phi} \hat{Q}_{n,k}(r,t)\exp[ik\phi], \tag{3.80}$$

transformation of (3.78) in the azimuthal direction yields an ODE for $\hat{\psi}_{n,k}(r,t)$

$$r\frac{d^2\hat{\psi}_{n,k}}{dr^2}+\frac{d\hat{\psi}_{n,k}}{dr}-\left(\frac{k^2}{r}+r\frac{f^2}{N^2}\gamma_n^2\right)\hat{\psi}_{n,k} = \hat{Q}_{n,k}. \tag{3.81}$$

It is now necessary to discretize the domain in the r-coordinate by introducing a non-uniform radial grid. Macaskill et al. (2003) derived the discretized form of (3.81) given uniform grid spacing Δr in the r-direction. For the purposes of the current study, it desirable to utilize an arbitrary non-uniform grid spacing in r in order to retain the option of varying the grid resolution within the domain. In such a way, grid resolution can be increased in areas where PV is expected to be advected (in the polar regions, for example) while keeping lower grid resolution in regions in which PV is not expected to be advected, thereby optimizing the computational efficiency of the model.

The derivatives of $\hat{\psi}_{n,k}$ with respect to r at $r_{i+1/2}$, $r_{i-1/2}$ and r_i are taken to be

$$\left.\frac{\partial\hat{\psi}_{n,k}}{\partial r}\right|_{r_{i+\frac{1}{2}}} = \frac{\hat{\psi}^{(i+1)}_{n,k}-\hat{\psi}^{(i)}_{n,k}}{\Delta r_{i+1}},$$

$$\left.\frac{\partial\hat{\psi}_{n,k}}{\partial r}\right|_{r_{i-\frac{1}{2}}} = \frac{\hat{\psi}^{(i)}_{n,k}-\hat{\psi}^{(i-1)}_{n,k}}{\Delta r_{i}},$$

$$\left.\frac{\partial\hat{\psi}_{n,k}}{\partial r}\right|_{r_{i}} = \frac{\hat{\psi}^{(i+1)}_{n,k}-\hat{\psi}^{(i-1)}_{n,k}}{\Delta r_{i+1}+\Delta r_{i-1}},$$

$$\left.\frac{\partial^2\hat{\psi}_{n,k}}{\partial r^2}\right|_{r_{i}} = \frac{\left.\frac{\partial\hat{\psi}_{n,k}}{\partial r}\right|_{r_{i+\frac{1}{2}}}-\left.\frac{\partial\hat{\psi}_{n,k}}{\partial r}\right|_{r_{i-\frac{1}{2}}}}{r_{i+\frac{1}{2}}-r_{i-\frac{1}{2}}},$$

where $\Delta r_{i+1}=r_{i+1}-r_i$ and $r_{i+1/2}=r_i+(1/2)\Delta r_{i+1}$. Substitution of these expressions when differentiating $\hat{\psi}_{n,k}$ with respect to r in (3.81) gives

$$\frac{2}{\Delta r_{i+1}\Delta r_i\left(\Delta r_{i+1}+\Delta r_i\right)}\Bigg[r_{i+\frac{1}{2}}\Delta r_i\left(\hat{\psi}^{(i+1)}_{n,k}-\hat{\psi}^{(i)}_{n,k}\right)-$$

$$-r_{i-\frac{1}{2}}\Delta r_{i+1}\left(\hat{\psi}^{(i)}_{n,k}-\hat{\psi}^{(i-1)}_{n,k}\right)\Bigg]$$

$$-\left(\frac{k^2}{r_i}+r_i\frac{f^2}{N^2}\gamma_n^2\right)\hat{\psi}^{(i)}_{n,k} = Q^{(i)}_{n,k}. \qquad (3.82)$$

The tridiagonal system of equations in (3.82) can be solved for $\hat{\psi}_{n,k}$ at any grid location r_i given suitable boundary conditions imposed at $r_0=0$ and $r_{N_r}=r_D$. By introducing the outer wall at $r=r_D$ the domain of the CASL model is given by a cylinder with radius $r=r_D$ and height $z=D$. Following Macaskill et al. (2003), the boundary conditions in r are chosen to be

$$\hat{\psi}_{n,k} = 0 \quad k\neq 0, \qquad (3.83)$$

$$\hat{\psi}_{n,k}^{(0)} = 2\hat{\psi}_{n,k}^{(1)} - \hat{\psi}_{n,k}^{(2)}, \quad k = 0, \tag{3.84}$$

$$\hat{\psi}_{n,k}^{(N_r)} = 0. \tag{3.85}$$

The above enables the streamfunction to be calculated at any point on a cylindrical grid with N_r radial grid points, N_ϕ azimuthal grid points and N_z z grid points, given that the potential vorticity $q - f$ is defined at each point.

We summarize the CASL advection algorithm as follows:

- The location and potential vorticity $q - f$ of all bounding contour nodes is interpolated onto a nonuniform (r, ϕ, z) grid of dimension $N_r \times N_\phi \times N_z$.
- The velocity at each grid point is calculated using the spectral/finite difference technique above.
- The grid point velocities are interpolated back onto the nodes of each bounding contour.
- At each time step, the velocity at each node is used to perform a 4th order Runge-Kutta integration calculating the new node position.

The spectral/finite difference approach of the CASL algorithm above is considerably less computationally expensive than the contour integral method used by the QGCS algorithm. In particular, the fact that a relatively coarse cylindrical grid can be used to calculate the velocity at each node means many numerical experiments can be performed for a minimal sacrifice in accuracy. This is a considerable advantage

should many experimental runs need to be performed, which shall be seen to be the case in chapters 6 and 7.

However, inaccuracies arising from the relative coarseness of the spectral grid in the CASL algorithm may still render it unsuitable for certain applications. In particular, attempts to investigate the structure of three-dimensional stationary vortex structures have found the CASL algorithm to be unsuitable. Inaccuracies in velocity calculations on the vortex boundary, although small, were too large for stationary vortex solutions to be found without significantly increasing the grid-resolution, and hence the computational expense of the exercise.

Another concern when using the CASL model is the influence of the horizontal boundaries corresponding to the walls of the cylindrical domain. However, by taking the horizontal scale of the domain to be large in comparison to the polar vortex (typically $r_D \approx 10R$ where R is the typical vortex radius), the impact of these boundaries on the flow is seen to be negligible.

Chapter 4

Single Layer Shallow Water Model of Vortex-Splitting: the Kida Vortex Approximation

4.1 Introduction

Sudden stratospheric warmings (SSWs) of the winter polar stratosphere are accompanied by two distinct types of breakdown of the polar vortex. The first type are vortex-displacement SSWs which are characterized by displacement of the polar vortex off the pole, with an accompanying westward tilt of the vortex with height (for examples in the northern hemisphere see Manney et al. (1999)). Displacement SSWs are commonly referred to as wave-1 SSWs as the polar vortex is characterized by a growth of zonal wavenumber-1 waves on the vortex edge. The second type are vortex-splitting SSWs, during which the polar vortex splits into two smaller, almost

columnar vortices of comparable magnitude (for examples in the northern hemisphere see Andrews et al. (1985) and Manney et al. (1994); for the southern hemisphere vortex split in September 2002 see *J. Atmos. Sci.* Special Issue, vol **62**). Vortex-splitting SSWs are commonly referred to as "wave-2" SSWs as the breakdown of the vortex is associated with the growth of zonal wavenumber-2 planetary-scale Rossby waves on the vortex edge. From the perspective of vortex dynamics, it is desirable to treat these two types of SSWs as distinct, yet related, phenomena (Charlton and Polvani 2007; Matthewman et al. 2009).

One possible mechanism for the breakdown of the polar vortex is that of a "self-tuning" resonant excitation of disturbances to the vortex due to external forcing, for example by tropospheric planetary waves. Resonance is characterized by an excitation of free-travelling waves of the system by forcing. In a linear system, maximum excitation of these waves occurs at linear resonance, when the forcing frequency measured relative to the background flow is the same as the frequency of the linear mode being excited. Charney and DeVore (1979) investigated the resonant excitation of large amplitude waves by topographic forcing as a possible mechanism for planetary scale tropospheric disturbances, or "tropospheric blockings", in the atmosphere. Using an idealized β-plane model of the troposphere, it was found that the largest disturbance amplitudes were observed in flows close to linear resonance, with the scale and magnitude of such disturbances being similar to that observed during blocking events.

However, in a nonlinear system, nonlinear interactions can correct the frequency of the free mode being excited. In such systems, if forcing is initiated near linear reso-

nance it is possible for the system to become more strongly resonant, as amplitudes of excited waves increase and nonlinear frequency corrections become important. This phenomenon is known as "self-tuning" resonance (see Plumb 1981). If the self-tuning resonance mechanism is present, the maximum amplitude of excited waves in the system will arise when the forcing is initially not at linear resonance.

Plumb (1981) applied the concept of self-tuning resonance to an idealized β-plane model of the extratropical stratosphere. By investigating the interaction of free-travelling weakly nonlinear waves with a non-stationary topographic forcing, a relationship was derived which determined how far from linear resonance the system must be in order for nonlinear corrections to the free-travelling wave frequency to self-tune the system towards resonance.

This chapter will use a fully nonlinear two-dimensional vortex model to investigate the role of self-tuning resonance as the underlying mechanism of SSWs. We concentrate here on vortex-splitting SSWs, as the height-independent, barotropic behaviour of the polar vortex which characterizes these SSWs suggests that a suitable conceptual model is a barotropic single layer model of the stratosphere (see Liberato et al. 2007; Matthewman et al. 2009, as shown in chapter 2).

A hierarchy of models approach is used; the dynamics of elliptical vortices in a uniform strain flow in the two-dimensional Euler equations (Kida 1981) is used to gain insight into the dynamics of topographically forced vortices in a fully nonlinear shallow water model framework. Behaviour in the topographically forced vortex model is then investigated from the perspective of self-tuning resonance. In particular, the motivating questions for this and the following chapter are:

Q1. Is vortex-splitting behaviour, similar to that observed in vortex-splitting SSWs, observed in the topographically forced vortex in the shallow water model?

Q2. For a topographically forced circular vortex, does the excitation of waves on the vortex edge display features of a linear resonance of the vortex with the background flow? If so, from the perspective of self-tuning resonance, what is the role of nonlinearity in bringing the vortex into, or out of, resonance with the background flow?

Q3. Under what circumstances, and to what extent, can the behaviour of the vortex in the fully nonlinear shallow water model be understood by reference to the Kida vortex model? Can a mapping between the two models be established?

To answer these questions the chapter will proceed as follows: We start by introducing an elliptical vortex patch in the absence of background flow, the Kirchhoff ellipse, and discuss the conditions leading to instability of this vortex. Section 4.3 will provide a thorough review of the results of Kida (1981) for a vortex patch subject to uniform background strain and rotation velocity fields. A topographically forced circular vortex in a shallow water model is introduced in section 4.4. The relevance of the Kida vortex to the topographically forced model is then discussed and it is shown that the uniform strain in the Kida system acts as a good approximation to the topographic forcing in the shallow water system. Within section 4.4, a wave-activity diagnostic (Dritschel and Saravanan 1994) quantifying the disturbance to the vortex is introduced. In section 4.5 this wave-activity diagnostic is used to analyze vortex evolution in fully nonlinear numerical model experiments. The utility of the vortex-splitting diagnostic κ_4, introduced in chapter 2, in objectively classifying splitting of

the vortex is also demonstrated. Finally, section 4.6 provides discussion of the fully nonlinear numerical model results in relation to the key questions above.

4.2 The Kirchhoff ellipse and its instabilities

In purely two-dimensional vortex dynamics, the streamfunction ψ induced by a uniform vortex patch with interior vorticity Δ and area D in the absence of background flow is governed by

$$\nabla^2_{\mathbf{x}}\psi = \begin{cases} \Delta & \mathbf{x} \in D \\ 0 & \mathbf{x} \notin D. \end{cases} \tag{4.1}$$

Before proceeding, it is convenient to nondimensionalize all quantities using $\mathbf{x} = a_0\tilde{\mathbf{x}}$, $t = \tilde{t}/\Delta$ and $\psi = \Delta a_0^2\tilde{\psi}$, where a_0 is a typical vortex radius, reducing the problem to that of a vortex patch with unit interior vorticity

$$\nabla^2_{\tilde{\mathbf{x}}}\tilde{\psi} = \begin{cases} 1 & \tilde{\mathbf{x}} \in \tilde{D} \\ 0 & \tilde{\mathbf{x}} \notin \tilde{D}, \end{cases} \tag{4.2}$$

where the $\tilde{D}$ has area π by construction. For the remainder of the chapter all tildes will be dropped on nondimensional quantities with the proviso that, unless specified, all variables are in nondimensional form.

Consider first an elliptical vortex patch of unit interior vorticity, with major and minor semi-axes of length a and b, in the absence of background flow. The centre of the patch is located at the origin, with the major axis making an angle of $\phi_e(t)$ with the x-axis. Such a vortex configuration is known as the Kirchhoff vortex which

rotates with angular velocity

$$\dot{\phi}_e = \frac{r_e}{(r_e+1)^2}, \tag{4.3}$$

where $r_e = b/a \leq 1$ is the aspect ratio of the elliptical vortex (Saffman 1992).

The linear stability of the Kirchhoff ellipse to infinitesimal disturbances was studied by Love (1893), and recently revisited by Guo et al. (2004). A very brief overview of their analysis will follow. Transforming to a coordinate system (ξ, ν), where ξ acts as an "elliptic radial coordinate", and ν as an "elliptic azimuthal coordinate", the boundary of the undisturbed ellipse can be expressed as

$$\begin{aligned} x &= c\cosh\xi_0\cos\nu, \\ y &= c\sinh\xi_0\sin\nu, \end{aligned}$$

where $c = \sqrt{a^2 - b^2}$ and ξ_0 is a constant elliptic radial coordinate. In this coordinate system, the boundary of a perturbed ellipse can be expressed in complex form as

$$z(\xi,\nu) = c\cosh\left[\xi_0 + \xi(\nu,t) + i\nu\right],$$

where $\xi(\nu,t)$ is the disturbance to the ellipse. By rewriting this disturbance in the form

$$\eta(\nu,t) \equiv J_0(\nu)\xi(\nu,t),$$

where $J_0(\nu) = c^2(\cosh 2\xi - \cos 2\nu)/2$, it is possible to express $\eta(\nu,t)$ as a Fourier series in the ν coordinate

$$\eta(\nu,0) = \sum_{m=1}^{\infty}\Big[A_m\cos m\nu + B_m\sin m\nu\Big],$$

where m acts as an "elliptic azimuthal wavenumber". The following linear stability analysis finds that unstable growth of perturbations occurs whenever L is positive, where

$$L = \left(\frac{r_e - 1}{r_e + 1}\right)^{2m} - \left(\frac{2mr_e}{(r_e+1)^2} - 1\right)^2 .$$

Azimuthal wavenumber $m = 1$ gives $L < 0$, implying that the elliptical vortex is linearly stable for all aspect ratios, and azimuthal wavenumber $m = 2$ gives $L = 0$, implying neutral linear stability. When $m \geq 3$, aspect ratios $1/3 < r_e \leq 1$ give $L < 0$, corresponding to linearly stable elliptical vortices, whereas aspect ratios $r_e < 1/3$ give $L > 0$, corresponding to linearly unstable elliptical vortices. This result was first presented by Love (1893). The analysis of Love (1893) was modified to include nonlinear stability by Tang (1987), who found that the Kirchhoff ellipse is in fact nonlinearly stable whenever the aspect ratio satisfies $1/3 < r_e < 1$.

Using a fully nonlinear numerical model, Dritschel (1986) showed that for aspect ratios $r_e \lesssim 1/6$, much smaller than that of the stability threshold $r_e = 1/3$, nonlinear instability of elliptic azimuthal wavenumber-2 disturbances on the vortex edge leads to splitting of the vortex. For aspect ratios $1/6 \lesssim r_e \lesssim 1/5$, while elliptic azimuthal wavenumber-2 disturbances are stable, wavenumber-4 disturbances are unstable, and are characterized by erosion of vorticity from the end of the ellipse major axis via filamentation. For $1/5 \lesssim r_e < 1/3$, it was shown that wavenumber-3 disturbances are unstable (see table 3 and Fig. 12 in Dritschel 1986).

4.3 The Kida vortex

In order to understand the fully nonlinear shallow water dynamics of a circular vortex patch subject to topographic forcing, it will prove helpful to first investigate the behaviour of a class of elliptical vortices subject to uniform strain and rotation velocity fields. Kida (1981) derived a class of exact analytic solutions for the evolution of such elliptical vortices which will be referred to as the "Kida vortex" hereafter.

In real flows, any vortex is subject to the influence of other vortices, often much larger than itself. To investigate the influence of remote vortices on the Kirchhoff ellipse, Kida (1981) introduced an external background flow $\mathbf{u}_E = (u_E, v_E)$ to the Kirchhoff vortex system where

$$\begin{aligned} u_E &= \Lambda x - \Omega_b y, \\ v_E &= -\Lambda y + \Omega_b x. \end{aligned}$$

Note that $\nabla \cdot \mathbf{u}_E = 0$, i.e. the external background flow is nondivergent. In this formulation, the $\Lambda(> 0)$ terms correspond to an external straining field and the Ω_b terms correspond to a solid body rotation of vorticity $\mathbf{k} \cdot \nabla \times \mathbf{u}_E = \xi_b = 2\Omega_b$. The streamfunction satisfying $\mathbf{u}_E = -\nabla \times \psi_E \mathbf{k}$ is then given as

$$\psi_E = -\Lambda xy + \frac{\Omega_b}{2}\left(x^2 + y^2\right), \tag{4.4}$$

and including this external velocity contribution the total vorticity in the system is

$$q = \begin{cases} \xi_b + 1 & \tilde{\mathbf{x}} \in D \\ \xi_b & \tilde{\mathbf{x}} \notin D. \end{cases} \tag{4.5}$$

To analyze the evolution of an elliptical vortex patch in this external velocity field, Kida (1981) derived a system of equations for $\dot{r}_e$ and $\dot{\phi}$ which describe the

vortex behaviour at all times. Here, these equations are derived using the method of Schwarz functions as seen in Saffman (1992), which is an elegant alternative to the direct approach used in Kida (1981).

Following Saffman (1992), we now define the Schwarz function in terms of the elliptical vortex patch. Davis (1974) defines the Schwarz function of a closed curve ∂D bounding a simply connected region D as the unique, locally analytic function $S(z)$ that is equal to $\bar{z}$ everywhere on the curve ∂D. Taking the interior of the elliptical vortex patch to be the simply connected region D with bounding contour ∂D, the Schwarz function $S(z)$ may be written as

$$S(z) \;=\; F(z) + G(z), \tag{4.6}$$

where $F(z)$ and $G(z)$ are both functions of the complex variable z, with $F(z)$ defined as being analytic inside ∂D, and $G(z)$ as being analytic outside ∂D. The complex velocity field in the domain may then be expressed as

$$u - iv = \begin{cases} \frac{1}{2}iF(z) - \frac{1}{2}i\bar{z} & z \in D \\ -\frac{1}{2}iG(z) & z \notin D, \end{cases} \tag{4.7}$$

where the boundary of the ellipse acts as a material surface in the flow.

It is now useful to consider the rational function conformal map

$$z(\zeta) \;=\; \alpha(t)\zeta + \frac{\beta(t)}{\zeta}, \tag{4.8}$$

mapping the interior of the vortex patch in the z-plane to the interior of the unit disc in the complex ζ-plane. Making ζ the subject of (4.8) inverts this map giving

$$\zeta \;=\; \frac{z \pm (z^2 - 4\alpha\beta)^{\frac{1}{2}}}{2\alpha}.$$

Without loss of generality, the point $\zeta = 1$ is mapped to the end of the major axis of the elliptical patch and $\zeta = i$ is mapped to the end of the minor axis. This allows α and β to be expressed in terms of the parameters of the ellipse where

$$\alpha = \frac{1}{2}(a+b)\,e^{i\phi_e}, \tag{4.9}$$

$$\beta = \frac{1}{2}(a-b)\,e^{i\phi_e}. \tag{4.10}$$

The Schwarz function for the ellipse can now be constructed explicitly giving

$$\begin{aligned} S(z) &= \frac{\bar{\alpha}}{\zeta} + \bar{\beta}\zeta, \\ &= \frac{ab}{\alpha\zeta} + \frac{\bar{\beta}z}{\alpha}, \end{aligned}$$

where $F(z) = \bar{\beta}z/\alpha$ and $G(z) = ab/\alpha\zeta(z)$ from (4.6). In conjunction with (4.7) this gives the velocity induced by the ellipse in the region $z \notin D$ as

$$u - iv = -\frac{iab}{2\alpha\zeta}.$$

To progress, we exploit the fact that the boundary of the ellipse acts as a material surface in the flow. As the boundary of the ellipse corresponds to the curve $|\zeta| = 1$ in the ζ-plane, the ellipse boundary at $z = Z$ may be parametrised using $\zeta = e^{is}$ to give

$$Z(s,t) = \alpha e^{is} + \beta e^{-is} \quad \text{where} \quad 0 \le s < 2\pi.$$

As the ellipse boundary is a streamline of the flow, the difference between $\partial Z/\partial t$ and the velocity outside the ellipse is parallel to $\partial Z/\partial s$, that is

$$\frac{\partial Z}{\partial s} \parallel \frac{\partial Z}{\partial t} - (u + iv). \tag{4.11}$$

In complex form the external straining field is expressed as $u_E + iv_E = \Lambda\bar{z} + i\Omega_b z$, giving the total velocity field on the ellipse boundary as

$$u + iv = \frac{iab\zeta}{2\bar{\alpha}} + \left(\Lambda\bar{\beta} + i\Omega_b\alpha\right)\zeta + \frac{(\Lambda\bar{\alpha} + i\Omega_b\beta)}{\zeta},$$

which when substituted into (4.11) yields

$$i\alpha\zeta - \frac{i\beta}{\zeta} \quad \| \quad \left(\dot{\alpha} - \frac{iab}{2\bar{\alpha}} - \Lambda\bar{\beta} - i\Omega_b\alpha\right)\zeta - \frac{\left(\dot{\beta} - \Lambda\bar{\alpha} - i\Omega_b\beta\right)}{\zeta}. \tag{4.12}$$

The form of (4.12) has been chosen to emphasise that ζ and $1/\zeta$ are orthogonal coordinates. To satisfy (4.12), the ratio of coefficients of ζ and $1/\zeta$ on each side must be equal, which by using expressions (4.9) and (4.10) to calculate $\dot{\alpha}$ and $\dot{\beta}$ yields

$$\frac{1}{2}\left[(a\dot{a} - b\dot{b})\cos 2\phi_e - \Lambda(a^2 + b^2) + \left(\Omega_b(a^2 - b^2) + \frac{ab(a-b)}{a+b} - (a^2 - b^2)\dot{\phi_e}\right)\sin 2\phi_e\right]$$

$$+ \frac{i}{2}\left[(a\dot{a} - b\dot{b})\sin 2\phi_e + \left((a^2 - b^2)\dot{\phi_e} - \frac{qab(a-b)}{a+b} - \Omega_b(a^2 - b^2)\right)\cos 2\phi_e\right] = 0.$$

By equating real and imaginary parts and enforcing conservation of the patch area $A_e = \pi ab$, this reduces to give two equations describing the evolution of the ellipse parameters $r_e = a/b$ and ϕ_e in the Kida system, that is

$$\dot{r_e} = 2\Lambda r_e \cos 2\phi_e, \tag{4.13}$$

$$\dot{\phi_e} = -\frac{\Lambda(r_e^2 + 1)}{r_e^2 - 1}\sin 2\phi_e + \frac{r_e}{(r_e + 1)^2} + \Omega_b. \tag{4.14}$$

By constructing the function $\partial\phi_e/\partial r_e$ from (4.13) and (4.14) and integrating with respect to r_e, the relationship

$$\sin 2\phi_e = \frac{1}{\Lambda}g(r_e, s, \Omega_b), \tag{4.15}$$

$$g(r_e, s, \Omega_b) = \frac{r_e}{r_e^2 - 1} \ln \left[\frac{(r_e + 1)^2}{4 s r_e} \right] + \Omega_b \frac{(r_e - 1)}{(r_e + 1)}, \tag{4.16}$$

is derived, where $s > 0$ is a constant of integration. The rotation of the vortex can also be written in terms of $g(r_e, s, \Omega_b)$ as

$$\dot{\phi_e} = r_e \frac{\partial g}{\partial r_e}. \tag{4.17}$$

4.3.1 Regimes of behaviour in the Kida vortex model

Equations (4.13), (4.14) and (4.16) are now used to investigate the evolution of an elliptical vortex patch in detail. It can be seen that $g(r_e, s, \Omega_b)$ satisfies the properties

$$g(0, s, \Omega_b) = -\Omega_b,$$

$$\lim_{r_e \to 1} [g(r_e, s, \Omega_b)] = \begin{cases} -\infty & s < 1 \\ 0 & s = 1 \\ \infty & s > 1. \end{cases} \tag{4.18}$$

Given any particular pair of the parameters Λ and Ω_b, the aspect ratio r_0 is now introduced as the maximum value of r_e achieved over the vortex evolution. The behaviour of (4.16), and its consequence for the evolution of the vortex patch, is discussed in Kida (1981) and Dritschel (1990) for general r_0 values. For any choice of s, from (4.15) it is clear that $g(r_e, s, \Omega_b)$ must satisfy

$$|g(r_e, s, \Omega_b| \leq \Lambda. \tag{4.19}$$

Therefore, for all Λ and Ω_b the resulting evolution of the vortex can be described by the portion of $g(r_e, s, \Omega_b)$ enclosed within this region.

Of particular interest here is the special case of Kida vortex with $r_0 = 1$, that is a vortex which can be thought of as being initially circular. To gain an understanding of the types of vortex evolution we investigate the function $g(r_e, s, \Omega_b)$ in more detail for this choice of r_0. From (4.18) it is seen that for $g(r_e, s, \Omega_b)$ to satisfy (4.19) when $r_e = r_0 = 1$, we must have $s = 1$.

Evolution of the Kida vortex

When $r_0 = 1$ and the vortex is initially circular, it is helpful to define the function $g_1(r_e, \Omega_b) = g(r_e, 1, \Omega_b)$. Fig. 4.1 shows $g_1(r_e, \Omega_b)$ for a selection of the parameters (Ω_b, Λ). As the vortex evolves to more elliptical shapes, the value of r_e decreases, corresponding to movement to the left along the graph of $g_1(r_e, \Omega_b)$. However, it is known from the relationship in (4.19) that $|g_1(r_e, \Omega_b)| < \Lambda$, meaning that movement along the graph must halt once $g_1(r_e, \Omega_b) = \pm\Lambda$. The value of r_e at this point is the minimum aspect ratio and $\dot{r}_e = 0$. As the vortex continues to evolve, movement along the graph of $g_1(r_e, \Omega_b)$ reverses, with movement now to the right until $r_e = 1$, at which point the process repeats. At any point during the evolution, the magnitude and direction in which the vortex is rotating $\dot{\phi}$ is given by the gradient of the graph for $g_1(r_e, \Omega_b)$ at that point.

Panel A of Fig. 4.1 shows a graph of $g(r_e, \Omega_b)$ which is representative of purely anticlockwise rotating vortex states. In this regime, $\partial g_1/\partial r_e > 0$ at all points on the portion of the graph to which motion is restricted (the heavy dark line); that is, the direction of rotation is purely anticlockwise.

Panel B shows a graph which is representative of parameters (Ω_b, Λ) lying on the

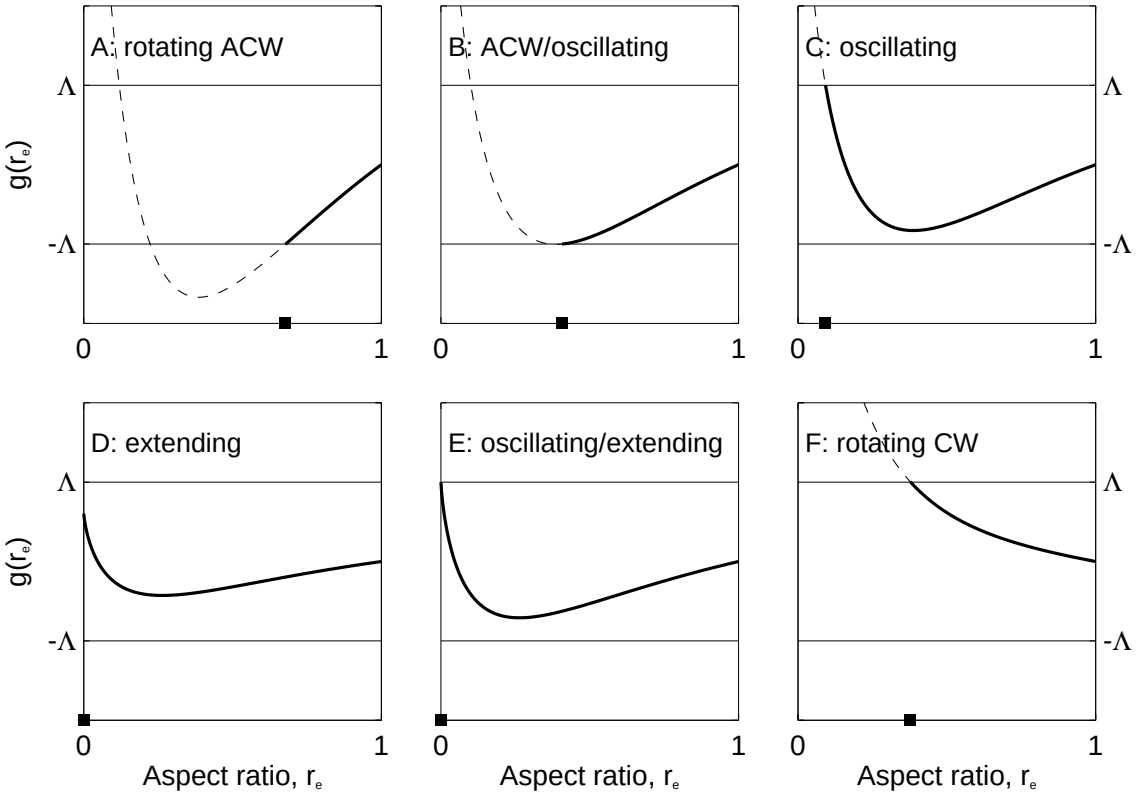

Figure 4.1: Representative $g_1(r_e, \Omega_b)$ plots of the four regimes of the Kida vortex when $r_0 = 1$, with $r_e = a/b$ the aspect ratio, Ω_b the background solid body rotation and Λ the external straining field. A: $(\Omega_b, \Lambda) = (-0.168, 0.015)$ representative of anticlockwise rotating solutions (ACW). B: $(\Omega_b, \Lambda) = (-0.168, 0.025)$ representative of solutions lying on the ACW/oscillating regime boundary. C: $(\Omega_b, \Lambda) = (-0.168, 0.030)$ representative of oscillating solutions. D: $(\Omega_b, \Lambda) = (-0.091, 0.151)$ representative of extending solutions. E: $(\Omega_b, \Lambda) = (-0.091, 0.091)$ representative of solutions lying on the oscillating/extending regime boundary. F: $(\Omega_b, \Lambda) = (-0.300, 0.035)$ representative of clockwise rotating solutions (CW). In all plots, the portion of the graph on which motion is restricted is denoted by the heavy solid line, and the minimum aspect ratio by the dark box.

regime boundary between purely anticlockwise rotating and oscillating vortex states, with the graph just touching the line at $-\Lambda$. As function $g_1(r_e, \Omega_b)$ has a stationary point when touching the line at $-\Lambda$, the vortex ceases to rotate as $\dot{\phi}_e = \partial g_1/\partial r_e = 0$.

For a small increase in Λ as shown in panel C, the graph no longer touches the line at $-\Lambda$. For this set of graphs, it is seen that $\partial g_1/\partial r_e$ takes both positive and negative values on the portion of the graph to which motion is restricted. This represents oscillating vortex states. From the location of the dark boxes in panels B and C, it is seen that for an infinitesimal increase in Λ from that shown in panel B, corresponding to a crossing of the regime boundary, the minimum aspect ratio will instantaneously jump to much smaller values. This instantaneous jump in the minimum aspect ratio of the vortex is discussed in the next section. Also, when considering the rotating behaviour of the vortex for the limiting case in panel B between anticlockwise rotating and oscillating vortex states, it must be the case that for parameters lying *exactly* on the regime boundary between the two, the vortex will stop rotating but will take an infinite amount of time to do so.

Panel D is representative of graphs which correspond to (Ω_b, Λ) lying in the extending Kida regime. As $g_1(r_e, \Omega_b)$ never intersects with the lines at $\pm\Lambda$ there is no reversal in the movement along the graph as seen in panels A-C. Instead, as time increases the minimum aspect ratio continues to decrease ($\dot{r} < 0$), but with $r_e = 0$ not being reached in finite time.

The regime boundary between extending and oscillating solutions is shown in panel E when the graph intersects the line for Λ when $r_e = 0$ exactly. That is, the elliptical vortex will eventually achieve $\dot{r}_e = 0$ when $r_e = 0$, but will take an infinite

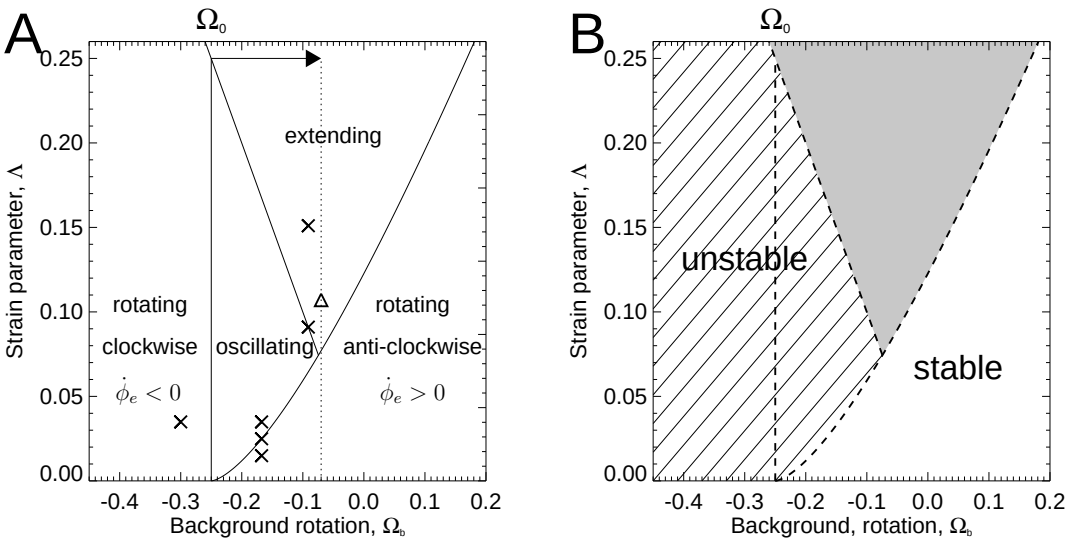

Figure 4.2: Panel A: the four regimes of the Kida system with $r_0 = 1$ in (Ω_b, Λ) parameter space. Crosses correspond to each of the parameter pairs (Ω_b, Λ) shown in Fig. 4.1. The heavy dashed line denotes the location of the background rotation $\Omega_0 = -0.25$ calculated from (4.42) with $k = 2$ and $\mathcal{B} = 0$. The arrow denotes the shift between Ω_0 and the tuning parameter $\Omega_b = -0.07$ at which the peak vortex response occurs in the fully nonlinear model experiments of section 4.5, as measured by the wave-activity $\mathcal{W}(t)$. The exact location $(\Omega_b, \Lambda) = (-0.07, 0.11)$ of the peak response is shown by the triangle. Panel B: the linear stability of the Kida vortex in (Ω_b, Λ) parameter space. No linear stability analysis exists in the extending Kida vortex regime which is shaded.

amount of time to do so. The final panel F shows a graph of $g_1(r_e, \Omega_b)$ representative of (Ω_b, Λ) lying within the clockwise rotation regime. The derivative $\partial g_1/\partial r_e < 0$ for the entirety of the motion, before the minimum aspect ratio is attained when the graph intersects with the line at Λ.

Regime boundaries in parameter space

We now determine the location of each of the four regimes of anticlockwise rotation, clockwise rotation, oscillation and extension, in (Ω_b, Λ) parameter space.

Panel B of Fig. 4.1 gives an example of the function $g_1(r_e, \Omega_b)$ for parameters $(\Omega_b, \Lambda) = (-0.168, 0.025)$ lying on the regime boundary of the anticlockwise and oscillating regimes. This graph is characterized by its intersection of both lines corresponding to $\pm\Lambda$. It is also seen that the function just touches the lower of these lines $(-\Lambda)$ and intersects the upper line (Λ) when $r_e > 0$. Therefore given Ω_b, the corresponding Λ value of the anticlockwise and oscillating regime boundary is found by solving

$$\frac{\partial g_1}{\partial r_e} = 0, \qquad \text{when} \qquad g_1(r_e) = -\Lambda. \tag{4.20}$$

As $g_1(0, \Omega_b) = -\Omega_b$, it is straightforward to see that the process above can also be used to identify the boundary of the anticlockwise rotation and extension regimes by noticing that if

$$-\Omega_b = \quad g_1(0; \Omega_b) \quad < \Lambda, \tag{4.21}$$

is satisfied, then the boundary lies with the extending regime, and if it is not satisfied then the boundary lies with the oscillating regime.

The function $g_1(r_e, \Omega_b)$ for parameters $(\Omega_b, \Lambda) = (-0.091, 0.091)$ lying on the boundary between the oscillating and extending regimes is shown in panel E of Fig. 4.1. This class of functions is characterized by an intersection of $g_1(r_e, \Omega_b)$ with the line for $+\Lambda$, rather than $-\Lambda$, with this intersection occurring when $r_e = 0$. Therefore, for a parameter pair to lie on this regime boundary the necessary conditions are $|g_1(r_e, \Omega_b)| < \Lambda$ for $0 < r_e \leq 1$ and that

$$-\Omega_b = \quad g_1(0; 1, \Omega_b) \quad = \Lambda, \tag{4.22}$$

is satisfied. The point which serves as the boundary of the anticlockwise, oscillating, and extending regimes, must therefore satisfy both (4.20) and (4.22). When $\partial g_1/\partial r_e < 0$ at all times on the graph of $g_1(r_e)$, it is seen that oscillating solutions make way to purely clockwise rotating states of the vortex. This regime boundary is given by the value of Ω_b such that $\partial g_1(r_e)/\partial r_e = 0$ when $r_e = 1$. This value of Ω_b is then the location of the oscillating and clockwise rotation regime boundary for all $\Lambda < -\Omega_b$.

The regimes and their respective boundaries are shown explicitly in panel A of Fig. 4.2, with each set of the parameters (Ω_b, Λ) corresponding to the $g_1(r_e, \Omega_b)$ graphs discussed above denoted by crosses. This regime diagram is similar to that given in Kida (1981), with the choice of $r_0 = 1$ and $s = 1$ allowing explicit regime boundaries to be defined.

Although it was predicted by Kida (1981) that oscillating vortex states can exist within the oscillating and clockwise rotating regimes seen in Fig. 4.2, for the special case $r_0 = 1$ at least, the existence of a vertical regime boundary separating oscillating and clockwise rotating vortex states, and the existence of a regime in which *only* oscillating vortex states are observed, is a new development in this work. As a result, the regime diagram shown in Fig. 4.2 differs slightly from that seen in Dritschel (1990), who also considered the special case $r_0 = 1$, by the inclusion of an oscillating sub-regime lying within the clockwise rotation regime of Dritschel (1990).

4.3.2 Stability of the Kida vortex

A linear stability analysis of the periodic Kida regimes, specifically the clockwise and anticlockwise rotation and oscillating regimes when $r_0 = 1$, was performed in Dritschel (1990) using Floquet theory. The monotonic, aperiodic extension of the vortex in the extending regime, and the reliance of Floquet theory on periodic vortex evolution, meant that a linear stability analysis does not make sense for flows in the extending regime.

Panel B of Fig. 4.2 gives an overview of the linearly stable and unstable regions of parameter space, as found in Dritschel (1990). For the periodic Kida regimes, it was determined that the anticlockwise rotation ($\dot{\phi}_e > 0$) regime is linearly stable, whereas the clockwise rotation regime (collectively the clockwise rotation and oscillating regimes in panel A) was found to be linearly unstable. Although, as previously stated, this analysis draws no conclusions regarding the linear stability of the extending regime, it is perhaps to be expected that the regime boundary of the anticlockwise rotation regime may form a barrier between linearly stable and unstable states of the vortex, with stable solutions lying in the anticlockwise rotation regime and unstable solutions lying in the remaining three regimes in panel A of Fig. 4.2. This final prediction, although not formally derived for the extending vortex case, may provide insight into the stability of vortices in a shallow water system as seen in section 4.5.

4.3.3 The potential function for the Kida vortex

Another description of the behaviour of the Kida vortex can be derived by expressing the Kida theory in the form of a fully nonlinear oscillator. Using the expression

for $\dot{r}_e$ as given in equation (4.13), it is straightforward to show that

$$\left(\frac{dr_e}{dt}\right)^2 = 4\Lambda^2 r_e^2 \left(1 - \sin^2 2\phi_e\right),$$

which by demanding that the vortex be initially circular gives

$$\left(\frac{dr_e}{dt}\right)^2 + V_e(r_e; \Omega_b, \Lambda) = 0, \tag{4.23}$$

where $V_e(r_e, \Omega_b, \Lambda)$ acts as a potential and is defined as

$$V_e(r_e; \Omega_b, \Lambda) = 4r_e^2 \left(\left[\frac{r_e}{r_e^2 - 1} \ln\left(\frac{(r_e+1)^2}{4r_e}\right) + \Omega_b \frac{r_e - 1}{r_e + 1} \right]^2 - \Lambda^2 \right).$$

It is seen that in order for dr_e/dt to take real values, the potential V_e must be negative. When $r_e = 1$ and the vortex is circular, the value of the potential is $V_e = -4\Lambda^2$.

Fig. 4.3 shows the maximum disturbance amplitude $A = (1 - r_e)/2r_e$ attained during the Kida vortex evolution. The reason $A > 0$ is chosen as a diagnostic, rather than $r_e < 1$, is to enable direct comparison with disturbance amplitudes due to wave-like perturbations to a circular vortex which will be the focus of section 4.4.

The region of parameter space to the right of the figure is characterized by small maximum disturbance amplitudes, with the contours only becoming tightly clustered in the vicinity of the curved regime boundary. In contrast, maximum disturbance amplitudes in the region of parameter space on the left of the curved regime boundary are noticeably higher, with maximum amplitudes $A_{max} > 2$ occurring for relatively small values of the strain rate Λ. Furthermore, contours of A_{max} are much tighter in this region than in the region to the right of the curved regime boundary. The regime boundary between the two regions is marked by a discontinuity in A_{max}, as also seen in the jump in $\min[r_e]$ between panels B and C in 4.1. Interestingly, the

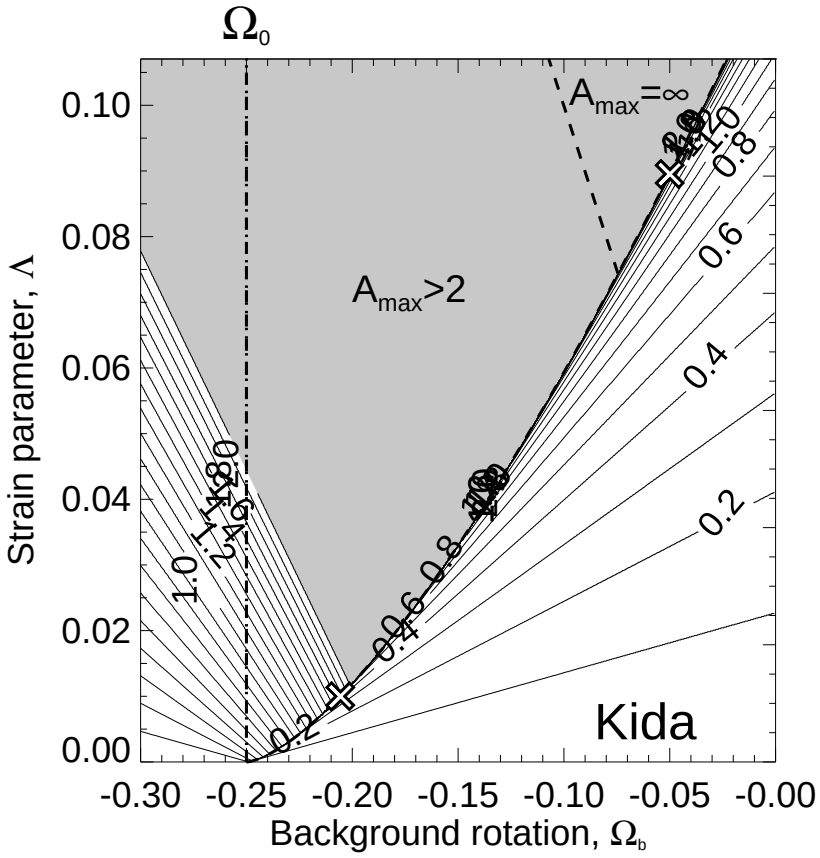

Figure 4.3: Contours of maximum disturbance amplitude A_{max} as a function of the background rotation Ω_b and Kida strain Λ. Shaded regions denote parameter space in which $A_{max} > 2$, with no contours shown in this region. Parameters which are shown in panels A2 and B2 of Fig. 4.4 are identified by crosses.

regime boundary between the oscillating and clockwise rotation regimes cannot be discerned from the behaviour of A_{max}.

A prediction of the jump in A_{max} across the curved regime boundary comes naturally from studying the behaviour of $V_e(r_e, \Omega_b, \Lambda)$. Should the evolution of the elliptical vortex be periodic, then $dr_e/dt = 0$ at the extremal aspect ratio, which is equivalent to $V(r_e; \Omega_b, \Lambda) = 0$ from (4.23). Fig. 4.4 gives examples of the Kida potential function $V_e(r_e; \Omega_b, \Lambda)$ for various values of Ω_b and Λ lying in the anticlockwise rotation, oscillating and infinite extension Kida regimes. In all panels, at $t = 0$ the initial aspect ratio is $r_e = 1$. As the vortex evolves, one travels along the graph for V_e in the region $0 < r_e \leq 1$ until the graph intersects with the r_e axis, at which point $V_e = 0$. The r_e value of this root is then the minimum aspect ratio of the vortex, and the evolution continues by retracing the V_e graph back towards $r_e = 1$. Once the graph reaches $r_e = 1$, the evolution continues by reversing direction and travelling back towards smaller values, giving a periodic motion. The only exception to this is if the minimum aspect ratio is $r_e = 0$. In this instance, motion along the V_e graph takes infinite time to reach $r_e = 0$.

Panels A1-A3, giving the Kida potential function near the regime boundary between the periodic clockwise and anticlockwise rotation regimes, show that there is a jump in the minimum aspect ratio $\min[r_e]$ achieved by the vortex depending on which regime the parameters (Ω_b, Λ) are in, with the jump occurring when the second stationary point of V_e is also a root of V_e. It is worth noting that although this implies a transition in the minimum aspect ratio of the vortex, we cannot make any conclusions about the direction of rotation, or whether the rotation has changed direction, based

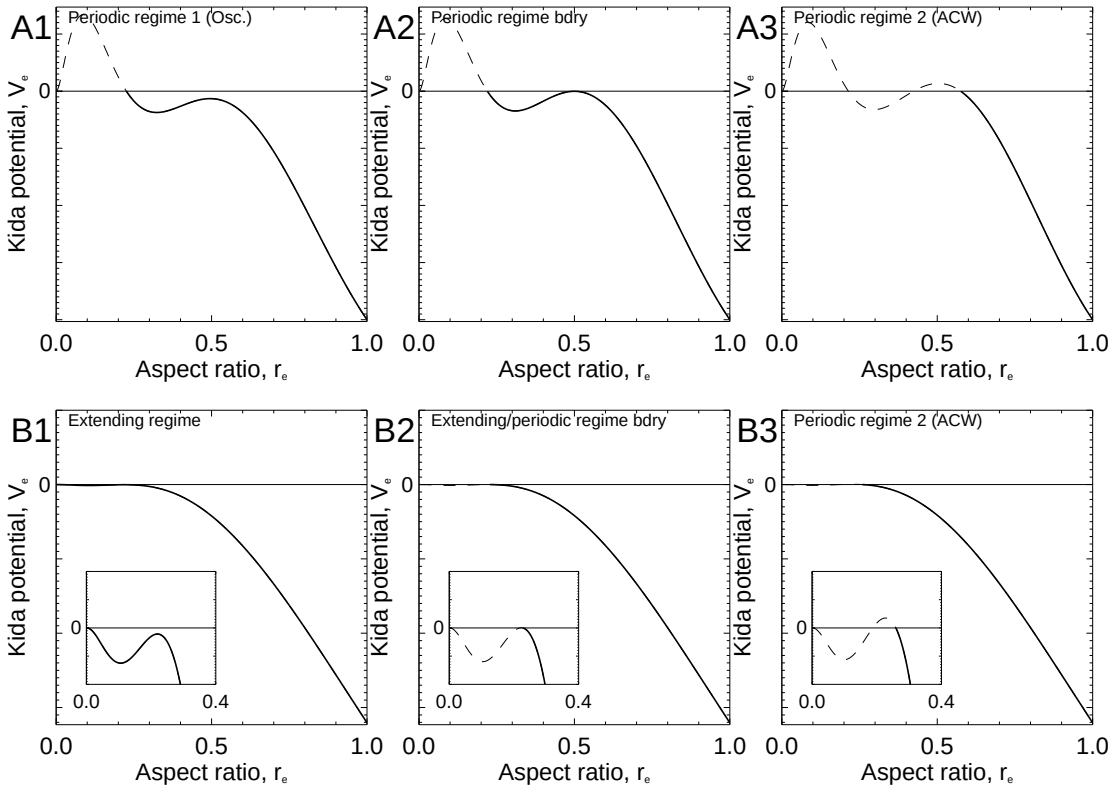

Figure 4.4: The Kida potential $V_e(r_e, \Omega_b, \Lambda)$ as a function of the Kida ellipse aspect ratio $r_e < 1$. A1: V_e for (Ω_b,Λ)=(-0.2036,0.0100) lying in the oscillating regime. A2: (Ω_b,Λ)=(-0.2056,0.0100) lying on the regime boundary between the oscillating and anticlockwise rotation regimes. A3: (Ω_b,Λ)=(-0.2076,0.0100) lying in the anticlockwise rotation regime. B1: (Ω_b,Λ)=(-0.0490,0.0895) lying in the extending regime. B2: (Ω_b,Λ)=(-0.0500,0.0895) lying on the regime boundary extending and anticlockwise rotation regimes. B3: (Ω_b,Λ)=(-0.0515,0.0895) lying in the anticlockwise rotation regime. In all panels, the heavy solid line denotes the portion of the graph for V_e along which motion is permitted during the vortex evolution, starting from $r_e = 1$. In plots B1-B3, magnifications are given for V_e in the interval $0 \leq r_e \leq 0.4$.

solely on the information presented here.

A similar transition in behaviour is shown in panels B1-B3. This transition in vortex behaviour, which occurs when the second stationary point of V_e is also a root of V_e, represents a regime boundary between anticlockwise rotating vortex states and infinitely extending vortex states. Once again, although we have identified the periodic vortex evolution as an anticlockwise rotating vortex state, this information cannot be derived solely from the plots of V_e shown, but is used to clarify which of the periodic/extension regime boundaries we are referring to in the Kida system. Although not shown, the final regime boundary between the clockwise rotation/oscillating vortex states and infinitely extending vortex states occurs when the first stationary point of V_e coincides with only real root of V_e at $r_e = 0$.

This study of V_e is analogous to that of $g(r_e, \Omega_b)$ given earlier in this section although, as has been stated, the direction of vortex rotation cannot be as easily deduced from V_e. However, as shall be seen in chapter 5, the concepts surrounding the derivation of the Kida potential V_e are important when deducing A_{max} in a shallow water model, and as such it has been included here as a precursor to that work.

4.4 A topographically forced vortex model

In this section, the behaviour of a somewhat more realistic member of our hierarchy of models is examined. Stratospheric dynamics are influenced by finite Rossby radius effects, and it is unclear that the influence of 'forcing' from the troposphere can be accurately represented by a uniform strain flow such as that present in the Kida vortex system. Hence we next study the dynamics of a vortex patch in the pres-

ence of topographic forcing using the quasi-geostrophic shallow water f-plane model described in chapter 3.

In dimensional form, the shallow water representation of an idealized vortex patch subject to topographic forcing $h_T(\mathbf{x})$ (see, for example Swanson 2000) is given by

$$\nabla^2_{\mathbf{x}}\psi - L_R^{-2}\psi + \frac{f_0 h_T}{H} + f_0 = \quad q = \quad \begin{cases} f_0 + \Delta_b + \Delta & \mathbf{x} \in D \\ f_0 + \Delta_b & \mathbf{x} \notin D. \end{cases} \tag{4.24}$$

where ψ is the streamfunction of the flow and q is the PV. In equation (4.24), D is the patch interior, L_R the Rossby radius of deformation, H the reference depth of the single layer model, f_0 the Coriolis parameter evaluated at the pole, ψ the streamfunction of the flow, Δ the uniform PV anomaly inside the vortex, and $\nabla^2_{\mathbf{x}}$ the two-dimensional horizontal Laplacian in terms of the planar coordinates $\mathbf{x}$. The parameter Δ_b represents a solid body rotation which is added to the flow. In the dimensional system, topographic forcing is introduced in polar coordinates (r, ϕ) as

$$h_T(r, \phi) \quad = \quad h_0 J_k(lr) \exp\{ik\phi\}, \tag{4.25}$$

where J_k is the Bessel function of the first kind, h_0 is the forcing height, $k \geq 2$ is an integer, and l is a parameter enabling the topography to be scaled in the radial coordinate.

As the operator $\nabla^2_{\mathbf{x}} - L_R^{-2}$ in (4.24) is linear, it is possible to express the streamfunction and PV of the system as linear combinations $\psi = \psi_d + \psi_b + \psi_T$ and $q = q_d + q_b + q_T$, where ψ_d is the streamfunction due to the disturbed vortex satisfying

$$\nabla^2_{\mathbf{x}}\psi_d - L_R^{-2}\psi_d = \quad q_d = \quad \begin{cases} \Delta & \mathbf{x} \in D \\ 0 & \mathbf{x} \notin D, \end{cases} \tag{4.26}$$

ψ_b is the streamfunction of the solid body background rotation

$$\nabla^2_{\mathbf{x}}\psi_b - L_R^{-2}\psi_b = \quad q_b = \quad \Delta_b, \tag{4.27}$$

and ψ_T is the streamfunction component induced by the topographic forcing which satisfies

$$\nabla^2_{\mathbf{x}}\psi_T - L_R^{-2}\psi_T = \quad q_T = \quad -\frac{f_0 h_T(r,\phi)}{H}. \tag{4.28}$$

At this stage it is convenient to nondimensionalize these equations in the same way as section 4.3 where $\mathbf{x} = a_0\tilde{\mathbf{x}}$, where a_0 is a typical vortex radius, $t = \tilde{t}/\Delta$ and $\psi = \Delta a_0^2\tilde{\psi}$. Equations (4.26) and (4.28) are then rewritten as

$$\nabla^2_{\tilde{\mathbf{x}}}\tilde{\psi}_d - \mathcal{B}^2\tilde{\psi}_d \quad = \quad \begin{cases} \xi_b + 1 & \tilde{\mathbf{x}} \in \tilde{D} \\ \xi_b & \tilde{\mathbf{x}} \notin \tilde{D}, \end{cases} \tag{4.29}$$

$$\nabla^2_{\tilde{\mathbf{x}}}\tilde{\psi}_b - \mathcal{B}^2\tilde{\psi}_b \quad = \quad \xi_b \tag{4.30}$$

$$\nabla^2_{\tilde{\mathbf{x}}}\tilde{\psi}_T - \mathcal{B}^2\tilde{\psi}_T \quad = \quad -\tilde{h}_T(\tilde{r},\phi), \tag{4.31}$$

where $\mathcal{B} = a_0/L_R$. The nondimensional topography which appears in (4.31) is

$$\tilde{h}_T(\tilde{r},\phi) \quad = \quad \mathcal{H}J_k(\lambda\tilde{r})\exp\{ik\phi\} \tag{4.32}$$

where $\mathcal{H} = f_0 h_0/\Delta H$ is a nondimensional forcing amplitude and $\lambda = la_0$ is a nondimensional radial wavenumber. The nondimensional background PV which appears in (5.4) is $\xi_b = 2\Omega_b = \Delta_b/\Delta$. The nondimensional angular velocity Ω_b is similar to the background rotation in the Kida vortex model, with the two being identical when $\mathcal{B} = 0$ ($L_R \to \infty$).

The nondimensional system can therefore be described by five nondimensional parameters:

- Ω_b: a tuning parameter with which the vortex can be brought into, and out of, resonance with the topographic forcing.
- $\mathcal{H}, k, \lambda$: forcing parameters with which the amplitude, azimuthal structure and radial structure of the topographic forcing are controlled.
- $\mathcal{B}$: the inverse of the nondimensional Rossby radius.

Now that the nondimensional equations of the system have been derived, for aesthetic reasons we drop all tildes on nondimensional quantities and specify that all variables from here on are assumed to be nondimensional.

4.4.1 The Kida vortex as an approximation to vortex dynamics in a shallow water model

When the nondimensional parameter $\mathcal{B} = 0$, the dynamics in the shallow water model reduce to that of purely two-dimensional flow. Therefore, as part of our hierarchy of models approach, we now look for a link between the topographic forcing in the shallow water system with $\mathcal{B} = 0$ and the Kida strain forcing in the purely two-dimensional system.

When $\mathcal{B} = 0$ in the shallow water model, the streamfunction component ψ_T which is induced by the topography is

$$\psi_T(r, \phi, t) \quad = \quad -\frac{\mathcal{H}}{\lambda^2} J_k(\lambda r) \exp\{ik\phi\}. \tag{4.33}$$

In the purely two-dimensional system, we now recall the streamfunction due to the strain component of the forcing in the Kida vortex model

$$\psi_E(r,\phi) \;=\; -\frac{\Lambda r^2}{2}\exp\{2i\phi\}. \tag{4.34}$$

By taking $k = 2$ in (4.33), it is seen that the azimuthal structures of ψ_T and ψ_E are the same, allowing a direct comparison to be drawn between the two streamfunctions. We now demand that the forcing streamfunction, when calculated at the vortex edge $r = 1$, be the same in both models. Thus, by equating (4.34) and (4.33) at $r = 1$, a relationship is derived between the nondimensional parameters $\mathcal{H}$, λ and Λ giving

$$\Lambda \;=\; \frac{2\mathcal{H}J_2(\lambda)}{\lambda^2} = F. \tag{4.35}$$

Fig. 4.5 gives a comparison of the radial structure of the Kida and topographic streamfunctions $|\psi_E| = \Lambda r^2/2$ and $|\psi_T| = \mathcal{H}J_2(\lambda r)/\lambda^2$ for various values of the topographic radial wavenumber λ, when $\mathcal{B} = 0$. It is seen that as long as λ is not too large, the Kida strain streamfunction acts as a good approximation to the topographic forcing streamfunction in the region of the vortex $r \lesssim 1$. Therefore it is to be expected that given λ is not too large, an initially circular vortex in the topographically forced model will evolve in a similar way to the Kida vortex; that is clockwise rotating, anticlockwise rotating or oscillating vortex states should be observed. It is clear, however, that the extending vortex solutions of the Kida vortex cannot exist in the topographically forced system, as the vortex, however small initially, must eventually extend to values of r at which the Bessel function structure of the topographic streamfunction becomes prominent.

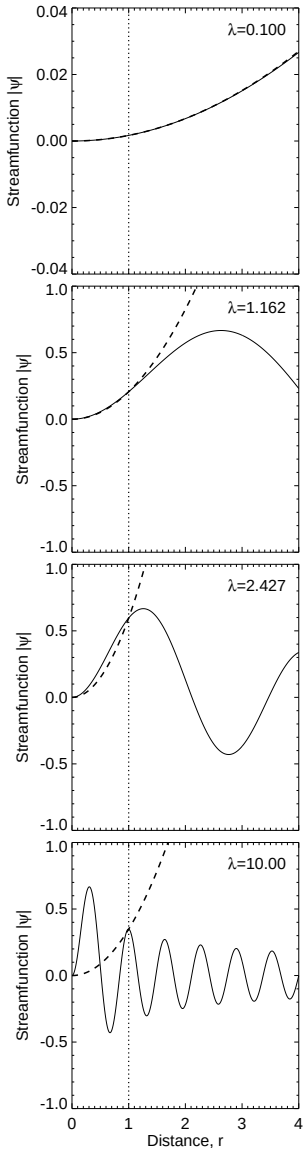

Figure 4.5: Magnitude of the streamfunction for the Kida strain $|\psi_E|$ (heavy dashed line) and topographic forcing $|\psi_T|$ with $k = 2$ (solid line) as a function of the radial coordinate r for four different values of radial wavenumber λ. The location of the undisturbed vortex edge is marked at $r = 1$ by the dotted line. The values $\lambda = 1.162$ and $\lambda = 2.427$ correspond to "hemispheric scale" and "vortex scale" forcing respectively (see Esler and Scott 2005). A factor of $\mathcal{H}/\lambda^2$ is omitted from $|\psi|$ for both streamfunctions.

Stability predictions for the shallow water model

One of the key objectives of this chapter is to investigate how instabilities leading to vortex-splitting in the topographically forced shallow water model may be predicted by corresponding instabilities of the Kida vortex. As has been shown above, for a sufficiently large value of λ in the shallow water model forced by (4.32) with $k = 2$ and $\mathcal{B} = 0$, the resulting vortex shape will be approximately that of the elliptical Kida vortex. Therefore, it is expected that the topographically forced model will exhibit unstable behaviour when parameters $(\mathcal{H}, \lambda, \Omega_b)$ correspond to values of (Λ, Ω_b) in the Kida model for which the elliptical vortex is unstable, as previously discussed in sections 4.2 and 4.3.2.

As discussed in section 4.2, for a Kirchhoff ellipse with $r_e \lesssim 1/6$, disturbances with elliptic azimuthal wavenumber-2 become nonlinearly unstable, leading to a split of the elliptical vortex. Hence, for (Λ, Ω_b) lying in unstable Kida regimes where the minimum aspect ratio of the ellipse is $r_e < 1/6$, it is proposed that vortex-splitting behaviour may be observed in the topographically forced model for parameters $(\mathcal{H}, \lambda, \Omega_b)$ corresponding to (Λ, Ω_b). On the other hand, for (Λ, Ω_b) lying in unstable Kida regimes where the minimum aspect ratio of the ellipse satisfies $1/6 \lesssim r_e \lesssim 1/4$, it is proposed that unstable growth of elliptic azimuthal wavenumber-4 disturbances will result in vortex filamentation rather than vortex-splitting behaviour.

4.4.2 Excitation of linear disturbances on a circular vortex

To understand the predictions of the nonlinear Kida model, we next investigate the properties of an unforced circular vortex in the shallow water model. We now derive

a linear dispersion relation for infinitesimal waves on the edge of a circular vortex in the nondimensional quasi-geostrophic f-plane model. The working presented here will closely follow that seen in Swanson (2000).

Wave-like disturbances on the vortex edge are expressed using a linear normal mode of the form

$$\eta(\phi) \quad = \quad \epsilon\alpha_1 \exp\{ik(\phi - \omega t)\}, \tag{4.36}$$

where ω is the frequency at which the wave translates around the vortex boundary.

The streamfunction ψ_d resulting from this disturbance to the vortex boundary satisfies equation (4.29) and can be expressed as $\psi_d(r,\phi) = \Psi(r) + \psi'(r,\phi)$, where $\Psi(r)$ is the contribution due to the basic state and ψ' is the contribution due to the infinitesimal perturbations. When partitioned in this way, the basic state streamfunction satisfies

$$\nabla_H^2\Psi - \mathcal{B}^2\Psi \quad = \quad \begin{cases} 1 & r < 1 \\ 0 & r > 1, \end{cases}$$

giving

$$\Psi(r) \quad = \quad \frac{1}{B}\begin{cases} K_1(\mathcal{B})I_0(\mathcal{B}r) - I_1(\mathcal{B})K_0(\mathcal{B}) & r < 1 \\ K_1(\mathcal{B})I_0(\mathcal{B}) - I_1(\mathcal{B})K_0(\mathcal{B}r) & r > 1, \end{cases}$$

$$U(r) \quad = \quad \frac{2\Omega_b + 1}{B}\begin{cases} K_1(\mathcal{B})I_1(\mathcal{B}r) & r < 1 \\ I_1(\mathcal{B})K_1(\mathcal{B}r) & r > 1, \end{cases}$$

where I and K are the modified Bessel functions of the first and second kind respectively and $U(r)$ is the velocity of the basic state, which is purely azimuthal. The streamfunction due to the infinitesimal perturbations then satisfies

$$\nabla_H^2\psi' - \mathcal{B}^2\psi' \quad = \quad 0. \tag{4.37}$$

Following Swanson (2000), the perturbation is expressed as $\psi'(r,\phi) = \tilde{\psi}(r)\eta(\phi)$ satisfying the jump condition

$$[\psi'_r]_{r=1-}^{r=1+} = -\eta, \tag{4.38}$$

where the bracket denotes the jump in the radial derivative of ψ at the vortex edge when $r = 1$. Substituting this streamfunction back into (4.37) yields an ODE for $\tilde{\psi}(r)$

$$r^2\tilde{\psi}_{rr} + r\tilde{\psi}_r - \left(\mathcal{B}^2r^2 + k^2\right)\tilde{\psi} = 0, \tag{4.39}$$

which when used with (4.38) and the fact that ψ must be continuous on $r = 1$ yields

$$\tilde{\psi}(r;k) = \begin{cases} K_k(\mathcal{B})I_k(\mathcal{B}r) & \text{for} \quad r < 1 \\ I_k(\mathcal{B})K_k(\mathcal{B}r) & \text{for} \quad r > 1. \end{cases} \tag{4.40}$$

As the vortex boundary is a material surface in the flow, the kinematic condition

$$\frac{\partial\eta}{\partial t} + \frac{u_d + u_b}{r}\frac{\partial\eta}{\partial\phi} - v_d = 0 \quad \text{on} \quad r = 1 + \eta(\phi,t),$$

must be satisfied on the vortex edge. Here, $u_b = 2\Omega_b I_1(\mathcal{B}r)K_1(\mathcal{B})$ is the azimuthal velocity due to the solid body rotation streamfunction ψ_b satisfying equation (4.30). By linearizing around $r = 1$, and noticing that the radial velocity component is $v_d = -(1/r)\psi'_\phi = ikI_k(\mathcal{B})K_k(\mathcal{B})\eta$, the linear dispersion relation is then given by

$$\omega = I_1K_1 - I_kK_k + 2\Omega_bI_1K_1, \tag{4.41}$$

where all Bessel functions are evaluated at $\mathcal{B}$.

Resonance of a wave-like disturbance on the vortex edge in the topographically forced system is expected to occur whenever the disturbance appears stationary with

respect to the forcing. For a system in which the forcing is fixed, this is equivalent to setting $\omega = 0$ in (4.41), such that the tuning parameter necessary for these stationary disturbances is

$$\Omega_b = \Omega_0(\mathcal{B}) = \frac{I_k K_k - I_1 K_1}{2 I_1 K_1}. \tag{4.42}$$

For a particular choice of $\mathcal{B}$, the function $\Omega_0(\mathcal{B})$ returns a single scalar value. Therefore, for convenience, the value $\Omega_0(\mathcal{B})$ will simply be referred to as Ω_0 when the value of the variable $\mathcal{B}$ has been given explicitly.

By utilizing the asymptotic forms of the I and K modified Bessel functions

$$I_k(\mathcal{B}) \sim \frac{1}{\Gamma(k+1)} \left(\frac{\mathcal{B}}{2}\right)^k, \tag{4.43}$$

$$K_k(\mathcal{B}) \sim \frac{\Gamma(k)}{2} \left(\frac{\mathcal{B}}{2}\right)^{-k}, \tag{4.44}$$

when $\mathcal{B} \to 0$, where $\Gamma(k) = (k-1)!$ is the Gamma function, it is seen that in the appropriate limit, $\Omega_0(\mathcal{B})$ agrees with that corresponding to purely two-dimensional disturbances

$$\Omega_0(0) = -\frac{k-1}{2k}$$

as seen in Su (1979) and Saffman (1992, equation 6, page 172). Looking at the rotation rate of the Kida ellipse in zero strain by taking $\Lambda = 0$ in equation (4.14), if the ellipse is nonrotating we have

$$\Omega_b = -\frac{r_e}{(r_e+1)^2},$$

which, for r_e close to 1 gives

$$\Omega_b \approx -\frac{1}{4}.$$

This is exactly the equal to $\Omega_0(0)$ when $k = 2$.

4.5 Nonlinear model results

4.5.1 Topographic forcing and vortex conditions

The evolution of an initially circular vortex patch subject to topographic forcing of the form

$$h_T(r, \phi, t) \;=\; \mathcal{H} J_k(\lambda r) \exp\{ik\phi\}, \tag{4.45}$$

is now investigated using a fully nonlinear quasi-geostrophic (QG) f-plane model discussed in section 4.4.

By virtue of the nondimensionalization of section 4.4, the circular vortex patch is initialized with unit radius and unit interior potential vorticity. To ensure that the velocity field associated with the topographic forcing (4.45) can be approximated by that of the Kida system, $k = 2$ is used in (4.45) and $\mathcal{B} = 0$ in (4.29) for all experiments, giving $\Omega_0 = -0.25$ corresponding to purely two-dimensional dynamics.

In addition to the topographic forcing, a background solid body rotation with angular velocity Ω_b is added to the flow. When investigating the role of resonant growth of disturbances on the vortex edge, the angular velocity Ω_b acts as a tunable parameter of the system, the varying of which brings the vortex into or out of resonance with the topographic forcing. A total of 31 values were chosen for the tuning parameter Ω_b in the fully nonlinear experiments. The smallest value is $\Omega_b = -0.30$, with experiments performed at increments of 0.01, giving the largest value of the tuning parameter as $\Omega_b = 0$.

When presenting our results, the topographic forcing amplitude is given in terms of $F = 2\mathcal{H}J_2(\lambda)/\lambda^2$ (see equation 4.35). The values of forcing parameter which are used in the experiments are partitioned into two sets. The first consists of four experiments representing small F values, $F = [0.00178, 0.00357, 0.00535, 0.00714]$. The second set starts at $F = 0.00892$ and includes 12 experiments, going up to $F = 0.10706$ in increments of 0.00892.

For all experiments, the radial dependence of the topographic forcing is scaled using $\lambda = 1.162$, and results in the first peak in topographic forcing occurring at a distance from the origin of approximately 2.67 vortex radii, as seen in panel B of Fig. 4.5. This scale is chosen to agree with the 'hemispheric scale' forcing discussed in Esler and Scott (2005) and to ensure that in the region of the undisturbed vortex the topographic streamfunction can be approximated by that of the Kida strain.

4.5.2 Model details and numerical parameters

The numerical model used to perform these experiments is the contour dynamics method of Dritschel (1988) (see chapter 3 for an overview of the contour dynamics numerical algorithm).

The node resolution parameter used in the contour dynamics method of Dritschel (1988) is taken to be $\mu = 0.025$, the time interval used is $dt = 0.05$, and the maximum time of the experiment is taken as $t_{max} = 40$. Data are generated at time intervals of 0.25 nondimensional time units, giving a total of 160 time steps of data For the κ_4 vortex split diagnostic, the split threshold parameter is taken to be $\mu_4 = -0.6$, such that if $\kappa_{min} = \min[\kappa_4] < -0.6$ a vortex split is identified (see chapter 2.2.4 for a

detailed discussion of the κ_4 diagnostic). When interpreting the κ_4 diagnostic, high positive values indicate that the vortex is undergoing a process of filamentation.

4.5.3 Vortex moment diagnostics

To compare the behaviour of the Kida vortex with that of the fully nonlinear model, it is useful to employ the vortex moment diagnostics as described in section 2.2.4 of chapter 2. For convenience, these are repeated here in the context of a single layer numerical model.[1]

For a patch of constant interior vorticity with jump in vorticity Δ at the vortex boundary, the absolute and relative moments of vorticity are calculated using

$$M_{m\,n} \;=\; \Delta \iint x^m y^n dx\, dy, \tag{4.46}$$

$$J_{m\,n} \;=\; \Delta \iint (x-\bar{x})^m (y-\bar{y})^n \; dx\, dy, \tag{4.47}$$

where

$$(\bar{x}, \bar{y}) \;=\; \frac{1}{M_{0\,0}} (M_{0\,1}, M_{1\,0}).$$

In a numerical model in which the contour dynamics algorithm is used, the integrals in (4.46) and (4.47) are converted to integrals around the boundary of the patch using Green's theorem. These contour integrals are calculated from the locations of the PV nodes which define the vortex boundary in the model. The orientation ϕ_e and aspect

[1]Note: vortex moments in the three-dimensional models of chapters 6 and 7 are calculated by applying the single layer methodology presented here to each height layer of the three-dimensional model in turn.

ratio r_e of the equivalent ellipse associated with the vortex patch are then given by

$$\phi_e = \frac{1}{2}\arctan\left(\frac{2J_{11}}{J_{20}-J_{02}}\right), \tag{4.48}$$

$$r_e = \left|\frac{(J_{20}+J_{02})+\sqrt{4J_{11}^2+(J_{20}-J_{02})^2}}{(J_{20}+J_{02})-\sqrt{4J_{11}^2+(J_{20}-J_{02})^2}}\right|^{1/2}. \tag{4.49}$$

4.5.4 Measures of finite amplitude vortex disturbances

When presenting the results in this section, one of the primary diagnostics used in determining the behaviour of wavelike disturbances on the vortex edge of the form $\eta(\phi,t) = \epsilon\alpha_1 \exp\{ik\phi\}$ is the amplitude of the first harmonic, $\epsilon\alpha_1$. To remove the dependence on ϵ, which is an artifact of the treatment of the linear analysis, we define the overall amplitude of the first harmonic as $A = \epsilon\alpha_1$. As previously discussed in section 4.3.3, in the Kida vortex model an approximation to the amplitude of the harmonic $\exp\{2i\phi\}$, equivalent to taking $k = 2$ in the topographically forced model, is given by

$$A = \frac{1-r_e}{2r_e},$$

where $r_e < 1$ is the minimum aspect ratio attained by the Kida vortex.

For the fully nonlinear numerical model, the quantity A is derived from the angular impulse of the vortex

$$\mathcal{I} = \frac{1}{2}\iint r^2 dxdy,$$

which is an invariant of the system in the absence of topographic forcing. Note that nondimensionalization of time by the constant interior potential vorticity Δ removes

a factor of Δ from $\mathcal{I}$. By subtracting $\mathcal{I}$ for the undisturbed vortex $\mathcal{I}_0 = \frac{\pi}{2}$ from the angular impulse, the wave-activity (pseudomomentum) of the vortex is defined as (see Dritschel and Saravanan 1994)

$$\mathcal{W}(t) = \frac{1}{8}\int_0^{2\pi} \Big[(1+\eta)^2 - 1\Big]^2 d\phi,$$

$$= \frac{\pi}{2}\epsilon^2\alpha_1^2 + O(\epsilon^3).$$

So by calculating the wave-activity of the vortex in the fully nonlinear model, the corresponding amplitude of the first harmonic of the wave-like disturbance η is approximated by

$$A = \sqrt{\frac{2\mathcal{W}(t)}{\pi}}.$$

4.5.5 Results from the single layer model

Classification of fully nonlinear experiments

To assess the similarities between the Kida vortex model and the topographically forced model, we classify the vortex evolution in each experiment into one of six categories:

−	Clockwise rotation: $\dot{\phi_e} < 0$ throughout the motion.
+	Anticlockwise rotation: $\dot{\phi_e} > 0$ throughout the motion.
H	Highly disturbed: $A_{max} > 1$.
H−	Highly disturbed clockwise rotation: $A_{max} > 1$ with $\dot{\phi_e} < 0$ throughout the motion.
S	Split: $\kappa_{min} < -0.6$.
O	Oscillating/other: all other vortex behaviour.

Fig. 4.6 shows the vortex evolution in the topographically forced model in relation to

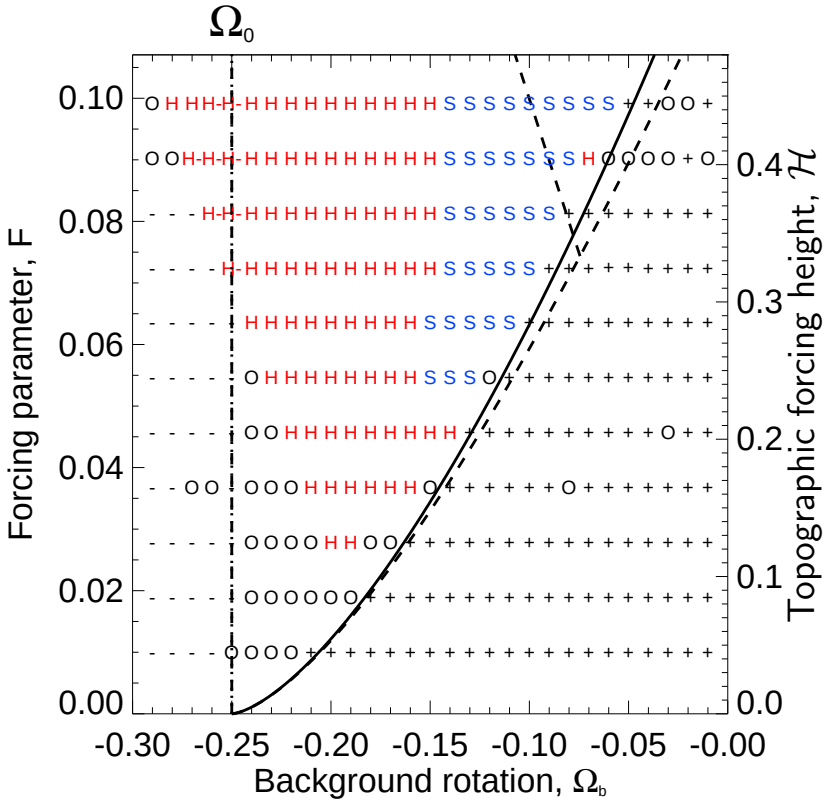

Figure 4.6: Classification of vortex evolution in the topographically forced model when $\mathcal{B} = 0$ as a function of Ω_b and F. Classification definitions are as described in the text. Dashed lines give the location of regime boundaries in the Kida vortex model, and the dotted line shows $\Omega_0 = -0.25$. The heavy solid line gives the location of the peak in A_{max} as predicted by the weakly nonlinear model in chapter 5.

the regime boundaries of the Kida vortex model. It is clear that the transition boundary between clockwise and anticlockwise rotating behaviour is accurately predicted by the Kida vortex model. Of the 55 experiments lying in the clockwise rotation regime of the Kida model, only 8 are not classified as clockwise rotating vortex states. Furthermore, no clockwise rotating states are observed for parameters outside of this region. Of the 122 experiments classified as anticlockwise rotating vortex states, only 5 lie outside of the anticlockwise rotation regime of the Kida model, with these 5 experiments occurring just to the left of the curved Kida regime boundary. For 125 experiments lying within the anticlockwise Kida regime, only 8 fully nonlinear experiments are not classified as anticlockwise rotating vortex states. Between the vertical (clockwise/oscillating) Kida regime boundary and the curved Kida regime boundary, almost all experiments are classified as either oscillating, highly disturbed or vortex-splitting vortex states. The sharp boundary observed between clockwise rotating and oscillating behaviour also bolsters confidence that the oscillating regime, which is a new finding of this work, having not been identified by Dritschel (1990), is a distinct regime in the Kida vortex model. It is therefore seen that there is a strong link between vortex behaviour in the topographically forced model and that predicted by the Kida model.

Panel B of Fig. 4.7 shows contours of the maximum disturbance amplitude A_{max} for a range of parameter values (Ω_b, F) in the fully nonlinear topographically forced model. For small values of the forcing parameter F, the peak in A_{max} occurs near $\Omega_b = \Omega_0$ as predicted by linear theory. However, as the forcing parameter F increases, the location of the peak in A_{max} migrates to the right, away from $\Omega_b = \Omega_0$. The

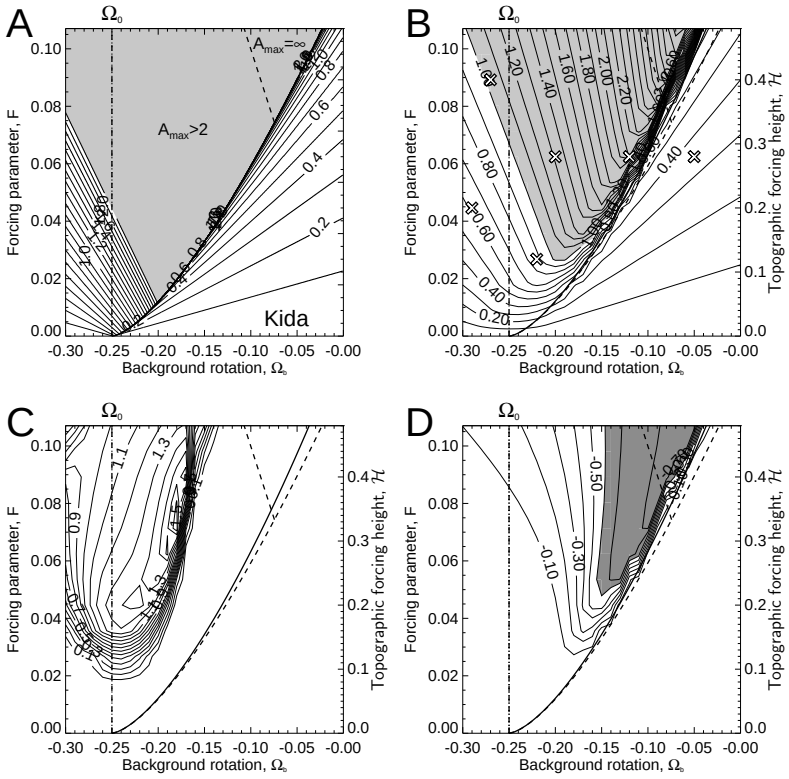

Figure 4.7: Panel A: contours of the maximum disturbance amplitude A_{max} as a function of background rotation Ω_b and forcing parameter F for the Kida model. The region in which $A_{max} > 2$ is shaded, and no contours are shown in this region. Panel B: as panel A but for the topographically forced fully nonlinear numerical model when $\mathcal{B} = 0$. Experiments which are shown in Fig. 4.8 are marked by crosses. The region in which $A_{max} > 1$ is shaded. Panel C: contours of κ_{max} as a function of Ω_b and F for the topographically forced fully nonlinear numerical model with $k = 2$ and $\mathcal{B} = 0$. Panel D: as panel C, but for κ_{min}. The region in which $\kappa_{min} < -0.6$, and a vortex split is detected, is shaded. In all panels, dashed lines give the location of regime boundaries in the Kida vortex model, the dotted line shows $\Omega_0 = -0.25$ and contours are shown at intervals of 0.1. In panels B-D, the heavy solid line gives the location of peak A_{max} which is discussed in chapter 5.

migration of the peak in A_{max} closely follows the curved anticlockwise rotation regime boundary of the Kida vortex model, with only small differences between the two at even the largest values of F. This difference between the Kida vortex model and the topographically forced model once F becomes large is to be expected, as the increasingly disturbed vortex states which correspond to such large forcing amplitudes are expected to outgrow the horizontal length scales over which the Kida strain acts as a good approximation to the topographic forcing (see Fig. 4.5).

In the region of parameter space to the right of the curved Kida regime boundary there is very good agreement between A_{max} in the topographically forced model and A_{max} as predicted by the Kida vortex model shown in panel A. In the region to the left of the curved Kida regime boundary, the values of A_{max} differ between the two models. However, there is still a good qualitative agreement such that when F is fixed, A_{max} decreases as Ω_b moves away from the the curved Kida regime boundary to more negative values. Furthermore, in both panels A_{max} values to the left of the curved Kida regime boundary are markedly higher than those to the right of the boundary. As these larger values of A_{max} correspond to vortex states which are approximately elliptical and with large aspect ratio, which are known to be linearly unstable to small perturbations (Dritschel 1986; Guo et al. 2004; Love 1893), it is expected that this region of parameter space will be populated by unstable vortex behaviour in the topographically forced model.

Panels C and D show the maximum and minimum values of the κ_4 parameter observed in each experiment during the vortex evolution. It is worth noting again that by design, $\kappa_4 = 0$ for both an undisturbed circular vortex with uniform interior

vorticity, and for an elliptical vortex. Therefore, $\kappa_{max} \geq 0$ and $\kappa_{min} \leq 0$. It is possible to observe vortex evolution exhibiting both large positive values of κ_{max}, indicating intense filamentation of the vortex, and large negative values of κ_{min}, indicating the vortex has become pinched. One such example would be a vortex which splits before undergoing intense filamentation and mixing with the background flow.

Looking at panels C and D, it is clear that large positive values of κ_{max} and large negative values of κ_{min} only occur to the left of the curved Kida regime boundary. That is, κ_4 remains almost zero throughout the vortex evolution in the region to the right of the curved regime boundary, indicating that the vortex remains elliptical at all times in this region. This behaviour is hinted at by the stability analysis of Dritschel (1990) for the Kida vortex model, as summarized schematically in panel B of Fig. 4.2. That is, stability of the Kida vortex model in this region of parameter space seems to imply stability of the topographically forced vortex in the same region.

In panel C, large values of κ_{max} observed to the left of the unstable region indicate that vortex instability in the topographically forced model leads to strong vortex filamentation for these experiments. For the values of forcing parameter F shown here, large negative values of κ_{min} are only observed in the region between the curved Kida regime boundary and the vertical (clockwise/oscillating) regime boundary. In particular, vortex splits, defined as occurring when $\kappa_{min} < -0.6$, are only observed in a narrow wedge of parameter space immediately to the left of the curved Kida regime boundary, and only for $F > 0.05$. As F increases, the vortex split sub-regime widens, indicating that values of Ω_b further from the curved boundary will also lead to vortex-splitting. Low values of κ_{max} in this vortex-splitting sub-regime indicate

that in this region, instability manifests itself primarily as large scale splitting of the vortex, rather than a combination of vortex-splitting and vortex filamentation. This is a direct consequence of the unstable nature of an approximately elliptical vortex state with large aspect ratio, and these are the states which are observed in this region of parameter space, as shown by the large values of A_{max} in panel B of Fig. 4.7. As previously discussed, nonlinear instability of a highly elongated elliptic vortex leads to vortex-splitting, corresponding to growth of elliptic azimuthal wavenumber-2 disturbances on the vortex edge. For less extreme aspect ratios, nonlinear instability occurs due to growth of elliptic azimuthal wavenumber-4 disturbances, leading to filamentation of the vortex (Dritschel 1986).

Vortex structure and evolution

It is now useful to compare the vortex behaviour and instabilities predicted by Fig. 4.6 and Fig. 4.7, with actual observations of the vortex evolution during the experiments. In order to do this, a phase space representation of the vortex motion is obtained by introducing the variables $X = (r_e - r_e^{-1})\cos 2\phi_e$ and $Y = (r_e - r_e)\cos 2\phi_e$ (Dritschel 1990). The different categories of vortex behaviour are then well illustrated by plotting trajectories in (X, Y) phase-space.

Fig. 4.8 shows snapshots of vortex evolution for each of the types of classification appearing in Fig. 4.6, and Fig. 4.9 shows the corresponding trajectories of $X = (r_e - r_e^{-1})\cos 2\phi_e$ and $Y = (r_e - r_e^{-1})\sin 2\phi_e$ over the vortex evolution. Note the different range for the (X, Y)-axes in each of the panels in Fig. 4.9. For the quasi-periodic motion exhibited by the topographically forced model in panels A-E, the

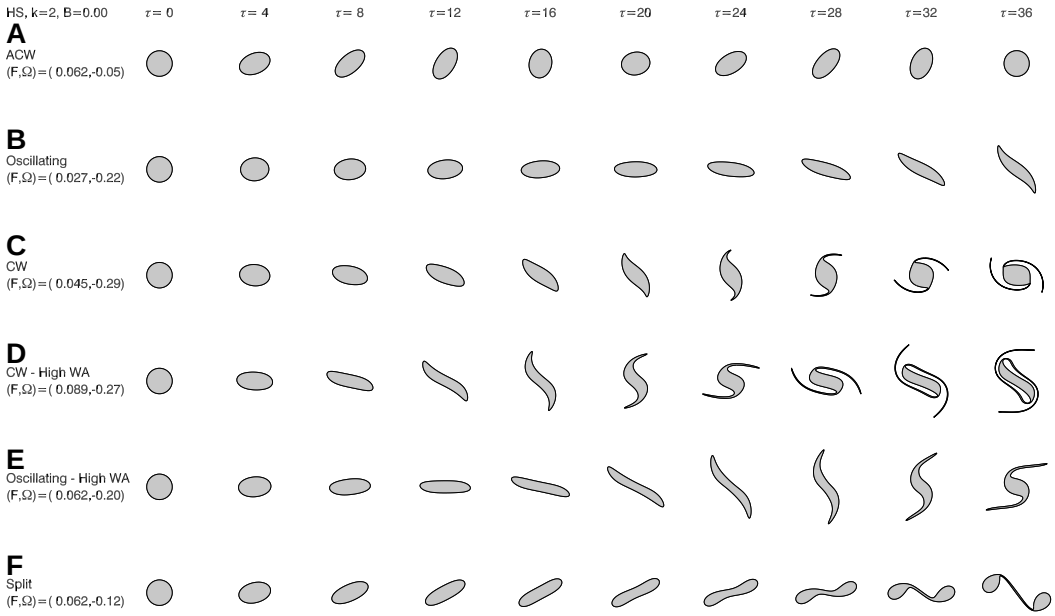

Figure 4.8: Snapshots of vortex evolution in the fully nonlinear numerical model when $\mathcal{B} = 0$ for the experiments marked by crosses in panel A of Fig. 4.7. Panel A: evolution classified as anticlockwise rotation when $F = 0.062$ and $\Omega_b = -0.05$. Panel B: evolution classified as oscillating/other when $F = 0.027$ and $\Omega_b = -0.22$. Panel C: evolution classified as clockwise rotation when $F = 0.045$ and $\Omega_b = -0.29$. Panel D: evolution classified as highly disturbed clockwise rotation when $F = 0.089$ and $\Omega_b = -0.27$. Panel E: evolution classified as highly disturbed oscillating/other when $F = 0.062$ and $\Omega_b = -0.20$. Panel F: evolution classified as a vortex-splitting event when with $F = 0.062$ and $\Omega_b = -0.12$.

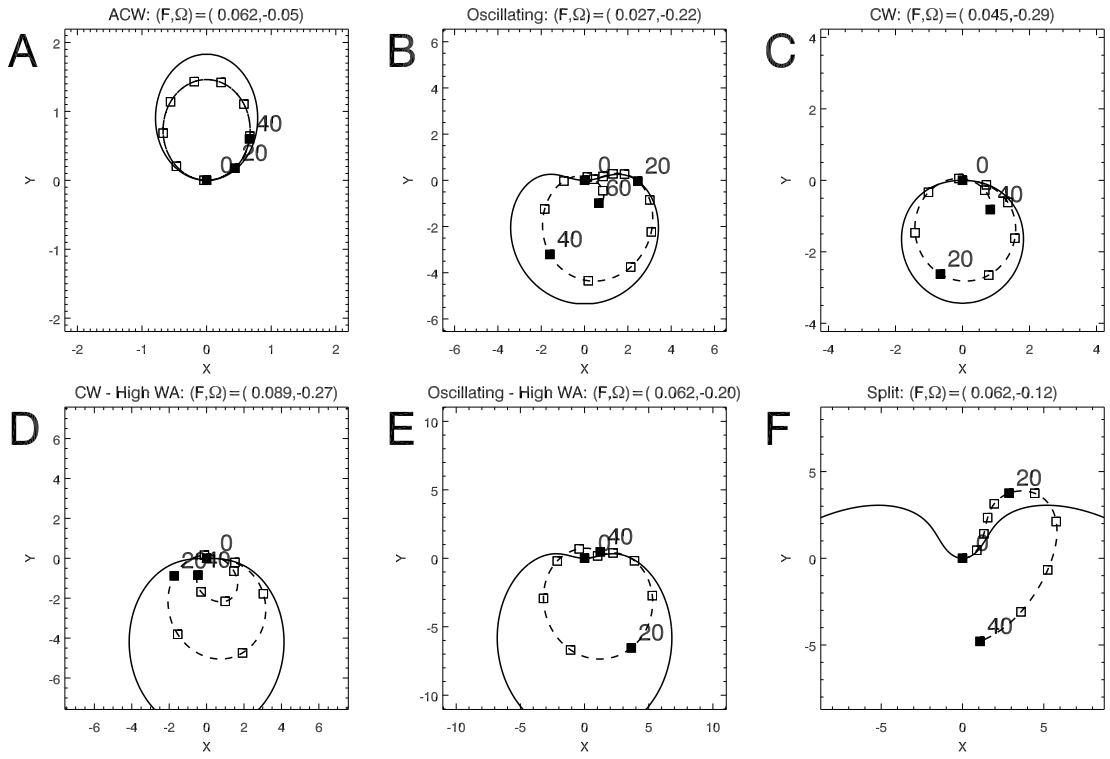

Figure 4.9: Phase plots of $X = (r_e - r_e^{-1})\cos 2\phi_e$ and $Y = (r_e - r_e^{-1})\sin 2\phi_e$ over the vortex evolution for the experiments marked by crosses in panel A of Fig. 4.7. Predictions for periodic evolution in the Kida vortex theory are shown by thick solid lines. Results from numerical experiments in the topographically forced model when $\mathcal{B} = 0$ are shown by dashed lines and boxes. Boxes are shown for every fourth time step, with filled boxes at $t = 0, 20, 40$. For the experiment in panel B, the maximum time of the experiment is $t = 60$.

(X, Y) trajectory represents at least one full period.

In panel A, the vortex undergoes anticlockwise rotation while maintaining an elliptic shape of varying aspect ratio, with the time taken for one complete revolution being approximately $t = 22$ as seen from panel A of Fig. 4.9.

Oscillating behaviour is observed in panel B, with the maximum aspect ratio of the vortex being much greater than that seen in panel A, even though the forcing parameter F is much smaller. In terms of the (X, Y) trajectory in panel B of Fig. 4.9, the oscillatory behaviour appears as a dip in the trajectory near to the origin. The shape of the vortex snapshot at $t = 36$ indicates that the tips of the major axes are beginning to roll up, and that the structure is unstable.

Clockwise rotation, and highly disturbed clockwise rotation are shown in panels C and D. It is seen from Fig. 4.9 that the vortex in panel C completes one full revolution at approximately $t = 32$, and in panel D at approximately $t = 24$. In both cases, vortex snapshots exhibit unstable behaviour of the vortex, with the instability being characterized by a growth of elliptical azimuthal wavenumber-4 disturbances on the vortex edge, leading to shedding of vorticity from the ends of the ellipse major axes (see, for example, Fig. 12c of Dritschel 1986). Similar behaviour is observed in panel E, classified as highly disturbed oscillatory motion. The vortex filamentation observed in panels C-E is in agreement with the large positive of values of κ_{max} for these experiments, as seen in panel C of Fig. 4.7.

In panel F of Fig. 4.8, snapshots indicate that the aspect ratio of the elliptical vortex becomes increasingly large, while exhibiting very little change in orientation, remaining almost fixed in relation to the topographical forcing. From the (X, Y)

trajectories in panel F of Fig. 4.9, it is seen that up until $t \approx 24$ the evolution of the vortex is roughly the same as that predicted by the Kida vortex. However, the snapshots show that by $t = 28$ an elliptical azimuthal wavenumber-2 disturbance is growing on the vortex edge (for example, Fig. 12a of Dritschel 1986) and the vortex becomes pinched. As indicated by the large negative value taken by κ_{min} for this experiment, growth of this disturbance results in a complete split of the vortex into two daughter vortices of comparable size. Although not shown, these daughter vortices then proceed to rotate clockwise about their common centroid. This splitting behaviour, in which the vortex orientation becomes fixed with respect to the topographical forcing while becoming increasingly elliptical, followed by a complete split of the vortex into two fragments, is highly reminiscent of the behaviour observed in the northern hemisphere polar vortex during vortex-splitting events, as seen in chapter 2.

4.6 Conclusions

In summarizing the results presented in this chapter, each of the questions Q1-Q3 presented as part of the introduction shall be addressed individually.

Q1: vortex-splitting in a single layer stratospheric model

In the topographically forcing shallow water model, in terms of vortex evolution, large values of A_{max} are observed as an elongated elliptical vortex. Dritschel (1986) showed that in the absence of background flow, very elongated elliptical vortices ($r_e \lesssim 1/6$) are unstable to elliptic azimuthal wavenumber-2 disturbances, the growth

of which pinches the vortex around its centre, eventually leading to vortex-splitting. For elliptical vortices which are moderately elongated ($1/6 \lesssim r_e \lesssim 1/5$), the instability of elliptic azimuthal wavenumber-4 disturbances dominate, leading to filamentation of vorticity from the ends of the ellipse major axes.

One expectation is that the unstable regimes of the topographically forced model, regions in which A_{max} is largest, will exhibit vortex-splitting instability, and regions with moderate to large A_{max} will exhibit instability dominated by filamentation. The numerical model results confirm that this is indeed the case, with vortex-splitting behaviour observed near the curved regime boundary, where A_{max} values are largest, and vortex filamentation behaviour observed in the unstable regimes away from the curved regime boundary.

Resonance alone is not sufficient for vortex-splitting behaviour. For vortex-splitting, Ω_b must be sufficiently close to resonance *and* the forcing F must be sufficiently large. For the choice of $\mathcal{B} = 0$ used in this chapter, it was seen that the forcing parameter F must exceed 0.05 for vortex-splitting. On the other hand, as F increases, the range of values for Ω_b resulting in vortex-splitting broadens, and splitting can occur further from resonance.

Q2: Excitation of a linear resonance and the role of nonlinearity

In the topographically forced model with $\mathcal{B} = 0$, for small values of the forcing parameter F, the peak in A_{max} is seen to occur when $\Omega_b = \Omega_0$ as predicted by a resonance of free-travelling small amplitude waves with the background flow. However, as F increases, the location of the peak in A_{max} migrates from Ω_0. This is due to

nonlinear effects, which modify the frequency of large amplitude free-travelling waves away from that predicted by linear theory.

This migration of A_{max} is accurately predicted by the curved Kida regime boundary when $\mathcal{B} = 0$ in the topographically forced model. This is due to the quasi-stationary behaviour of vortex states lying on this curved regime boundary in the Kida model, behaviour which is a necessary condition for resonance of the vortex with the background flow.

In chapter 5, weakly nonlinear theory is used to make quantitative predictions of the effects of nonlinearity on the migration of resonance observed in this chapter. This is particularly important for cases when $\mathcal{B} \neq 0$, where no direct comparison exists between the topographically forced model and the Kida vortex.

Q3: The Kida vortex model approximation and stability

The results of the previous section have shown that the Kida vortex model gives excellent predictions of the behaviour of an initially circular vortex subject to topographic forcing in a quasi-geostrophic shallow water model.

In the barotropic shallow water model with $\mathcal{B} = 0$ (which is equivalent to purely two-dimensional fluid flow) and for small forcing parameter F, the behaviour classification for the vortex evolution is almost exactly as predicted by the Kida model. This agreement is also observed for large forcing amplitudes should the flow parameters (Ω_b, F) lie in the anticlockwise rotation regime of the Kida model.

The stability of the topographically forced vortex can also be accurately predicted by analogy with the stable and unstable regimes of the Kida model. That is, anti-

clockwise rotating vortex states in the shallow water model are seen to be stable throughout their evolution, remaining almost elliptical at all times, as predicted by the stability of this regime in the Kida vortex model. In contrast, for flow parameters (Ω_b, F) lying in the clockwise rotation, oscillating and extending regimes, which are linearly unstable in the Kida vortex model, vortex evolution in the topographically forced model is also unstable, with both vortex filamentation and vortex-splitting type behaviour.

It has also been shown that for any choice of parameters (Ω_b, F), the maximum aspect ratio predicted by the Kida vortex model offers an insight into the maximum amplitude A_{max} of wave-like disturbances to the topographically forced vortex. In both models, values of A_{max} are largest in regimes to the left of the curved Kida regime boundary, with a jump in A_{max} at the curved boundary to much smaller values in the anticlockwise rotation regime on the right.

Chapter 5

Single Layer Shallow Water Model of Vortex-Splitting: Weakly Nonlinear Theory

5.1 Introduction

In chapter 4 it was shown that regimes of vortex behaviour in the Kida vortex model (Kida 1981) give accurate predictions for the behaviour of a circular vortex subject to forcing from topography. In particular, it was seen that the boundary between the anticlockwise rotation regime and oscillation or extension regimes in the Kida vortex system also plays a critical role in the topographically forced system. That is, forcing parameters (Λ, Ω) belonging to the linearly stable anticlockwise rotating regime in the analogous Kida vortex problem lead to anticlockwise rotating elliptical vortex states in the topographically forced system. Similarly, (Λ, Ω) belong-

ing to the linearly unstable regimes in the analogous Kida vortex problem lead to significant deformation of the circular vortex, in some cases resulting in a complete split of the vortex into two smaller fragments.

This chapter will investigate the possible role of a self-tuning resonance in explaining the abrupt transition between stable and unstable observed vortex evolution in the topographically forced model. It will be shown that the topographically forced vortex can be viewed as a forced weakly nonlinear oscillator, and consequently a regime boundary in the topographically forced system can be successfully predicted. This regime boundary in the nonlinear oscillator occurs as a result of initially off-resonant waves on the vortex edge self-tuning towards resonance. The weakly nonlinear predictions for self-tuning resonance are not limited to disturbances exhibiting 2-fold symmetry as typified by the elliptical vortex of chapter 4, but are shown to be valid for any disturbances exhibiting k-fold symmetry.

The chapter is structured as follows. In section 5.2.3, velocity fields are derived giving the velocity at points inside and outside a perturbed circular vortex in a quasi-geostrophic shallow water model. These velocity fields are then used in section 5.3 to derive a weakly nonlinear dispersion relation for wave-like disturbances on the vortex edge in the presence of a rotating background flow. The dispersion relation is then used in section 5.4 to derive theoretical predictions for the conditions leading to resonant wave growth in a linear, and then weakly nonlinear, shallow water model in the presence of topographic forcing. In the results sections, these predictions are used to interpret the findings of chapter 4, and are then used to interpret the results from two new sets of model runs. Finally, the conclusions of the chapter are presented.

5.2 Vorticity induced velocity fields in the single layer quasi-geostrophic f-plane

5.2.1 A uniform vortex patch in the shallow water equations

As discussed in chapters 3 and 4, in the shallow water f-plane representation of an idealized vortex patch, in the presence of topographic forcing $h_T(\mathbf{x})$, the streamfunction ψ is given in terms of the PV q as

$$\nabla^2_{\mathbf{x}}\psi - L_R^{-2}\psi + \frac{f_0 h_T}{H} + f_0 \;\; = \;\; \begin{cases} f_0 + \Delta_b + \Delta & \mathbf{x} \in D \\ f_0 + \Delta_b & \mathbf{x} \notin D. \end{cases} \tag{5.1}$$

In equation (5.1), D is the patch interior, L_R the Rossby radius of deformation, H the reference depth of the single layer model, f_0 the Coriolis parameter evaluated at the pole, ψ the streamfunction of the flow, Δ the uniform PV anomaly inside the vortex, and $\nabla^2_{\mathbf{x}}$ the two-dimensional horizontal Laplacian in terms of the planar coordinates $\mathbf{x}$. The parameter Δ_b represents a solid body rotation which is added to the flow.

Nondimensionalizing equation (5.1) in the same way as section 4.4 in chapter 4, using $\mathbf{x} = a_0\tilde{\mathbf{x}}$, $t = \tilde{t}/\Delta$, $q = \tilde{q}/\Delta$ and $\psi = \Delta a_0^2\tilde{\psi}$ where a_0 is a typical vortex radius, gives

$$\nabla^2_{\tilde{\mathbf{x}}}\tilde{\psi} - \mathcal{B}^2\tilde{\psi} + \tilde{h}_T(\tilde{\mathbf{x}}) = \;\; \tilde{q} = \;\; \begin{cases} \xi_b + 1 & \tilde{\mathbf{x}} \in \tilde{D} \\ \xi_b & \tilde{\mathbf{x}} \notin \tilde{D}. \end{cases} \tag{5.2}$$

with $\mathcal{B} = a_0/L_R$ defined as the inverse of the nondimensional Rossby radius. The nondimensionalized background rotation is $\xi_b = \Delta_b/\Delta$, and is related to the angular velocity of the solid body rotation Ω_b by $\xi_b = 2\Omega_b$. In the interests of clarity, in the

rest of this chapter all tildes will be omitted, and all variables are assumed to have been nondimensionalized as described above.

The linear property of the operator in (5.2) enables the streamfunction and PV of the flow to be partitioned into a component due to the disturbed vortex (ψ_d, q_d), a component due to the solid body rotation (ψ_b, q_b) and a component due to the topographic forcing (ψ_T, q_T) satisfying

$$\nabla^2\psi_d - \mathcal{B}^2\psi_d = \quad q_d = \quad \begin{cases} 1 & \mathbf{x} \in D \\ 0 & \mathbf{x} \notin D, \end{cases} \tag{5.3}$$

$$\nabla^2\psi_b - \mathcal{B}^2\psi_b = \quad q_b = \quad \xi_b, \tag{5.4}$$

$$\nabla^2\psi_T - \mathcal{B}^2\psi_T = \quad q_T = \quad -h_T(\mathbf{x}), \tag{5.5}$$

where $\psi = \psi_d + \psi_b + \psi_T$ and $q = q_d + q_b + q_T$ (see 4.4).

The topography considered in this chapter is of the same form as the nondimensional topography appearing in equation (5.5), and is given again for convenience as

$$h_T(r, \phi) \quad = \quad \mathcal{H} J_k(\lambda r) \exp\{ik\phi\}, \tag{5.6}$$

where (r, ϕ) are polar planar coordinates. Here, $\mathcal{H}$ is a nondimensional topographic forcing height, λ a nondimensional radial wavenumber for the topography and k an azimuthal integer wavenumber.

5.2.2 Velocity due to a uniform patch without topography

Considering a single vortex patch in the nondimensional system with interior area D, the streamfunction $\psi_d(\mathbf{x},t)$ induced in the flow by the vortex patch is found by inverting the operator in (5.3) giving

$$\psi_d(\mathbf{x},t) \;=\; \iint_D G(\mathbf{x},\mathbf{x}_1)d^2\mathbf{x}_1. \tag{5.7}$$

In (5.7), the Green's function $G(\mathbf{x},\mathbf{x}_1)$ for the shallow water model is defined as satisfying

$$\nabla^2 G - \mathcal{B}^2 G \;=\; \delta(\mathbf{x}-\mathbf{x}_1),$$

giving

$$G(\mathbf{x},\mathbf{x}_1) \;=\; -\frac{1}{2\pi}K_0\left(\mathcal{B}|\mathbf{x}-\mathbf{x}_1|\right), \tag{5.8}$$

where K_0 is the zero order modified Bessel function of the second kind.

An approach analogous to that of Su (1979) is now adopted in order to derive an integral expression for the velocity induced by the vortex patch at any point in the fluid. Utilizing the relationship between the streamfunction and the velocity $\mathbf{u}_d(\mathbf{x},t) = -\nabla\times\psi_d\mathbf{k}$, in conjunction with Green's Theorem in the plane, the integral (5.7) yields an expression for the velocity $\mathbf{u}_d$ in terms of an integral around the patch boundary ∂D

$$\mathbf{u}_d(\mathbf{x},t) \;=\; -\oint_{\partial D}\hat{\mathbf{t}}(\mathbf{x}_1)G(\mathbf{x},\mathbf{x}_1)ds_1, \tag{5.9}$$

where $\hat{\mathbf{t}}(\mathbf{x}_1)$ is the unit tangent vector counterclockwise along the boundary. Substituting $\hat{\mathbf{t}}(\mathbf{x}_1)ds_1 = d\mathbf{x}_1$ and letting $\mathbf{x}_1 = r(\phi)\cos\phi\mathbf{i} + r(\phi)\sin\phi\mathbf{j}$, (5.9) can be rewritten

as

$$\begin{aligned}\mathbf{u}_d(\mathbf{x},t) &= -\int_0^{2\pi} G(\mathbf{x},\mathbf{x}_1)\times \\ &\qquad \Big[\left(r'(\phi)\cos\phi - r(\phi)\sin\phi\right)\mathbf{i} + \left(r'(\phi)\sin\phi + r(\phi)\cos\phi\right)\mathbf{j}\Big]d\phi, \\ &= -\int_0^{2\pi} G(\mathbf{x},\mathbf{x}_1)\left(r(\phi)'\hat{\mathbf{r}}_1 + r(\phi)\hat{\boldsymbol{\phi}}\right)d\phi. \end{aligned} \tag{5.10}$$

By taking the dot product of (5.10) with $\hat{\mathbf{r}} = \cos\theta\mathbf{i} + \sin\theta\mathbf{j}$ and $\hat{\boldsymbol{\theta}} = -\sin\theta\mathbf{i} + \cos\theta\mathbf{j}$, the azimuthal and radial velocities $\mathbf{u} = u\hat{\boldsymbol{\theta}} + r\hat{\mathbf{r}}$ are given as

$$u_d(\mathbf{x},t) = \int_0^{2\pi} G(\mathbf{x},\mathbf{x}_1)\left(-r'(\phi)\sin(\theta-\phi) + r(\phi)\cos(\theta-\phi)\right)d\phi, \tag{5.11}$$

$$v_d(\mathbf{x},t) = \int_0^{2\pi} G(\mathbf{x},\mathbf{x}_1)\left(r'(\phi)\cos(\theta-\phi) + r(\phi)\sin(\theta-\phi)\right)d\phi. \tag{5.12}$$

It is now useful to define the function $\mathcal{T}_n(r_1,r_2)$ defined by

$$\mathcal{T}_n(r_1,r_2,\mathcal{B}) = \begin{cases} I_n(\mathcal{B}r_1)K_n(\mathcal{B}r_2) & r_1 < r_2 \\ K_n(\mathcal{B}r_1)I_n(\mathcal{B}r_2) & r_2 < r_1 \end{cases}$$

where I_n is the order n modified Bessel function of the first kind. For $r(\theta)$ and $r(\phi)$ shown in the schematic in Fig. 5.1, it is therefore seen that the form of $\mathcal{T}(r(\theta),r(\phi),\mathcal{B})$ is dependent on whether $r(\phi) < r(\theta)$ or $r(\theta) < r(\phi)$. The modified Bessel function in (5.8) can be expressed using Graf's formula (Watson 1944, §11.3) as the infinite sum

$$K_0(\mathcal{B}w) = \sum_{n=-\infty}^{\infty} \mathcal{T}_n\Big(r(\phi),r(\theta),\mathcal{B}\Big)\cos n(\theta-\phi). \tag{5.13}$$

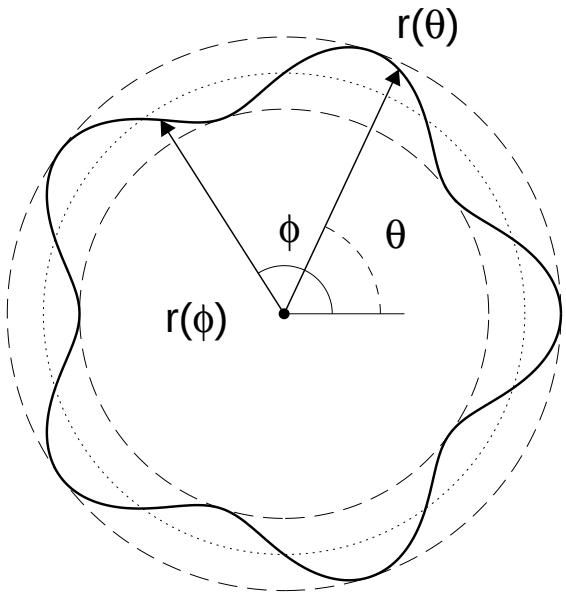

Figure 5.1: Schematic of a disturbed vortex boundary given by $r(\phi) = 1 + A\cos(k\phi)$ with $A = 0.15$ and $k = 5$. The region enclosing all points $r(\theta)$ which are not wholly inside or outside the boundary lies between the dashed lines. The dotted line shows $r = 1$. It is seen that $r(\theta) < r(\phi)$ or $r(\theta) > r(\phi)$ depending on the angle ϕ.

where $w = \sqrt{r(\phi)^2 + r(\theta)^2 - 2r(\phi)r(\theta)\cos(\theta - \phi)}$ and $\mu = \theta - \phi$. Substitution of (5.8) and (5.13) into the integral expressions for the azimuthal and radial velocities then gives

$$\begin{aligned} u_d(\mathbf{x},t) &= \frac{1}{4\pi}\sum_{n=-\infty}^{\infty}\int_0^{2\pi}\mathcal{T}_n\Big(r(\phi),r(\theta),\mathcal{B}\Big)\times \\ &\quad \Big[r'(\phi)\left(\sin\left[(n-1)\mu\right]-\sin\left[(n+1)\mu\right]\right) \\ &\quad + r(\phi)\left(\cos\left[(n-1)\mu\right]+\cos\left[(n+1)\mu\right]\right)\Big]d\phi, \end{aligned} \tag{5.14}$$

$$\begin{aligned} v_d(\mathbf{x},t) &= \frac{1}{4\pi}\sum_{n=-\infty}^{\infty}\int_0^{2\pi}\mathcal{T}_n\Big(r(\phi),r(\theta),\mathcal{B}\Big)\times \\ &\quad \Big[r'(\phi)\left(\cos\left[(n+1)\mu\right]+\cos\left[(n-1)\mu\right]\right) \\ &\quad + r(\phi)\left(\sin\left[(n+1)\mu\right]-\sin\left[(n-1)\mu\right]\right)\Big]d\phi. \end{aligned} \tag{5.15}$$

Utilizing the fact that $K_n = K_{-n}$ and $I_n = I_{-n}$, the infinite summation can be reduced to a summation over the positive integers $n \geq 2$ only, that is

$$\begin{aligned} \sum_{n=\infty}^{\infty} I_n K_n\left[\cos(n+1)\mu+\cos(n-1)\mu\right] &= 2I_1K_1 \\ &\quad +2\sum_{n=2}^{\infty}\left(I_nK_n+I_{n-2}K_{n-2}\right)\cos(n-1)\mu, \end{aligned}$$

and

$$\sum_{n=\infty}^{\infty} I_n K_n\left[\sin(n+1)\mu+\sin(n-1)\mu\right] = 2\sum_{n=2}^{\infty}\left(I_nK_n+I_{n-2}K_{n-2}\right)\sin(n-1)\mu,$$

the velocity integrals (5.14) and (5.15) are given in final form as

$$\begin{aligned} u_d(\mathbf{x},t) &= \frac{1}{2\pi}\int_0^{2\pi}\mathcal{T}_1\Big(r(\phi),r(\theta),\mathcal{B}\Big)r(\phi)d\phi \\ &\quad + \frac{1}{2\pi}\sum_{n=2}^{\infty}\Bigg\{\int_0^{2\pi}\Big[\mathcal{T}_n\Big(r(\phi),r(\theta),\mathcal{B}\Big)+\mathcal{T}_{n-2}\Big(r(\phi),r(\theta),\mathcal{B}\Big)\Big]\times \\ &\qquad\qquad r(\phi)\cos\left[(n-1)(\theta-\phi)\right]d\phi \\ &\quad + \int_0^{2\pi}\Big[\mathcal{T}_n\Big(r(\phi),r(\theta),\mathcal{B}\Big)-\mathcal{T}_{n-2}\Big(r(\phi),r(\theta),\mathcal{B}\Big)\Big]\times \\ &\qquad\qquad r'(\phi)\sin\left[(n-1)(\theta-\phi)\right]d\phi\Bigg\}, \end{aligned} \tag{5.16}$$

$$\begin{aligned} v_d(\mathbf{x},t) &= \frac{1}{2\pi}\int_0^{2\pi}\mathcal{T}_1\Big(r(\phi),r(\theta),\mathcal{B}\Big)r'(\phi)d\phi \\ &\quad + \frac{1}{2\pi}\sum_{n=2}^{\infty}\Bigg\{\int_0^{2\pi}\Big[\mathcal{T}_{n-2}\Big(r(\phi),r(\theta),\mathcal{B}\Big)+\mathcal{T}_n\Big(r(\phi),r(\theta),\mathcal{B}\Big)\Big]\times \\ &\qquad\qquad r'(\phi)\cos\left[(n-1)(\theta-\phi)\right]d\phi \\ &\quad + \int_0^{2\pi}\Big[\mathcal{T}_{n-2}\Big(r(\phi),r(\theta),\mathcal{B}\Big)-\mathcal{T}_n\Big(r(\phi),r(\theta),\mathcal{B}\Big)\Big]\times \\ &\qquad\qquad r(\phi)\sin\left[(n-1)(\theta-\phi)\right]d\phi\Bigg\}. \end{aligned} \tag{5.17}$$

Therefore, the velocity at any point $\mathbf{x}$ in the domain can be derived by evaluating the integrals (5.16) and (5.17) around the vortex boundary.

5.2.3 Velocities resulting from infinitesimal weakly nonlinear perturbations of a circular vortex

The integral expressions for the azimuthal and radial velocities given in (5.16) and (5.17) respectively are now used to derive the velocity field due to weakly nonlinear perturbations of a circular vortex patch of unit radius with unit interior vorticity. Without loss of generality the vortex patch is taken to be centred at the origin and the distance between the origin and a general point $\mathbf{x}_1 = r(\phi)\hat{\mathbf{r}}_1$ on the vortex boundary is given by

$$r(\phi) \;=\; 1 + \epsilon\alpha_1 \exp\{ik\phi\} + \epsilon^2\alpha_2 \exp\{2ik\phi\} + \epsilon^3\alpha_3 \exp\{3ik\phi\}, \qquad (5.18)$$

where $\epsilon \ll 1$, α_i are complex coefficients and it is implied that the real part of $r(\phi)$ is taken. Note that in the dimensional system the vortex has radius a_0, the length scale used when performing the nondimensionalization.

Consider a reference point in the domain $\mathbf{x} = r\hat{\mathbf{r}}$. It is necessary to demand that $\mathbf{x}$ be wholly inside or outside the perturbed vortex boundary ∂D, such that r can be taken to be either $r < r(\phi)$ or $r > r(\phi)$ for every $0 \leq \phi < 2\pi$ in the integrals of (5.16) and (5.17). This is shown in schematic 5.1 where any r lying in the region enclosed by the two dashed lines can be seen to be either inside or outside of the vortex boundary depending on the angle ϕ. Choosing r lying wholly inside or outside the vortex boundary, the velocity at any point $\mathbf{x}$ in the domain can be expressed as a weakly nonlinear asymptotic expansion of the form

$$\begin{aligned} u_d &= u_0 + \epsilon u_1 + \epsilon^2 u_2, \\ v_d &= \epsilon v_1 + \epsilon^2 v_2 + \epsilon^3 v_3. \end{aligned}$$

Evaluating the integrals in (5.16) and (5.17) when $r < r(\phi)$ gives

$$u_0 = I_1(\mathcal{B}r)K_1(\mathcal{B}), \tag{5.19}$$

$$u_1 = -\frac{\mathcal{B}\alpha_1}{2}K_k(\mathcal{B})\Big(I_{k-1}(\mathcal{B}r) + I_{k+1}(\mathcal{B}r)\Big)\exp\{ik\theta\}, \tag{5.20}$$

$$\begin{aligned} u_2 = & -\frac{\mathcal{B}\alpha_2}{2}K_{2k}(\mathcal{B})\Big(I_{2k-1}(\mathcal{B}r) + I_{2k+1}(\mathcal{B}r)\Big)\exp\{2ik\theta\} \\ & + \frac{\mathcal{B}^2\alpha_1^2}{32k}\Big(I_{2k+1}(\mathcal{B}r) + I_{2k-1}(\mathcal{B}r)\Big) \times \\ & \Big(\,(2k+1)\,K_{2k-1}(\mathcal{B}) + (2k-1)\,K_{2k+1}(\mathcal{B})\Big)\exp\{2ik\theta\} \\ & + \frac{\mathcal{B}|\alpha_1|^2}{4}I_1(\mathcal{B}r)\Big(\mathcal{B}K_1(\mathcal{B}) - K_0(\mathcal{B})\Big), \end{aligned} \tag{5.21}$$

and

$$v_1 = \frac{ik\alpha_1}{r}I_k(\mathcal{B}r)K_k(\mathcal{B})\exp\{ik\theta\}, \tag{5.22}$$

$$\begin{aligned} v_2 = & -\frac{i}{8r}\Big[\mathcal{B}\alpha_1^2 I_{2k}(\mathcal{B}r)\left((2k-1)\,K_{2k+1}(\mathcal{B}) + (2k+1)\,K_{2k-1}(\mathcal{B})\right) \\ & - 16k\alpha_2 I_{2k}(\mathcal{B}r)K_{2k}(\mathcal{B})\Big]\exp\{2ik\theta\}, \end{aligned} \tag{5.23}$$

$$v_3 = -\frac{i\mathcal{B}\bar{\alpha}_1\alpha_2}{4r}I_k(\mathcal{B}r)\Big[\,(k-1)\,K_{k+1}(\mathcal{B}) + (k+1)\,K_{k-1}(\mathcal{B})\Big]\exp\{ik\theta\}$$

$$- \frac{ik|\alpha_1|^2\alpha_1}{16r}\Big[\mathcal{B}I_k(\mathcal{B}r)\Big(K_{k-1}(\mathcal{B}) + K_{k+1}(\mathcal{B})\Big)$$

$$- 2\left(k^2 + \mathcal{B}^2\right) I_k(\mathcal{B}r)K_k(\mathcal{B})\Big] \exp\{ik\theta\}$$

$$- \frac{i\mathcal{B}\alpha_1\alpha_2}{4r} I_{3k}(\mathcal{B}r)\Big[\left(3k-1\right) K_{3k+1}(\mathcal{B}) + \left(3k+1\right) K_{3k-1}(\mathcal{B})\Big] \exp\{3ik\theta\}$$

$$- \frac{ik\alpha_1^3}{16r}\Big[\mathcal{B}I_{3k}(\mathcal{B}r)\Big(K_{3k-1}(\mathcal{B}) + K_{3k+1}(\mathcal{B})\Big)$$

$$- 2\left((3k)^2 + \mathcal{B}^2\right) I_{3k}(\mathcal{B}r)K_{3k}(\mathcal{B})\Big] \exp\{3ik\theta\}$$

$$+ \frac{3ik\alpha_3}{r} I_{3k}(\mathcal{B}r)K_{3k}(\mathcal{B}) \exp\{3ik\theta\}. \tag{5.24}$$

Similarly, taking $r > r(\phi)$ gives

$$u_0 = K_1(\mathcal{B}r)I_1(\mathcal{B}), \tag{5.25}$$

$$u_1 = \frac{\mathcal{B}\alpha_1}{2} I_k(\mathcal{B})\Big(K_{k-1}(\mathcal{B}r) + K_{k+1}(\mathcal{B}r)\Big) \exp\{ik\theta\}, \tag{5.26}$$

$$u_2 = \frac{\mathcal{B}\alpha_2}{2} I_{2k}(\mathcal{B})\Big(K_{2k-1}(\mathcal{B}r) + K_{2k+1}(\mathcal{B}r)\Big) \exp\{2ik\theta\}$$

$$+ \frac{\mathcal{B}^2\alpha_1^2}{32k}\Big(K_{2k+1}(\mathcal{B}r) + K_{2k-1}(\mathcal{B}r)\Big) \times$$

$$\Big(\left(2k+1\right) I_{2k-1}(\mathcal{B}) + \left(2k-1\right) I_{2k+1}(\mathcal{B})\Big) \exp\{2ik\theta\}$$

$$+\frac{\mathcal{B}|\alpha_1|^2}{4}K_1(\mathcal{B}r)\Big(\mathcal{B}I_1(\mathcal{B})+I_0(\mathcal{B})\Big), \tag{5.27}$$

and

$$v_1 = \frac{ik\alpha_1}{r}K_k(\mathcal{B}r)I_k(\mathcal{B})\exp\{ik\theta\}, \tag{5.28}$$

$$\begin{aligned} v_2 = \frac{i}{8r}\Big[&\mathcal{B}\alpha_1^2K_{2k}(\mathcal{B}r)\left((2k-1)\,I_{2k+1}(\mathcal{B})+(2k+1)\,I_{2k-1}(\mathcal{B})\right)\\ &+16k\alpha_2K_{2k}(\mathcal{B}r)I_{2k}(\mathcal{B})\Big]\exp\{2ik\theta\}, \end{aligned} \tag{5.29}$$

$$\begin{aligned} v_3 = {}& \frac{i\mathcal{B}\bar{\alpha}_1\alpha_2}{4r}K_k(\mathcal{B}r)\Big[(k-1)\,I_{k+1}(\mathcal{B})+(k+1)\,I_{k-1}(\mathcal{B})\Big]\exp\{ik\theta\}\\ &+\frac{ik|\alpha_1|^2\alpha_1}{16r}\Big[\mathcal{B}K_k(\mathcal{B}r)\Big(I_{k-1}(\mathcal{B})+I_{k+1}(\mathcal{B})\Big)\\ &+2\left(k^2+\mathcal{B}^2\right)K_k(\mathcal{B}r)I_k(\mathcal{B})\Big]\exp\{ik\theta\}\\ &+\frac{i\mathcal{B}\alpha_1\alpha_2}{4r}K_{3k}(\mathcal{B}r)\Big[(3k-1)\,I_{3k+1}(\mathcal{B})+(3k+1)\,I_{3k-1}(\mathcal{B})\Big]\exp\{3ik\theta\}\\ &+\frac{ik\alpha_1^3}{16r}\Big[\mathcal{B}K_{3k}(\mathcal{B}r)\Big(I_{3k-1}(\mathcal{B})+I_{3k+1}(\mathcal{B})\Big)\\ &+2\left((3k)^2+\mathcal{B}^2\right)K_{3k}(\mathcal{B}r)I_{3k}(\mathcal{B})\Big]\exp\{3ik\theta\} \end{aligned}$$

$$+ \frac{3ik\alpha_3}{r} K_{3k}(\mathcal{B}r) I_{3k}(\mathcal{B}) \exp\{3ik\theta\}. \tag{5.30}$$

Once again, it is implied that the real part is taken in all velocity components given in (5.19)-(5.30).

Comparisons of the weakly nonlinear velocity calculations with the actual velocity observed in a fully nonlinear shallow water model are shown in Figs. 5.2-5.4. In each of these figures, the top six panels show the coefficients $u^{(i)}$ and $v^{(i)}$ of the ith harmonic in the discrete Fourier transform of u_d and v_d

$$u_d = u^{(0)} + u^{(1)} \exp\{ik\phi\} + u^{(2)} \exp\{2ik\phi\} + \ldots, \tag{5.31}$$

$$v_d = v^{(1)} \exp\{ik\phi\} + v^{(2)} \exp\{2ik\phi\} + v^{(3)} \exp\{3ik\phi\} + \ldots, \tag{5.32}$$

with the weakly nonlinear predictions denoted by solid lines and the fully nonlinear model results by diamonds and dashed lines. The bottom two panels show contours of the azimuthal (bottom left) and radial (bottom right) velocity fields as a function of distance from the origin r and azimuthal angle ϕ.

Fig. 5.2 shows the velocity components in (5.31) and (5.32) when $\mathcal{B} = 0$ and $\epsilon = 0.15$, corresponding to the purely two-dimensional perturbed vortex problem discussed by Su (1979). For the harmonic coefficients $u^{(i)}$ and $v^{(i)}$ and full velocity fields u_d and v_d, results are only shown for points which are wholly inside or outside the vortex patch. For the full velocity fields u_d and v_d, the region between the heavy solid lines denotes the region between the dashed lines in Fig. 5.1. There is excellent agreement between the Fourier coefficients $u^{(i)}$ and $v^{(i)}$ and the total azimuthal and radial velocity fields u_d and v_d for the weakly nonlinear theory and the fully nonlinear

numerical model. This agreement is also seen when $\mathcal{B} = 1.0$ in Fig. 5.3 and when $\mathcal{B} = 3.0$ and $\epsilon = 0.10$ in Fig. 5.4.

The velocity expansions given above are formally valid for all values of $\mathcal{B}$ as $\epsilon \to 0$. However, when considering finite amplitude disturbances, for the higher order terms in ϵ to remain small in the truncated expansions for u_d and v_d, we require that $\epsilon\mathcal{B} \ll 1$. The lower value of $\epsilon = 0.10$ used in the calculations of Fig. 5.4 when $\mathcal{B} = 3$ ensures that $\epsilon\mathcal{B} \ll 1$, such that the weakly nonlinear theory is expected to give a good approximation to the actual velocity.

5.3 Weakly nonlinear dispersion relation for infinitesimal disturbances

In this section, we derive a dispersion relation for weakly nonlinear wave-like perturbations to the edge of an unforced circular vortex. Considering disturbances of the form given in (5.18), translation of the wave around the vortex edge is introduced using a first harmonic wave frequency ω such that

$$r(\phi, t) \;=\; 1 + \eta(\phi, t),$$

where

$$\begin{aligned} \eta(\phi, t) \;=\; & \epsilon\alpha_1 \exp\{ik\,(\phi - \omega t)\} \\ & + \epsilon^2\alpha_2 \exp\{2ik\,(\phi - \omega t)\} \\ & + \epsilon^3\alpha_3 \exp\{3ik\,(\phi - \omega t)\}, \end{aligned} \tag{5.33}$$

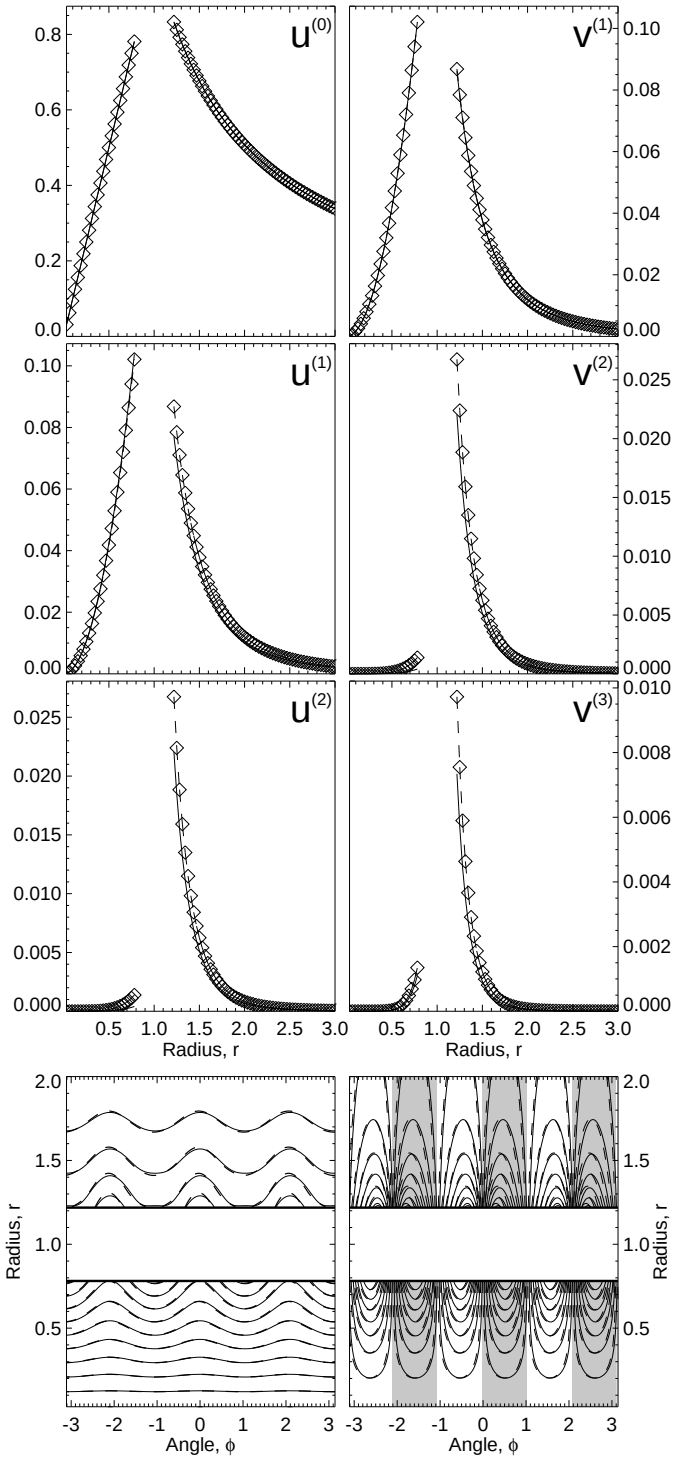

Figure 5.2: Velocity comparisons for $\mathcal{B} = 0$, $k = 3$, $\epsilon = 0.15$, $\alpha_1 = 1.13$, $\alpha_2 = 1.67$ and $\alpha_3 = 0.83$. Top Panels: coefficients $u^{(i)}$ and $v^{(i)}$ of the ith harmonic in the Fourier decomposition of the azimuthal and radial velocities for $i = 1, 2, 3$. Weakly nonlinear predictions are given by solid lines, fully nonlinear results by diamonds and dashed lines. Bottom panels: contours of the azimuthal and radial velocity fields u_d and v_d as a function of azimuthal and radial coordinates (ϕ, r) for the weakly nonlinear theory (solid lines) and fully nonlinear numerical model (dashed lines). Contours are shown at intervals of 0.05 for u_d and 0.01 for v_d. Shaded regions in the contour plot for v_d denote regions in which the radial velocity is negative.

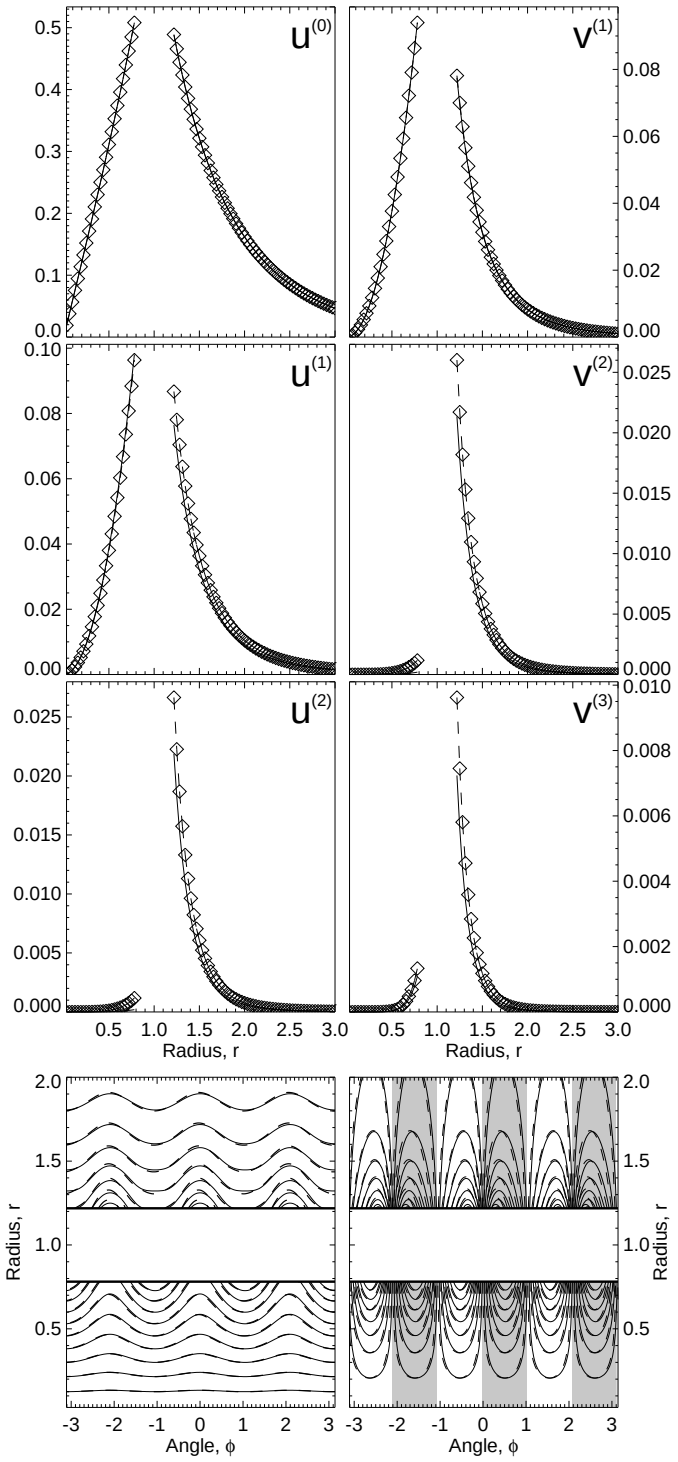

Figure 5.3: As 5.2 but for $\mathcal{B} = 1.0$.

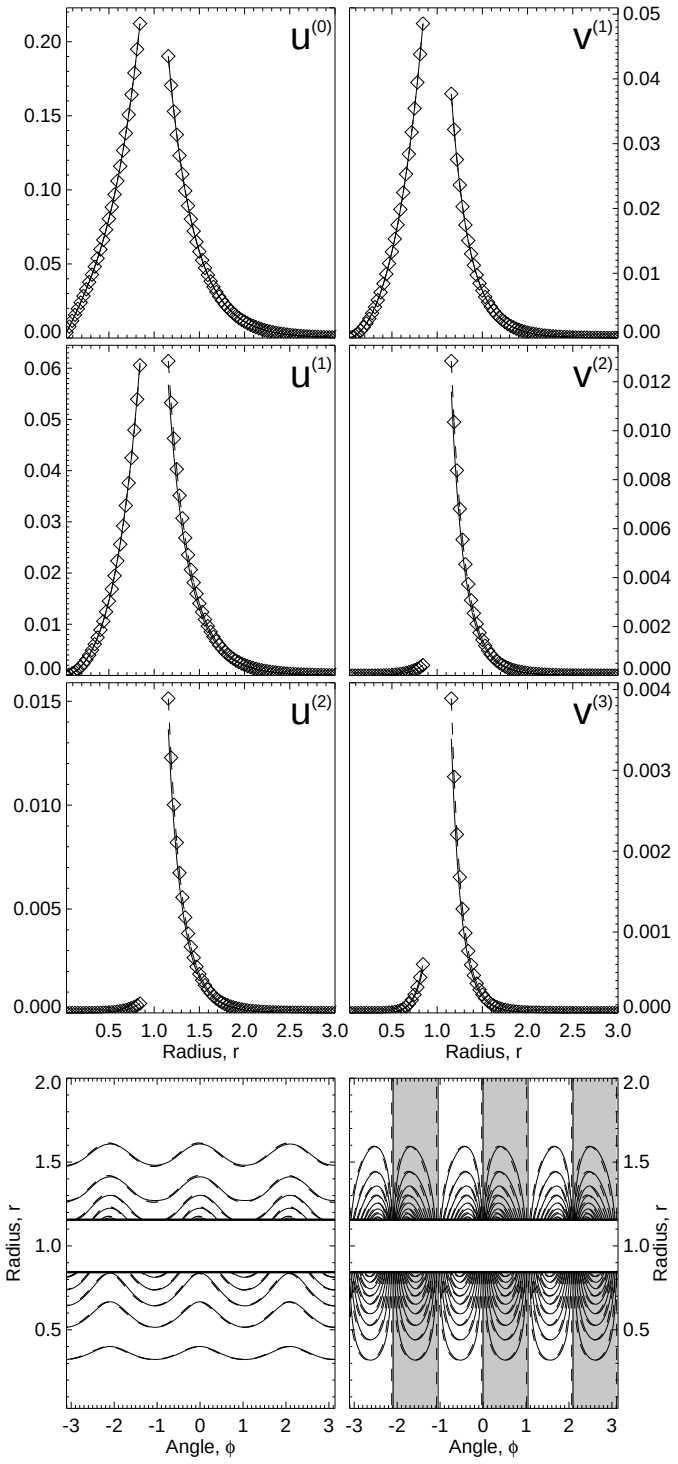

Figure 5.4: As 5.2 but for $\mathcal{B} = 3.0$ and $\epsilon = 0.10$. Contours are shown at intervals of 0.005 for v_d.

with α_i being complex amplitudes and with the real part implied. It is straightforward to show that the weakly nonlinear u_d and v_d velocity fields corresponding to this disturbance are identical to those calculated for (5.18) but with exponential terms modified to include the linear wave frequency as seen above. The frequency is now expressed as an expansion in terms of the small parameter ϵ, that is $\omega = \omega_0 + \epsilon^2\omega_2$.

In addition to the velocities (u_d, v_d) due to the disturbed vortex patch, the velocity due to a solid body rotation of the flow corresponding to the streamfunction ψ_b satisfying equation (5.4) is

$$u_b \quad = \quad 2\Omega_b I_1(\mathcal{B}r)K_1(\mathcal{B}). \tag{5.34}$$

In a similar vein to the first harmonic wave frequency, it is convenient to express Ω_b as an expansion in powers of ϵ using $\Omega_b = \Omega_b^{(0)} + \epsilon^2\Omega_b^{(2)}$.

The boundary of the vortex acts as a material boundary for the flow such that the fully nonlinear kinematic condition

$$\frac{\partial \eta}{\partial t} + \frac{u_d + u_b}{r}\frac{\partial \eta}{\partial \phi} \quad = \quad v_d \quad \text{on} \quad r = 1 + \eta(\phi, t), \tag{5.35}$$

is satisfied on the vortex edge (note that the absence of topographic forcing means $(u_T, v_T) = 0$). Using the weakly nonlinear expansions for u_d and v_d from section 5.2.3 on the vortex edge, along with u_b, the kinematic condition on the vortex edge can be expressed as a weakly nonlinear expansion in powers ϵ.

There is, however, one small complication in using the kinematic condition on the vortex edge. As discussed in section 5.2.3, the weakly nonlinear analytic forms for the azimuthal and radial velocities u_d and v_d are only strictly valid for reference points either wholly inside or outside the vortex patch, as seen in Fig. 5.1. Therefore, when

using these weakly nonlinear expansions for u_d and v_d to calculate velocities on the vortex edge, questions must be asked as to whether the inner or outer velocity representations are used, and even then there may be concerns given that neither expansion is strictly valid. This consideration was omitted in the work of Su (1979), who used the inner velocity when deriving the dispersion relation for a perturbed vortex in a purely two-dimensional vortex model, equivalent to setting $\mathcal{B} = 0$ in our representation. However, although detailed calculations are not shown here, it is found that the dispersion relation is independent of the choice of either inner or outer velocity when representing the actual velocity on the vortex edge in (5.35). Furthermore, Figs. A.1-A.3 in Appendix A, testing the results of the weakly nonlinear dispersion relation against a fully nonlinear numerical model, confirm that the frequencies derived using either the inner or outer velocities agree with results of the fully nonlinear numerical simulations. A mathematical justification for this interesting property has, as yet, proved elusive, and has therefore been left to the realm of future study. However, as Figs. A.1, A.2 and A.3 indicate, we can be confident that the resulting weakly nonlinear frequencies are correct.

Due to the independence of the final dispersion relation on the choice of either the inner or outer velocity expansions for u_d and v_d in (5.35), it is assumed that when referring to the velocity at the vortex edge a choice is made between the two possible expansions, and that choice is carried throughout the entire calculation.

Following substitution of the expansions for u_d, v_d and u_b into (5.35) and expanding modified Bessel functions as Taylor series in the vicinity of $\mathcal{B}$, equating powers of ϵ and $\exp\{ik(\phi - i\omega t)\}$ gives the second harmonic amplitude and first harmonic wave

frequency as

$$\alpha_2 = \frac{4k\mathcal{B}I_1K_2 + \mathcal{B}K_{2k}\mathcal{C}_{2k} - 4kI_kK_k - 4kI_k\mathcal{B}\mathcal{R}_k - 8k\mathcal{B}\Omega_b^{(0)}I_2K_1}{16k\left(\mathcal{P}_k - \mathcal{P}_{2k}\right)}\alpha_1^2, \quad (5.36)$$

$$\omega_0 = I_1K_1 - I_kK_k + 2\Omega_b^{(0)}I_1K_1, \quad (5.37)$$

$$\begin{aligned}\omega_2 = {} & \frac{\mathcal{B}|\alpha_1|^2}{4}\Big[-2I_1K_2 - K_k\mathcal{S}_k + I_k\mathcal{R}_k + I_{2k}\mathcal{R}_{2k} + 4\Omega_b^{(0)}I_2K_1\Big]\left(\frac{\alpha_2}{\alpha_1^2}\right) \\ & + \Big[8\mathcal{B}I_0K_1 + 11\mathcal{B}^2I_1K_1 + \mathcal{B}I_1K_3 \\ & - 2\mathcal{B}K_kS_k - 16(k^2+\mathcal{B}^2)I_kK_k - 6\mathcal{B}I_k\mathcal{R}_k \\ & + \frac{\mathcal{B}^2}{2k}\mathcal{C}_{2k}\mathcal{R}_{2k} + 6\mathcal{B}^2\Omega_b^{(0)}I_1K_1 + 2\mathcal{B}^2\Omega_b^{(0)}I_3K_1\Big]\frac{|\alpha_1|^2}{32} \\ & + 2\Omega_b^{(2)}I_1K_1, \end{aligned} \quad (5.38)$$

where all modified Bessel functions have been evaluated at $\mathcal{B}$ with the following functions defined in the interests of brevity

$$\begin{aligned}\mathcal{C}_k &= (k+1)I_{k-1} + (k-1)I_{k+1}, \\ \mathcal{D}_k &= (k+1)K_{k+1} + (k-1)K_{k-1}, \\ \mathcal{S}_k &= I_{k-1} + I_{k+1}, \\ \mathcal{R}_k &= K_{k-1} + K_{k+1}, \\ \mathcal{P}_k &= K_kI_k.\end{aligned}$$

Verification of these results with a fully nonlinear numerical model can be found in appendix A. Also, the linear frequency in equation (5.37) is seen to agree with that already calculated in section 4.4.2 of chapter 4.

It is now useful to consider disturbances which appear stationary in the fixed frame. By setting $\omega_0 = \omega_2 = 0$ in (5.37) and (5.38), this gives the background angular velocity components which are required for stationary waves as

$$\Omega_b^{(0)} = \Omega_0(k,\mathcal{B}) = \frac{I_k K_k - I_1 K_1}{2 I_1 K_1}, \tag{5.39}$$

$$\begin{aligned}\Omega_b^{(2)} = \Omega_2(k,\mathcal{B}) = |\alpha_1|^2 \Bigg(8 \Bigg[& I_1^2 \mathcal{R}_1 - 2 I_2 \mathcal{P}_{2k} \\ & + I_1 \Big(\mathcal{S}_1 K_1 + \mathcal{S}_k K_k - \mathcal{R}_k I_k - \mathcal{R}_{2k} I_{2k} \Big) \Bigg] \left(\frac{\alpha_2}{\alpha_1^2} \right) \\ & \Bigg[\mathcal{B}^2 I_1 \left(I_1 K_3 - K_1 I_3 \right) - 8 \mathcal{B} \mathcal{P}_1 \left(I_0 + \mathcal{B} I_1 \right) - \mathcal{B}^2 \mathcal{P}_k I_3 \\ & + I_1 \Bigg(\left(16 k^2 + 13 \mathcal{B}^2 \right) \mathcal{P}_{2k} + 6 \mathcal{B} \mathcal{R}_k I_k + 2 \mathcal{B} \mathcal{S}_k K_k \\ & + \mathcal{B} \mathcal{R}_{2k} \left(\mathcal{R}_{2k} + 2 I_{2k} \right) \Bigg) \Bigg] \Bigg) \Bigg/ \Big(I_1 \mathcal{P}_1 \Big). \end{aligned} \tag{5.40}$$

Considering the asymptotic forms of the modified Bessel functions I_1 and K_1, as given in equations (4.43) and (4.44), it is straightforward yet lengthy to show that when $\mathcal{B} = 0$, the wave amplitudes and frequency components of the stationary

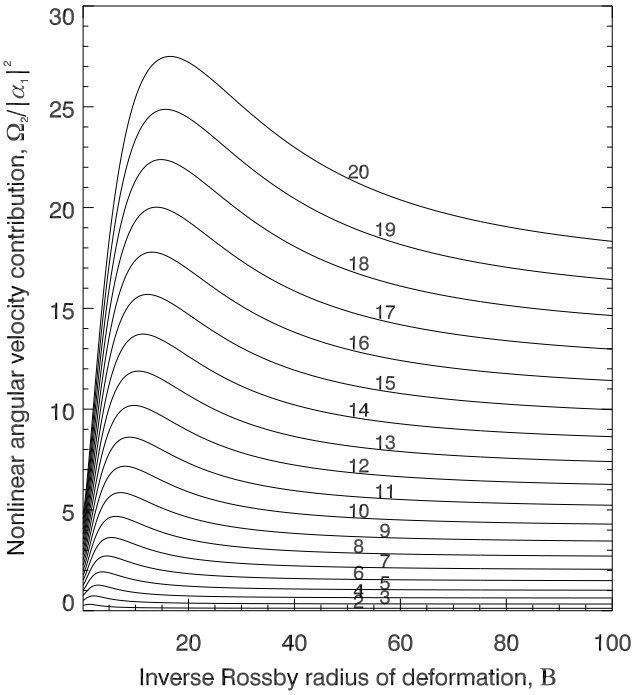

Figure 5.5: Nonlinear component of the background solid body rotation giving stationary disturbances, $\Omega_2(k,\mathcal{B})/|\alpha_1|^2$, as a function of inverse Rossby radius $\mathcal{B}$ shown for the first 20 azimuthal wavenumbers $k \geq 2$.

disturbances agree with those of Su (1979), that is

$$\alpha_2(k,0) = \frac{2k-1}{4}\alpha_1^2,$$

$$\Omega_0(k,0) = -\frac{(k-1)}{2k},$$

$$\Omega_2(k,0) = \frac{(k-1)}{4}|\alpha_1|^2.$$

One important consideration to the work which follows is whether Ω_2 takes both positive and negative values as the parameters k and $\mathcal{B}$ are varied. From Fig. (5.5), it is seen that for any choice of k and $\mathcal{B}$, $\Omega_2(k,\mathcal{B})$ is positive. As a result, the background

rotation leading to stationary disturbances on the vortex edge, $\Omega_b = \Omega_0 + \epsilon^2\Omega_2$, will always be greater than Ω_0 for any choice of k or $\mathcal{B}$.

5.4 Wave amplification

5.4.1 Linear resonance

In order to understand and appreciate the role of nonlinearity in determining the resonant behaviour of a circular vortex subject to topographic forcing, it is first helpful to investigate a linear model of the response to forcing, analogous to that of Tung and Lindzen (1979a). We now introduce a topographic forcing

$$h_T(r,\phi) \;=\; \mathcal{H}J_k(\lambda r)\exp\{ik\phi\}, \tag{5.41}$$

with corresponding streamfunction ψ_T which is found by inverting 5.5 giving

$$\psi_T(r,\phi) \;=\; -\frac{\mathcal{H}}{\lambda^2+\mathcal{B}^2}J_k(\lambda r)\exp\{ik\phi\}, \tag{5.42}$$

where it is implied that the real part has been taken. Therefore, utilizing the relationship between the streamfunction and the velocity field $\mathbf{u}_T = -\nabla\psi_T \times \mathbf{k}$, the azimuthal and radial velocities induced in the flow by the topography are

$$u_T(r,\phi) \;=\; -\frac{\mathcal{H}\lambda}{\lambda^2+\mathcal{B}^2}J_k'(\lambda r)\exp\{ik\phi\}, \tag{5.43}$$

$$v_T(r,\phi) \;=\; \frac{ik\mathcal{H}}{\lambda^2+\mathcal{B}^2}J_k(\lambda r)\exp\{ik\phi\}. \tag{5.44}$$

We now consider a system in which an initially circular vortex in a solid body rotating background flow is subject to topographic forcing of the form (5.41) such

that the forcing is stationary with respect to the fixed frame of the domain. In the linear treatment, topographic forcing of the form (5.41) will excite waves on the vortex edge of the form

$$\eta(\phi,t) \;=\; \epsilon\alpha_1(t)\exp\{ik(\phi-\omega t)\}, \tag{5.45}$$

where the linear wave amplitude α_1 is allowed to vary as a function of time t.[1] Resonant excitation of this disturbance occurs when there is no translation of the disturbance relative to the fixed topographic forcing. Therefore, we now set $\omega = 0$ in (5.45). To investigate linear resonance, we now specify that the topography enter the analysis at $O(\epsilon)$ by setting

$$\mathcal{H} \;=\; \epsilon\mathcal{H}_0 \tag{5.46}$$

In order to study the growth of linear waves on the vortex edge, it is necessary to construct an amplitude equation for α_1 describing the evolution of α_1 with time t. As the vortex boundary $r = 1+\eta(\phi,t)$ acts as a material boundary in the flow, linearizing the kinematic condition (5.35) on the vortex edge gives

$$\frac{\partial\eta}{\partial t} + \frac{(u_d+u_b)}{r}\frac{\partial\eta}{\partial\phi} \;=\; v_d + v_T \quad \text{on} \quad r=1. \tag{5.47}$$

Here, the azimuthal and radial velocities

$$u_d \;=\; I_1K_1, \tag{5.48}$$

[1]Formally, the nondimensional wave amplitude and topography should consist of a single nondimensional parameter, rather than ϵ *and* α_1 or $\mathcal{H}$. However, to enable the results of section 5.3 to be used directly, the form given here has been chosen for convenience.

and

$$v_d \;=\; ik\alpha_1(t) I_k K_k \exp\{ik(\phi - \omega t)\},$$

are the velocities due to the disturbed vortex patch in the absence of topographic forcing and rotating background flow, as given by the expansions in section 5.2.3 with α_1 replaced everywhere by $\alpha_1(t)$. The velocity $u_b = 2\Omega_b I_1 K_1$ is the linear contribution to the background solid body rotation (5.34), with angular velocity Ω_b acting as a tunable parameter of the system. All modified Bessel functions are evaluated at $\mathcal{B}$.

Solution of the linearized kinematic condition yields an ODE for $\alpha_1(t)$

$$\frac{d\alpha_1}{dt} + iq\alpha_1 \;=\; if,$$

where

$$q \;=\; 2k\Big[\Omega_b - \Omega_0(k,\mathcal{B})\Big] I_1 K_1, \tag{5.49}$$

and

$$f \;=\; \frac{k\mathcal{H}_0}{\lambda^2 + \mathcal{B}^2} J_k(\lambda). \tag{5.50}$$

The function $\Omega_0(k,\mathcal{B})$ is as defined in equation (5.39), and takes a single scalar value for a particular choice of k and $\mathcal{B}$ (i.e. when referring to Ω_0 it is assumed to be the scalar value corresponding to a particular choice of k and $\mathcal{B}$).

Solution of this ODE for $\alpha_1(t)$ is straightforward using an integrating factor of $\exp\{iqt\}$ giving

$$\alpha_1(t) \;=\; \frac{f}{q}\left(1 - e^{-iqt}\right).$$

It is clear from looking at the solution for $\alpha_1(t)$ that the system becomes resonant when the value of the tuning parameter is $\Omega_b = \Omega_0$ for which $q = 0$. When $\Omega_b \neq \Omega_0$, the maximum disturbance amplitude $\max|\alpha_1|$ is found from

$$\begin{aligned} |\alpha_1| &= \left|\frac{f}{q}\right| |1 - (\cos qt - i \sin qt)| , \\ &= \left|\frac{f}{q}\right| \sqrt{2 - 2\cos qt}, \end{aligned}$$

giving

$$\max|\alpha_1| = \left|\frac{2f}{q}\right| .$$

By writing $A = \epsilon\alpha_1$ and $F = \epsilon f$, the maximum disturbance of the first harmonic is related to the physical forcing amplitude of the system by

$$A_{max} = \left|\frac{2F}{q}\right| . \tag{5.51}$$

Contours of the maximum linear amplitude A_{max} as a function of the topographic forcing F and background flow tuning parameter Ω_b are shown in panel C of Fig. 5.6 for $\mathcal{B} = 0$ and $k = 2$. The most striking feature is that A_{max} is symmetric, becoming infinite when $\Omega_b = \Omega_0 = -0.25$, that is, linear resonance theory predicts that for off resonant Ω_b, the maximum amplitude of linear disturbances at the vortex edge will be the same irrespective of which side of resonance Ω_b lies.

5.4.2 Nonlinear resonance

In the previous section, it was shown that resonant excitation of waves on the vortex edge in the linearized problem occurs when the tuning parameter Ω_b is equal

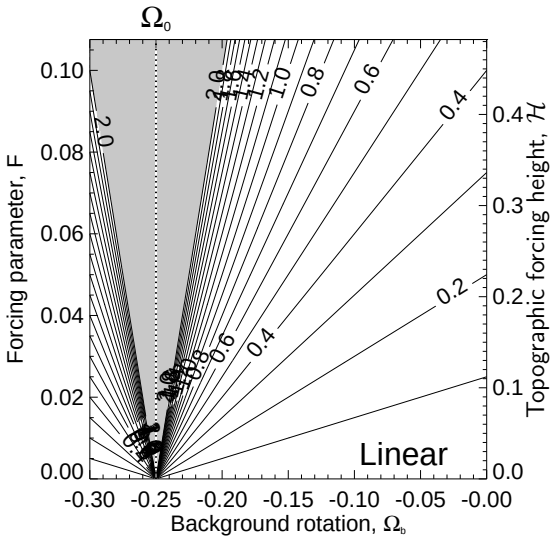

Figure 5.6: Contours of $\max|\alpha_1|$ as a function of topographic forcing f and tuning parameter Ω_b when $\mathcal{B} = 0$ and $k = 2$. Regions where $\max|\alpha_1| > 2$ are shaded.

to $\Omega_0(k, \mathcal{B})$, the linear component of the background angular velocity for which stationary disturbances are observed on the vortex edge.

Resonance of waves on the vortex edge is characterized by a significant increase in the linear wave amplitude α_1. When α_1 is large, it is to be expected that nonlinear interactions will become increasingly important. As a result, linear predictions of resonant wave growth predict the breakdown of the very conditions for which the linear theory is assumed to be valid. Therefore, in order to gain a more detailed view of the mechanisms governing resonant wave growth, it is necessary to expand upon the existing linear theory to investigate the impact of nonlinearity in the system. The treatment here is similar to that of Plumb (1981), with the notable difference being that we consider only a single layer model and that the velocities required in the analysis have already been derived in section 5.2.3 and do not therefore need to be derived as part of the nonlinear analysis.

We start by considering a system with topography of the form (5.6) with topographic streamfunction satisfying (5.5). The resulting azimuthal and radial velocities are given by (5.43) and (5.43).

It is now useful to define a 'long time' variable in the problem by letting $T = \epsilon^2 t$ such that $T \sim 1$ when $t \sim \epsilon^{-2}$. In a similar way to the linear problem, Rossby waves excited on the vortex edge by the topography can be expected to exhibit the azimuthal wavenumber k structure taken in the topography. In addition, the amplitude of all excited waves are taken to be functions of the long time variable, that is $\alpha_i(T)$. Therefore, disturbances to the edge of an initially circular vortex which is subject to topographic forcing of the form (5.6) are given by

$$\begin{aligned} \eta(\phi, t, T) &= \epsilon\alpha_1(T)\exp\{ik(\phi - \omega t)\} \\ &\quad + \epsilon^2\alpha_2(T)\exp\{2ik(\phi - \omega t)\} \\ &\quad + \epsilon^3\alpha_3(T)\exp\{3ik(\phi - \omega t)\}. \end{aligned} \tag{5.52}$$

Once again, as we are looking for conditions leading to resonance of η with the topographic forcing, which occurs when the disturbances do not translate relative to the fixed frame, it is now necessary to set $\omega = 0$ in (5.52). Therefore, the only time-dependence in η is a dependence on the long time variable, T. In addition, as we are interested in the nonlinear effects of topography, we now express the topographic forcing height as

$$\mathcal{H} = \epsilon^3\mathcal{H}_0.$$

Introducing a solid body background rotation to the system with streamfunction governed by equation (5.4) and angular velocity Ω_b, which adds an azimuthal velocity

$u_b = 2\Omega_b I_1(\mathcal{B}r)K_1(\mathcal{B})$ to the flow, the fully nonlinear kinematic condition on the vortex boundary becomes

$$\epsilon^2\frac{\partial\eta}{\partial T} + \frac{(u_d + u_b + u_T)}{r}\frac{\partial\eta}{\partial\phi} = v_d + v_T \quad \text{on} \quad r = 1 + \eta(\phi, T). \tag{5.53}$$

In (5.53), azimuthal and radial velocities u_d and v_d are the velocities induced by the disturbed vortex in the absence of the topographic forcing and background rotation. These are therefore just the velocity expansions given in section 5.2.3, with coefficients α_i everywhere replaced by $\alpha_i(T)$. We also adopt the approach of section 5.3 and express the angular velocity Ω_b as a power series in ϵ

$$\Omega_b = \Omega_b^{(0)} + \epsilon^2\Omega_b^{(2)}.$$

It can immediately be seen that the radial velocity due to the topographic forcing will appear at $O(\epsilon^3)$ whereas the azimuthal velocity due to the topographic forcing will appear at $O(\epsilon^4)$ due to the fact that $\partial\eta/\partial\phi$ is at most $O(\epsilon)$.

By expanding the kinematic condition (5.53) in powers of ϵ, to leading order $O(\epsilon)$ we get

$$\Omega_b^{(0)} = \Omega_0(k, \mathcal{B}),$$

where $\Omega_0(k, \mathcal{B})$ is given in equation (5.39). Looking to $O(\epsilon^2)$ simply re-derives the expression for $\alpha_2(T)$ in terms of $\alpha_1(T)$ in exactly the same way as shown in equation (5.36). The long time dependence of $\eta(\phi, T)$ and the topographic forcing finally enter the analysis at $O(\epsilon^3)$ giving

$$\frac{d\alpha_1}{dT} + \Big(\Omega_b^{(2)} - \Omega_2(k, \mathcal{B})\Big)ik\alpha_1 = \frac{ik\mathcal{H}_0 J_k(\lambda)}{\lambda^2 + \mathcal{B}^2}, \tag{5.54}$$

where the function $\Omega_2(k, \mathcal{B})$ is given in equation (5.40). Equation (5.54) is now rewritten as

$$\frac{d\alpha_1}{dT} + iq\alpha_1 - i\sigma(k, \mathcal{B})|\alpha_1|^2\alpha_1 = if, \tag{5.55}$$

where

$$\begin{aligned} q &= \Omega_b^{(2)} k, \\ &= \epsilon^{-2}\Big[\Omega_b - \Omega_0(k, \mathcal{B})\Big]k, \\ \sigma(k, \mathcal{B}) &= \frac{k\Omega_2(k, \mathcal{B})}{|\alpha_1|^2}, \\ f &= \frac{k\mathcal{H}_0 J_k(\lambda)}{\lambda^2 + \mathcal{B}^2}. \end{aligned}$$

The nonlinear ODE in (5.55) is analogous to the forced amplitude equation derived in Plumb (1981, equation 4.25), and describes a forced nonlinear oscillator. To determine the long time behaviour of $\alpha_1(T)$, it is now necessary to tackle (5.55). First, it is useful to re-scale (5.55) by taking $A = (f/\sigma)^{-1/3}\alpha_1$ and $\tau = (f^2\sigma)^{1/3}T$ giving

$$\frac{dA}{d\tau} + i\mu A - i|A|^2 A = i, \tag{5.56}$$

where $\mu = q/(f^2\sigma)^{1/3}$. Equation (5.56) is an example of a forced weakly nonlinear oscillator, so by expressing the complex amplitude as $A = a(\tau)\exp\{i\theta(\tau)\}$, where $a, \theta \in \mathbb{R}$, the nonlinear oscillator equation can be rewritten as

$$\left[\frac{da}{d\tau} + ia\frac{d\theta}{d\tau}\right]e^{i\theta} + i\mu a e^{i\theta} + ia^3 e^{i\theta} = i.$$

Multiplying by $\exp\{-i\theta\}$ and then equating the real and imaginary parts gives

$$\frac{da}{d\tau} = \sin\theta, \tag{5.57}$$

$$\frac{d\theta}{d\tau} = \frac{\cos\theta}{a} - \mu + a^2, \tag{5.58}$$

which can be combined to give

$$\frac{d\theta}{da} = \frac{\cos\theta + a(a^2-\mu)}{a\sin\theta}.$$

In order to integrate this ODE, we make a change of variables using $x = \cos\theta$ giving

$$\frac{dx}{da} + \frac{x}{a} = \mu - a^2,$$

which can be shown using the integrating factor method to have general solution

$$x = \frac{\mu a}{2} - \frac{a^3}{4} + \frac{C}{a},$$

where C is a constant of the integration. As $da/d\tau = \sin\theta$ from (5.57), $da/d\tau$ and x are related through $(da/d\tau)^2 + x^2 = 1$ giving

$$\left(\frac{da}{d\tau}\right)^2 + \left[\frac{\mu a}{2} - \frac{a^3}{4} + \frac{C}{a}\right]^2 - 1 = 0. \tag{5.59}$$

If the vortex is initially circular, then $\alpha_1 = 0$ when $t = 0$ in the original system, giving $a = 0$ and $\tau = 0$ in the rescaled system. Therefore, for the C/a term in (5.59) to be well behaved when $\tau = 0$, the constant of integration must be $C = 0$ and

$$\left(\frac{da}{d\tau}\right)^2 + V(a;\mu) = 0, \tag{5.60}$$

where

$$V(a;\mu) = a^2\left[\frac{\mu}{2} - \frac{a^2}{4}\right]^2 - 1. \tag{5.61}$$

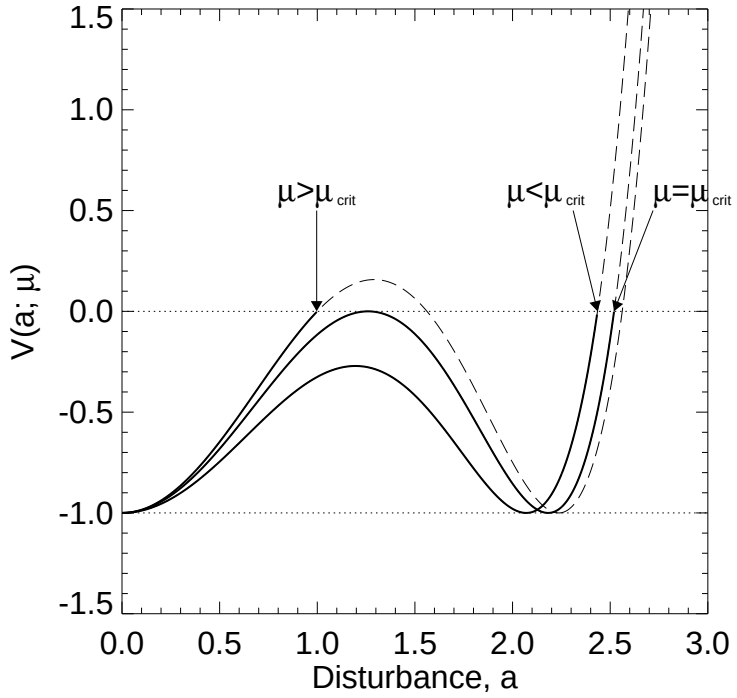

Figure 5.7: Plots of the potential function $V(a;\mu)$ for the weakly nonlinear oscillator as a function of a for $\mu < \mu_{crit}$, $\mu = \mu_{crit}$ and $\mu > \mu_{crit}$. In each case, the heavy solid line denotes the portion of the $V(a;\mu)$ curve corresponding to permitted vortex evolution.

For a given value of μ, the extremal values of a over the vortex evolution occur when $da/d\tau = 0$, which is equivalent to finding the roots of the function $V(a;\mu)$. This is similar to the potential function derived in section 4.3.3 of chapter 4 for the Kida vortex.

Plots of $V(a;\mu)$ as a function of a, for three sample values of μ, are shown in Fig. 5.7. The first thing to notice from the relationship in (5.60) is that in order for $da/d\tau$ to take real values, the function $V(a;\mu)$ must be negative. So when analyzing the behaviour of $V(a;\mu)$, it is necessary to restrict our attention to the branches of $V(a;\mu)$ lying in the negative half plane. As $a = 0$ when $\tau = 0$, it becomes apparent

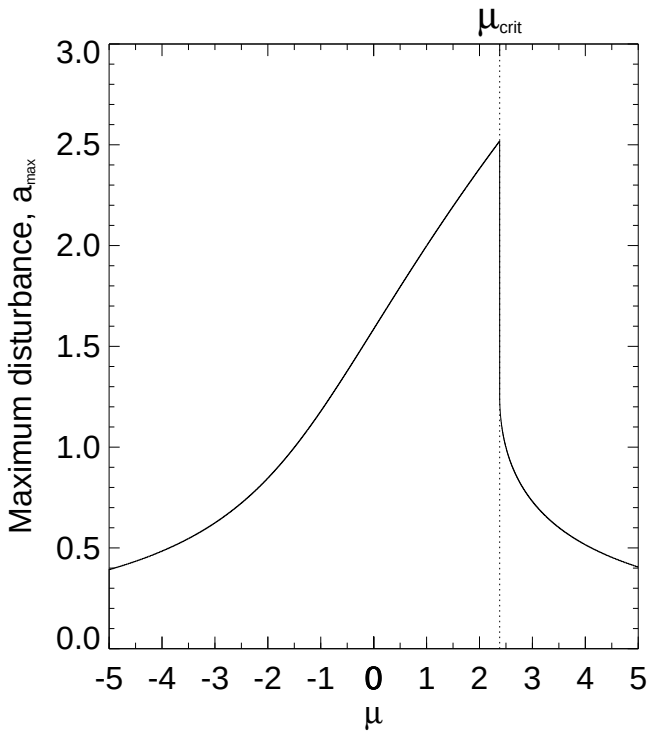

Figure 5.8: The maximum disturbance a_{max} as a function of μ. The dotted line denotes the location of the critical value $\mu = \mu_{crit}$.

that the branch of $V(a;\mu)$ determining the evolution of the vortex for $\tau > 0$ is the branch of $V(a;\mu)$ lying between $a = 0$ and the point a_{max}, the smallest positive root of $V(a;\mu)$. For each value of μ in Fig. 5.7, this branch of $V(a;\mu)$ is denoted by a heavy solid line.

Fig. 5.8 shows the value a_{max} varying as a function of μ. It is immediately obvious that there is a discontinuity in a_{max} at $\mu = \mu_{crit}$. For $\mu < \mu_{crit}$, a_{max} increases monotonically, reaching a peak when $\mu = \mu_{crit}$. There is then an instantaneous drop in a_{max} to much smaller values, with a_{max} then decreasing as $\mu > \mu_{crit}$ increases further.

An explanation for this discontinuity in a_{max} can be deduced from Fig. 5.7. For

$\mu < \mu_{crit}$, very close to the critical value μ_{crit}, it is seen that a local maximum in $V(a;\mu)$ exists in the negative half-plane. When $\mu = \mu_{crit}$, this peak just touches the a-axis, such that the local maximum is also a root of $V(a;\mu)$. For both of the curves $\mu < \mu_{crit}$ and $\mu = \mu_{crit}$, the maximum in a over the evolution is seen to occur at $a \approx 2.4$. However, as μ increases past μ_{crit}, the local maximum crosses the a-axis giving a sharp transition to a situation where the smallest root of $V(a;\mu)$ now occurs at $a \approx 1.0$. By differentiating $V(a;\mu)$ in (5.61) with respect to a, stationary points are found when

$$a = \sqrt{2\mu} \qquad \text{or} \qquad a = \sqrt{\frac{2\mu}{3}}.$$

For real values of a, this implies that stationary points only occur when $\mu > 0$. Finding the value of $V(a;\mu)$ at these stationary points gives

$$V(\sqrt{2\mu}) = -1 \qquad \text{and} \qquad V(\sqrt{2\mu/3}) = \frac{2\mu^3}{27} - 1.$$

Therefore, a local minimum occurs when $a = \sqrt{2\mu}$, and a local maximum when $a = \sqrt{2\mu/3}$. If the local maximum is also a root when the parameter $\mu = \mu_{crit}$ then

$$V(\sqrt{2\mu_{crit}/3}) \;=\; \frac{2\mu_{crit}^3}{27} - 1 = 0,$$

giving

$$\mu_{crit} \;=\; \frac{3}{\sqrt[3]{2}} \approx 2.38. \tag{5.62}$$

So when $\mu < \mu_{crit}$, the function $V(a;\mu)$ has one real root at a_{max}. When $\mu > \mu_{crit}$, the function $V(a;\mu)$ has three real roots, with a_{max} defined as being the smallest of the three. By rewriting $V(a;\mu)$ as

$$V(a;\mu) \;=\; \frac{1}{16}\left(a^6 - 4\mu a^4 + 4\mu^2 a^2 - 16\right), \tag{5.63}$$

and letting $y = a^2$, the roots are then given by solving the cubic equation

$$g(y) \quad = y^3 - 4\mu y^2 + 4\mu^2 y - 16 \quad = 0. \tag{5.64}$$

When $\mu < \mu_{crit}$ Cardano's formula gives the only real root of (5.64) as

$$y \quad = \quad \frac{4\mu}{3} + \frac{2}{3}\left[\mu^2\left(27 - \mu^3 + 3\sqrt{3}\sqrt{27 - 2\mu^3}\right)^{-1/3} + \left(27 - \mu^3 + 3\sqrt{3}\sqrt{27 - 2\mu^3}\right)^{1/3}\right].$$

When $\mu > \mu_{crit}$, the smallest root of $g(y)$ is found as

$$y \quad = \quad \frac{4\mu}{3} - \frac{4\mu}{3}\cos\left[\frac{1}{3}\arccos\left\{\frac{\mu^3 - 27}{\mu^3}\right\}\right].$$

So the smallest root of $V(a;\mu)$, at which the disturbance amplitude a is a maximum, is

$$a_{max} \quad = \quad \sqrt{\frac{4\mu}{3} + \frac{2}{3}\left[\mu^2\left(27 - \mu^3 + 3\sqrt{3}\sqrt{27 - 2\mu^3}\right)^{-1/3} + \left(27 - \mu^3 + 3\sqrt{3}\sqrt{27 - 2\mu^3}\right)^{1/3}\right]},$$

$$a_{max} \quad = \quad \sqrt{\frac{4\mu}{3} - \frac{4\mu}{3}\cos\left[\frac{1}{3}\arccos\left\{\frac{\mu^3 - 27}{\mu^3}\right\}\right]},$$

when $\mu < \mu_{crit}$ and $\mu > \mu_{crit}$ respectively. The maximum amplitude of the first harmonic of the disturbance in the original system is then given by

$$\max[\epsilon\alpha_1] \quad = \quad \epsilon\left(\frac{f}{\sigma}\right)^{1/3} a_{max},$$

$$= \quad \left(\frac{F}{\sigma}\right)^{1/3} a_{max} \tag{5.65}$$

where $F = \epsilon^3 f$ is the rescaled height of the topographic forcing in the nondimensionalized system.

5.5 Results

The weakly nonlinear theory of section 5.4 is now used to quantify the impact of self-tuning resonance on vortex-splitting in a quasi-geostrophic single layer f-plane model.

5.5.1 Topographic forcing and vortex conditions

The PV in the nondimensional fully nonlinear experiments is governed by equations (5.3)-(5.5). All experiments are initiated with an initially circular vortex of unit radius and unit interior vorticity, subject to topographic forcing of the form given in (5.6). Of the five nondimensional forcing parameters $(\mathcal{B}, \Omega_b, \mathcal{H}, k, \lambda)$, three sets of experiments were performed with parameter choices:

SWM-1:	$\mathcal{B} = 0$	$k = 2$	$\lambda = 1.162$
SWM-2:	$\mathcal{B} = 1.44$	$k = 2$	$\lambda = 1.162$
SWM-3:	$\mathcal{B} = 0$	$k = 3$	$\lambda = 2.575$

For each set of experiments, the background rotation tuning parameter Ω_b and the topographic height $\mathcal{H}$ were varied between model runs.

When presenting the results for each set of experiments, the forcing amplitude is given in terms of

$$F \;=\; \frac{k\mathcal{H}J_k(\lambda)}{\lambda^2 + \mathcal{B}^2}.$$

For SWM-1 experiments, F is equal to the straining parameter Λ in the Kida vortex model, as seen in section 4.3 of chapter 4.

In all experiments, vortex moment diagnostics are calculated as described in section 4.5.3 of chapter 4. Finite amplitude disturbances in the numerical experiments are calculated as described in section 4.5.4 or chapter 3.

5.5.2 Model details and numerical parameters

The contour dynamics numerical method of Dritschel (1988) was used for all experiments (see chapter 3 for an overview of the contour dynamics numerical algorithm).

In all experiments, the node resolution parameter used in the contour dynamics method of Dritschel (1988) is $\mu = 0.025$, the time interval is $dt = 0.05$, and the maximum time of the experiment is $t_{max} = 40$. Data are generated at time intervals of 0.25 nondimensional time units, giving a total of 160 time steps of data.

All vortex moment diagnostics are calculated as given in section 4.5.3 of chapter 4. For the κ_4 vortex split diagnostic, the split threshold parameter is $\mu_4 = -0.6$, such that if $\kappa_{min} = \min[\kappa_4] < -0.6$ a vortex split is identified (see chapter 2.2.4). When interpreting the κ_4 diagnostic, high positive values indicate that the vortex is undergoing a process of filamentation.

5.5.3 SWM-1 experiments

Experiments were performed for Ω_b between -0.30 and 0 at increments of 0.01. Values of the forcing parameter which are used in the experiments are partitioned into two sets. The first consists of four experiments representing small F values, $F = [0.00178, 0.00357, 0.00535, 0.00714]$. The second set starts at $F = 0.00892$ and includes 12 experiments, going up to $F = 0.10706$ in increments of 0.00892.

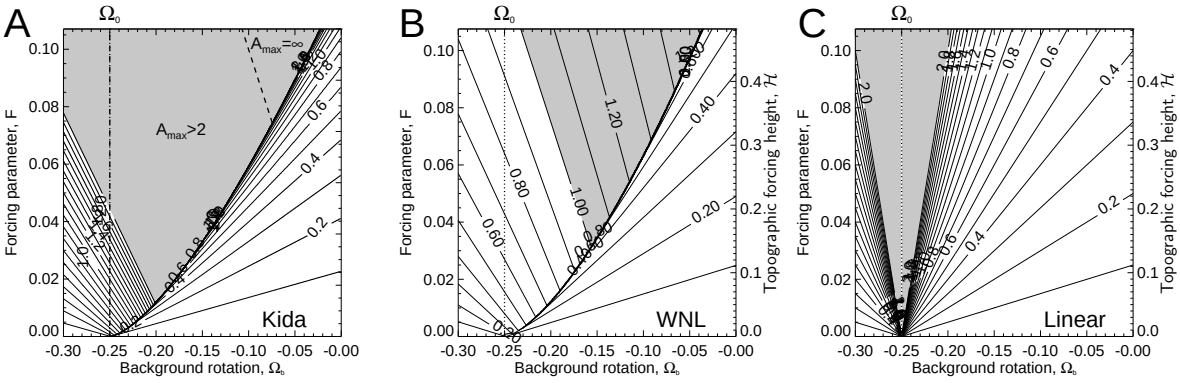

Figure 5.9: Panel A: contours of the maximum disturbance amplitude A_{max} as a function of forcing height F and background rotation Ω_b for the Kida model. Panel B: as panel A but for the weakly nonlinear model with $k = 2$ and $\mathcal{B} = 0$. Panel C: as panel A but for the linear model with $k = 2$ and $\mathcal{B} = 0$. In all panels, dashed lines denote the location of the regime boundaries in the Kida model and dotted lines denote the location of the linear component of the background rotation Ω_0. The shaded region is where $A_{max} > 2$ in Panels A and C and where $A_{max} > 1$ in Panel B. Contours are shown at intervals of 0.1.

Kida, weakly nonlinear and linear model comparisons

Predictions for the maximum disturbance of an initially circular vortex in the Kida model and linear and weakly nonlinear topographically forced models (referred to herein as the linear model and weakly nonlinear model in the interests of brevity) are now discussed. The maximum amplitude A_{max} as a function of parameters F and Ω_b is shown in Fig. 5.9 for the Kida theory (panel A), the weakly nonlinear theory (panel B) and the linear theory (panel C).

It is clear that the discontinuity in A_{max} in the weakly nonlinear model closely follows the curved regime boundary of the Kida model, as seen in panel B. That is, for a given forcing height F, very small changes in the background rotation Ω_b across the discontinuity result in sudden increase (or decrease) in the maximum amplitude A_{max}.

This is in stark contrast to the linear theory in which the peak value of A_{max} occurs at $\Omega_b = \Omega_0$ for all forcing amplitudes. Also, the peak in A_{max} is characterized by a steady increase in A_{max} as one nears $\Omega_b = \Omega_0$ rather than a sudden jump in A_{max}. Therefore, as the forcing amplitude increases, the linear model does not reveal anything about the migration of the peak response in A_{max} away from Ω_0. Also, at any given forcing height, the peak A_{max} in the Kida and weakly nonlinear models occurs at the line of discontinuity, and is finite. This is once again in contrast to the linear model in which the peak value of A_{max} is infinite for all forcing amplitudes.

For disturbances which appear stationary with respect to the topography, the tuning parameter is $\Omega_b = \Omega_0 + \epsilon^2\Omega_2$, where $\Omega_2 > 0$ as shown in Fig. 5.5. Therefore, as resonance occurs when the wave-like disturbance is stationary with respect to the fixed frame, resonance is expected to occur when $\Omega_b = \Omega_0 + \epsilon^2\Omega_2 > \Omega_0$. This accounts for the rightward migration of the location at which the peak response occurs in panel B of Fig. 5.9, and explains why resonance, as indicated by a jump in A_{max}, does not occur at all in parameter space where $\Omega_b < \Omega_0$.

The A_{max} predictions in the region of parameter space to the right of the curved regime boundary are seen to agree reasonably well between all three models, with significant differences only appearing in the Kida model at large values of F. In the region to the left of the curved Kida regime boundary, all three models are seen to differ somewhat. Although the qualitative behaviour in A_{max} between the Kida and weakly nonlinear models agrees, there is a discrepancy in magnitude, with values in the Kida model exceeding those of the weakly nonlinear model. In this region the linear model is in poor agreement with the other two models due to the fact that the

location of the peak in A_{max} does not migrate from $\Omega_b = \Omega_0$. Another point worth noting is that the regime boundary between the oscillating and clockwise rotating regimes in the Kida model has no impact on the behaviour of A_{max}. That is, without prior knowledge of this qualitative change in the rotating nature of the vortex, it would be impossible to identify by looking at the behaviour of A_{max} alone.

From this comparison in the behaviour of A_{max} between the three models it is clear that the predictions of the weakly nonlinear model when $k = 2$ and $\mathcal{B} = 0$ accurately reproduces the major features of the fully nonlinear Kida model. The linear theory, while capturing some of the features of the other models at very small forcing amplitudes, is subject to two significant deficiencies; the location of the peak in A_{max} is fixed at $\Omega_b = \Omega_0$, and the peak in A_{max} is infinite. Therefore, it is anticipated that the A_{max} predictions of the weakly nonlinear topographically forced model will closely resemble those of the fully nonlinear topographically forced model for any given value of k and $\mathcal{B}$. This prediction is put to the test in the following section.

Weakly nonlinear and fully nonlinear model comparisons

Comparisons can now be drawn between predictions of the maximum disturbance amplitude A_{max} in the weakly nonlinear model with the fully nonlinear numerical model results of chapter 4.

In chapter 4 it was shown that the wave-activity of an initially circular vortex subject to topographic forcing of the form (5.6) with $k = 2$ and $\mathcal{B} = 0$ could be predicted by investigating the behaviour of the curved regime boundary in the Kida vortex model. In particular, for sufficiently large forcing amplitudes, vortex-splitting

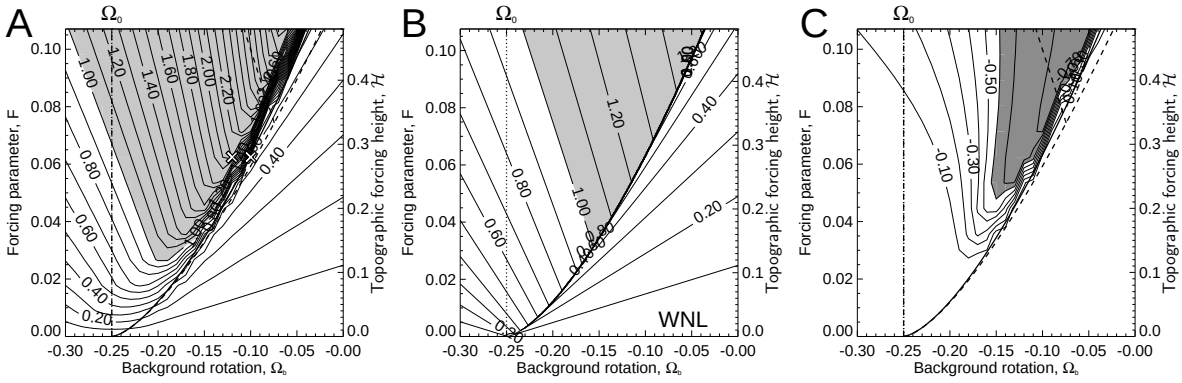

Figure 5.10: Panel A: contours of the maximum disturbance amplitude A_{max} as a function of forcing height F and background rotation Ω_b for SWM-1 experiments. Contours are shown at intervals of 0.1. The shaded region is where $A_{max} > 1$. The experiments shown in Fig. 5.11 are marked by crosses. Panel B: as panel A but for the weakly nonlinear model. Panel C: contours of κ_{min} for the fully nonlinear numerical model. Contours are shown at intervals of 0.1. The shaded region is where $\kappa_{min} < -0.6$ indicating that a vortex-splitting event has occurred. In all panels, the dashed line marks the regime boundaries in the Kida model, the dotted line where $\Omega_b = \Omega_0$, and the solid line is the line of discontinuity in the weakly nonlinear theory.

events were seen to lie in the unstable Kida regimes to the left of the curved regime boundary in panel A of Fig. 5.9.

Fig. 5.10 gives a comparison of the maximum disturbance amplitude A_{max} in the fully nonlinear model (panel A), A_{max} in the weakly nonlinear model (panel B), and κ_{min} in the fully nonlinear numerical model (panel C), for the SWM-1 set of experiments. In panels A and B, parameter space in which $A_{max} \geq 1$, for which large deformation of the vortex occurs, is shaded. The shading in panel C represents regions in which $\kappa_{min} < -0.6$ and a vortex-splitting event has been identified.

The line of discontinuity in A_{max} in the weakly nonlinear model is excellently reproduced in the fully nonlinear model results. In parameter space to the right of the line of discontinuity, the behaviour of A_{max} agrees between the the weakly

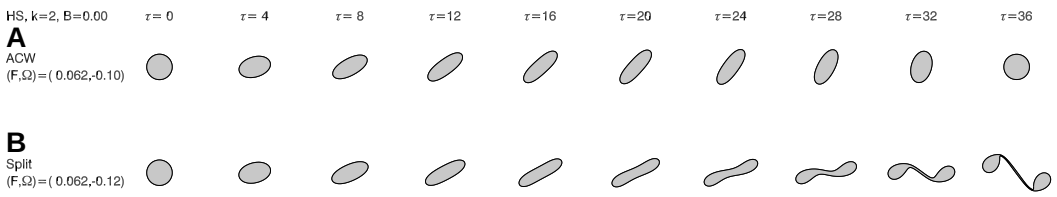

Figure 5.11: Snapshots of vortex evolution in the fully nonlinear numerical model in SWM-1 experiments. Panel A: evolution classified as anticlockwise rotation when $F = 0.062$ and $\Omega_b = -0.10$. Panel B: evolution classified as a vortex-splitting event when with $F = 0.062$ and $\Omega_b = -0.12$.

nonlinear and fully nonlinear models. To the left of the line of discontinuity there is still a high degree of agreement between the two models, with A_{max} magnitudes in the fully nonlinear model slightly exceeding those predicted by the weakly nonlinear theory. The location in parameter space of vortex-splitting events in panel C is seen to lie entirely within the shaded region in panel B, with the right hand boundary closely following the line of discontinuity. That is, prediction of large A_{max} values in the weakly nonlinear model is a harbinger of vortex-splitting events in the fully nonlinear model.

The existence of a sudden transition in behaviour due to very small changes in Ω_b (across the line of discontinuity) is shown by Fig. 5.11. These numerical experiments are marked on panel A of Fig. 5.10 by white bordered crosses on either side of the line of discontinuity. At a given forcing height $F = 0.062$, a small change in Ω_b from 0.12 to 0.10 results in a transition of the vortex from a stable anticlockwise rotating state to an unstable vortex-splitting event (see Fig. 4.6 in chapter 4, and discussion in the text, for the classification of these experiments). As the parameter Ω_b crosses the line of discontinuity from right to left, the maximum disturbance amplitude A_{max} jumps from small values to much larger values. It has been shown by Dritschel (1986)

that for large aspect elliptic vortices in the absence of background flow, instability of elliptic azimuthal wavenumber-2, leading to a complete split of the vortex (see section 4.2). Although our model includes a topographically forced background flow, it may well be the case that a similar process of instability is found here.

The agreement between the fully nonlinear numerical model and the weakly nonlinear model when $k = 2$ and $\mathcal{B} = 0$ is to be expected given the independent agreement between each of these models and the Kida model. In order to ascertain whether this agreement is retained for other values of k and $\mathcal{B}$, we now compare the fully nonlinear numerical model and weakly nonlinear model for the SWM-2 and SWM-3 experiment sets.

5.5.4 SWM-2 experiments

We now discuss the results of the weakly nonlinear theory in relation to the SWM-2 set of experiments in the fully nonlinear model.

Experiments were performed for Ω_b between -0.19 and 0.11 at increments of 0.01. Values of the forcing parameter which are used in the experiments are partitioned into two sets. The first consists of four experiments representing small F values, $F = [0.00178, 0.00357, 0.00535, 0.00714]$. The second set starts at $F = 0.00892$ and includes 12 experiments, going up to $F = 0.10706$ in increments of 0.00892.

Weakly nonlinear and fully nonlinear model comparisons

Fig. 5.12 shows contours of the maximum disturbance amplitude A_{max} as a function of forcing height F and background rotation Ω_b for the fully nonlinear (panel A)

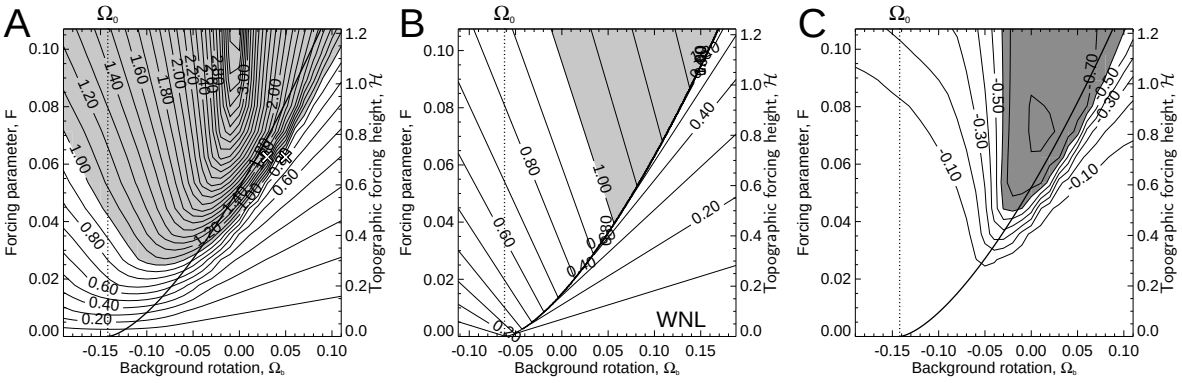

Figure 5.12: As figure 5.10 but for the SWM-2 experiments ($\mathcal{B} = 1.44$).

and weakly nonlinear (panel B) models for SWM-2 experiments.[2]

It is clear from Fig. 5.12 that, similar to the SWM-1 experiments seen in Fig. 5.10, there is good qualitative agreement in A_{max} between panels A and B. Although a slight offset is observed, the location of the line of discontinuity in A_{max} as predicted by the weakly nonlinear theory is reasonably well represented in the fully nonlinear model. A discrepancy does exist however in the magnitude of A_{max} in all regions, with A_{max} values in the fully nonlinear model being approximately double that predicted by the weakly nonlinear theory.

Panel C shows the diagnostic κ_{min} for the fully nonlinear numerical model experiments. In a similar way to the SWM-1 experiments, it is seen that the region of parameter space in which vortex-splitting events occur coincides with the shaded region of panel B in which large amplitude events with $A_{max} > 1$ are predicted. In

[2]The choice of $\mathcal{B} = 1.44$ in the SWM-2 model runs approximately corresponds to the nondimensional Rossby radius for the zero order Lamb mode in the three-dimensional quasi-geostrophic f-plane model with a physical lower boundary condition, as introduced in section 3.2.2 of chapter 3. The importance of the zero order Lamb mode on the dynamics of the stratospheric polar vortex during vortex-splitting events in a three-dimensional model will be the focus of chapter 7.

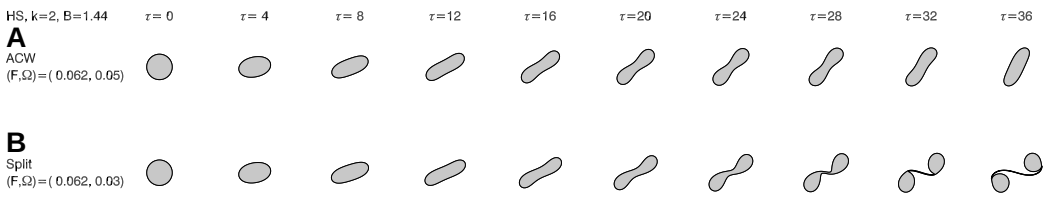

Figure 5.13: Snapshots of vortex evolution in the fully nonlinear numerical model experiments of SWM-2. Panel A: evolution classified as anticlockwise rotation when $F = 0.062$ and $\Omega_b = 0.05$. Panel B: evolution classified as a vortex-splitting event when with $F = 0.062$ and $\Omega_b = 0.03$.

fact, the correlation between the two is higher in this case than for the SWM-1 experiments, with the majority of the shaded region in panel B being identified as vortex splitting events in the numerical model. The primary difference in the behaviour of κ_{min} in these experiments, when compared to the SWM-1 experiments, is that 'leakage' of splitting-like behaviour is observed over the line of discontinuity. That is, larger negative values of κ_{min} are observed in the region to the right of the line of discontinuity in the SWM-2 experiments shown here than in the SWM-1 experiments, although the values are not so large and negative that vortex splits are identified. In overall shape and size, the shaded region in panel C which denotes vortex-splitting experiments is very similar to that seen in panel C of Fig. 5.10, with the region in this case being slightly wider. In particular, there is no discernible change in the minimum physical forcing height at which vortex splits are observed, with splits first occurring when $0.45 < F < 0.50$ in both SWM-1 and SWM-2 experiments.

A transition in vortex behaviour in the vicinity of the line of discontinuity in A_{max} is demonstrated in Fig. 5.13 by snapshots of vortex evolution. Once again, for a fixed forcing height $F = 0.062$, a relatively small change in the background rotation Ω_b from 0.05 to 0.03 results in a transition of the vortex from a stable anticlockwise

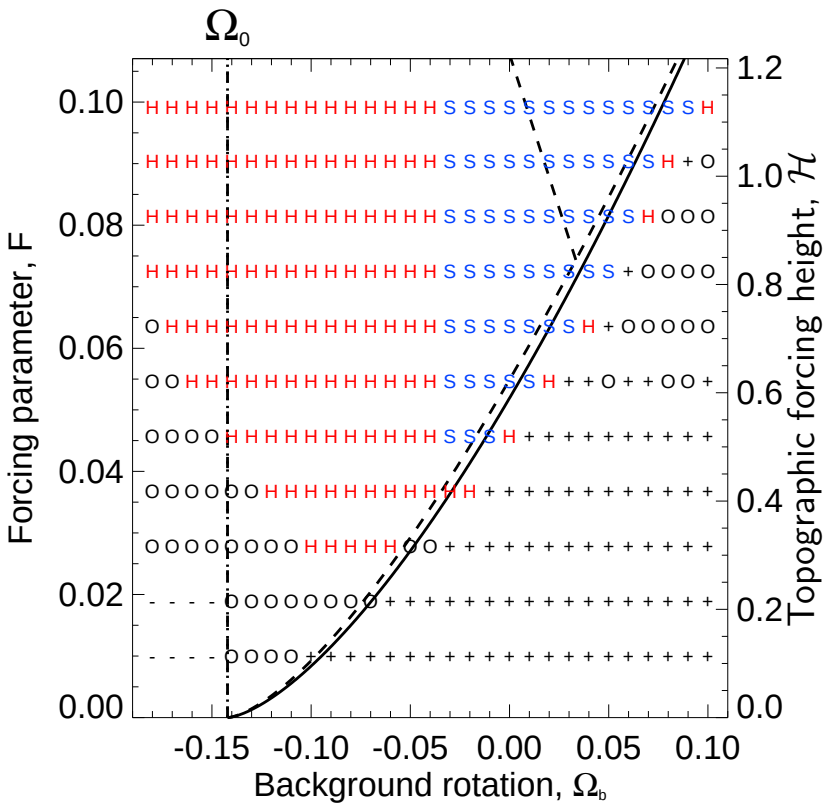

Figure 5.14: Class identification for the fully nonlinear numerical model during SWM-2 experiments. Classification definitions are as described in the text in section 4.5.5 of chapter 4. The solid dark line marks the line of discontinuity in A_{max} as predicted by the weakly nonlinear theory. The dotted line marks the location of $\Omega_b = \Omega_0$.

rotating state to a vortex-splitting state.[3] The existence of larger negative values of κ_{min} can be seen here in the slight pinching of the vortex over the course of its evolution when $(F, \Omega_b) = (0.062, 0.05)$. It should be pointed out however that the vortex evolution remains stable and periodic in spite of this slight perturbation.

A further demonstration of the good agreement in the line of discontinuity between the fully nonlinear and weakly nonlinear models is seen in Fig. 5.14, in which

[3]Vortex classification is as described in section 4.5.5 of chapter 4.

the classification of each numerical experiment is shown in the parameter space. It is clear that the discontinuity in A_{max} predicted by the weakly nonlinear theory gives an excellent indication of the region in which a transition in vortex behaviour occurs. Another interesting feature of Fig, 5.14 is that the vortex classification in the parameter space is similar to that of the Kida vortex regimes (see Fig. 4.2, particularly for small forcing heights). That is, the region of parameter space to the right of the line of discontinuity is mainly characterized by stable anticlockwise rotating vortex states, whereas the region to the left is characterized by over classifications of vortex states, many of which appear to be unstable. Furthermore, clockwise rotating vortex states only exist in the region to the left of the dotted line $\Omega_b = \Omega_0$, similar to the clockwise rotation regime of the Kida model.

It has been shown that the weakly nonlinear theory successfully predicts the discontinuity in the maximum disturbance A_{max} in the SWM-1 and SWM-2 sets of experiments in which $k = 2$. The SWM-3 experiments will now test the predictions of the weakly nonlinear model for azimuthal wavenumber $k = 3$ in the topographic forcing.

5.5.5 SWM-3 experiments

We now discuss the results of the weakly nonlinear theory in relation to the SWM-2 set of experiments in the fully nonlinear model.

Experiments were performed for Ω_b between -0.38 and -0.08 at increments of 0.01. Values of the forcing parameter which are used in the experiments are partitioned into three sets. The first consists of one experiment representing small F values at

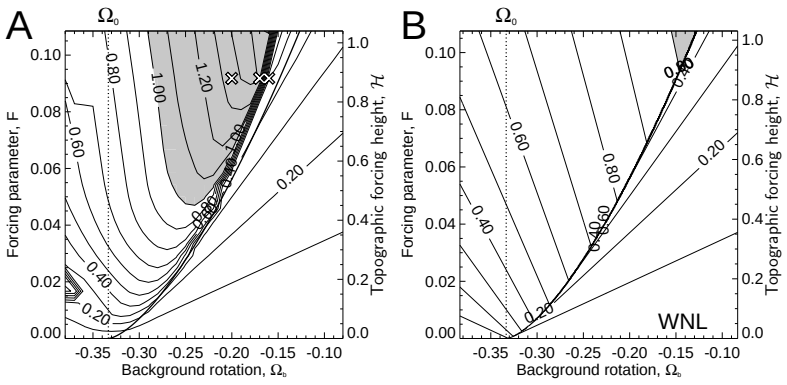

Figure 5.15: Panel A: contours of the maximum disturbance amplitude A_{max} as a function of forcing height F and background rotation Ω_b for the fully nonlinear numerical model SWM-3 experiments. Contours are shown at intervals of 0.1. The shaded region is where $A_{max} > 1$. The experiments shown in Fig. 5.16 are marked by crosses. Panel B: as panel A but for the weakly nonlinear model. The solid dark line marks the line of discontinuity in A_{max} as predicted by the weakly nonlinear theory. The dotted line marks the location of $\Omega_b = \Omega_0$.

$F = 0.0008$. The second set starts at $F = 0.0042$ and includes 14 experiments, going up to $F = 0.0584$ in increments of 0.0041. The third set starts at $F = 0.0668$ and includes 5 experiments, going up to $F = 0.1085$ in increments of 0.0083.

Weakly nonlinear and fully nonlinear model comparisons

Fig. 5.10 shows contours of A_{max} as a function of forcing height F and background rotation tuning parameter Ω_b. Once again, it is seen that the line of discontinuity in A_{max} exists in the fully nonlinear model, and its location is predicted by the weakly nonlinear theory. To the right of the line of discontinuity, the agreement in A_{max} is excellent between the two models. In the region to the left, A_{max} in the numerical model is larger than predicted by the weakly nonlinear model. Large amplitude events where $A_{max} > 1$ occur for forcing heights of approximately $F = 0.05$, rather than $F = 0.10$ as predicted in the weakly nonlinear theory. Even so, the most important

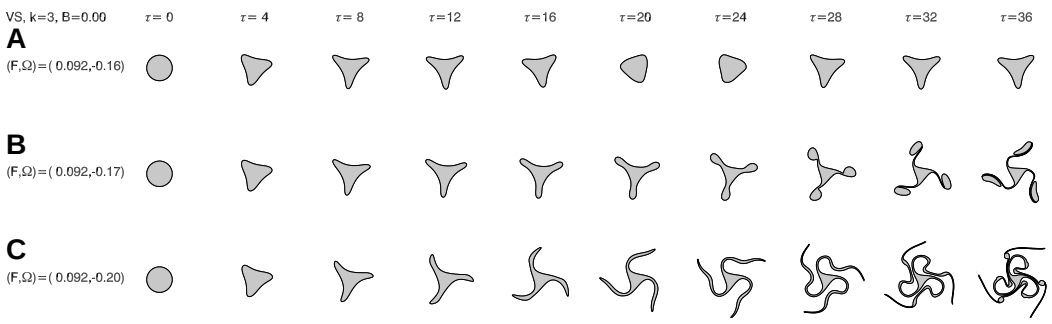

Figure 5.16: Snapshots of vortex evolution in the fully nonlinear numerical model during SWM-3 experiments. Panel A: evolution when $F = 0.092$ and $\Omega_b = 0.16$. Panel B: evolution when $F = 0.092$ and $\Omega_b = 0.17$. Panel C: evolution when $F = 0.092$ and $\Omega_b = 0.20$.

feature, the location of the line of discontinuity in A_{max} characterized by a transition in vortex behaviour, is well predicted by the weakly nonlinear theory.

For a given forcing height of $F = 0.092$, transitions in vortex behaviour for small changes in Ω_b (across the line of discontinuity) are shown in Fig. 5.16. When $\Omega_b = -0.16$, the vortex in panel A exhibits stable periodic motion. With a small change in tuning parameter $\Omega_b = -0.17$, crossing the line of discontinuity, panel B shows that the stable periodic motion is replaced with unstable wave growth on the vortex edge leading to a three-way split of the vortex around a small triangular vortex core. When $\Omega_b = -0.20$ in panel C, still relatively close in terms of parameter space to the experiments in panels A and B, a different type of unstable vortex behaviour is observed. Rather than a three-way split, as seen in panel B, the vortex undergoes intense filamentation which almost entirely destroys the organized vortex structure to leave a diminished triangular vortex core. This highlights the importance of the line of discontinuity in determining vortex behaviour in response to topographic forcing with azimuthal wavenumber $k = 3$, in addition to that already shown when $k = 2$.

5.6 Conclusions

5.6.1 Self-tuning resonance in the weakly nonlinear model

In a weakly nonlinear theory, as the forcing height increases, migration of the tuning parameter Ω_b at which A_{max} occurs, away from $\Omega_b = \Omega_0$ at which resonance occurs in a linear theory, can be thought of as a self-tuning resonance of the system. Resonant behaviour of the first harmonic wave on the vortex edge occurs when the wave becomes stationary with respect to the fixed frame, that is stationary with respect to the topographical forcing. It was shown in section 5.3 that for any choice of k and $\mathcal{B}$, the angular velocity of the background rotation for which wave-like disturbances are stationary is $\Omega_b = \Omega_0 + \epsilon^2\Omega_2 > \Omega_0$. Hence, the behaviour of the weakly nonlinear oscillator as shown in Figs. 5.10, 5.12 and 5.15, in which the line of discontinuity in A_{max} migrates to values $\Omega_b > \Omega_0$, are representative of the weakly nonlinear predictions for any value of k or $\mathcal{B}$. That is, the line of discontinuity, which represents the location of perfect resonance in the parameter space, will never migrate to values $\Omega_b < \Omega_0$.

A visualization of resonant mechanisms in the linear and weakly nonlinear models is presented in schematic form in Fig. 5.17. In all panels, the vortex is represented by perturbations to a circular boundary

$$r \;=\; 1 + A\exp\{ik\phi\},$$

where $k = 3$.

Panel A describes the resonant process in the linear model. The background flow is given by $\Omega_b = \Omega_0$. For resonance growth of the first harmonic wave to occur, the

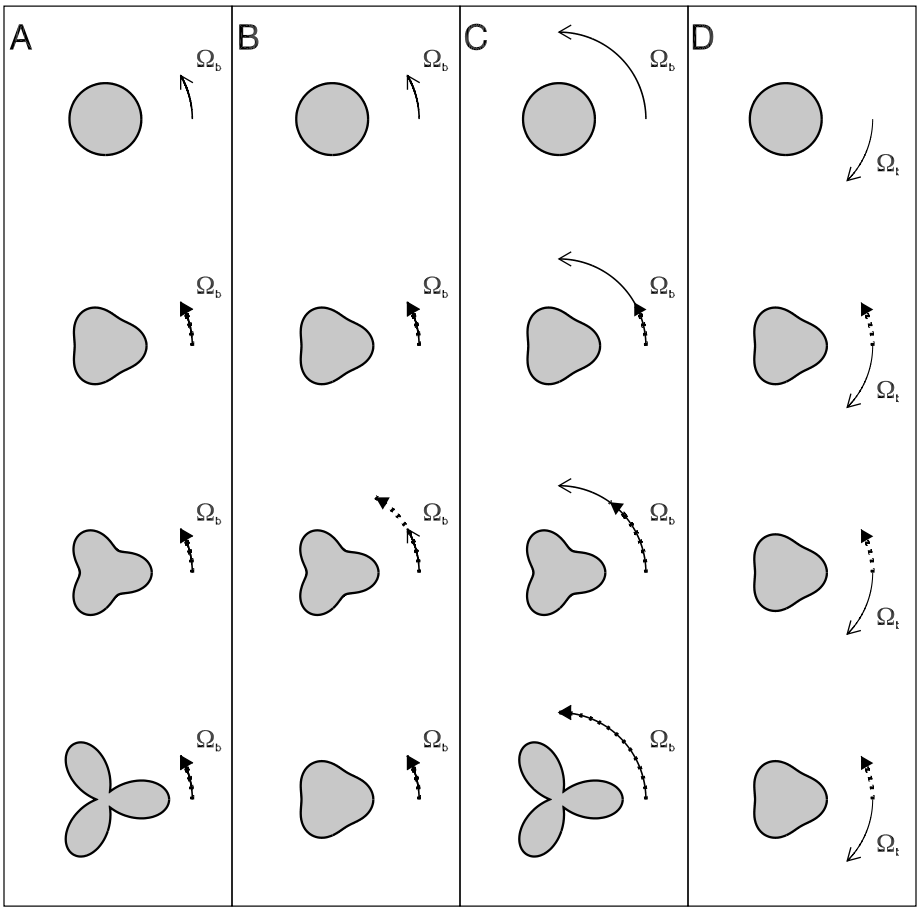

Figure 5.17: Schematic visualizing resonant and non-resonant vortex states in a vortex subject to azimuthal wavenumber-3 type forcing. In all panels, the thin solid line represents the direction and magnitude of the background rotation Ω_b. For each disturbed vortex state, the dotted line represents the direction and magnitude of $\Omega = \Omega_0 + \epsilon^2\Omega_2 > \Omega_0$, the background rotation for which the first harmonic of the disturbance would appear stationary in the fixed frame. Waves are in perfect resonance when $\Omega = \Omega_b$. Panel A: linear resonance when $\Omega_b = \Omega_0$. Panel B: weakly nonlinear theory when $\Omega_b = \Omega_0$ leading to non-resonant behaviour. Panel C: weakly nonlinear theory when $\Omega_b > \Omega_0$ leading to resonant behaviour. Panel D: weakly nonlinear theory when $\Omega_b < \Omega_0$ leading to non-resonant behaviour.

angular velocity of the background rotation must be such that the first harmonic wave becomes stationary with respect to the topography in the fixed frame. In linear theory, the angular velocity for which the first harmonic is stationary is Ω_0 irrespective of wave amplitude. Therefore, as the background rotation $\Omega_b = \Omega_0$ the condition for resonance in the linear model is satisfied. As this angular velocity does not change as the first harmonic amplitude increases, the resonance is sustained and the amplitude of the first harmonic grows indefinitely.

The remaining panels will focus on the weakly nonlinear theory. In the weakly nonlinear theory of section 5.3, it was shown that the angular velocity of the background flow which ensures that the first harmonic is stationary with respect to the forcing is

$$\begin{aligned} \Omega &= \Omega_0 + \epsilon^2 \Omega_2, \\ &= \Omega_0 + \hat{\Omega}_2 A^2 > \Omega_0, \end{aligned} \tag{5.66}$$

where $\hat{\Omega}_2 = \Omega_2/|\alpha_1|^2$ and $A = \epsilon|\alpha_1|$. In panel B, background flow in the model $\Omega_b = \Omega_0$ is the same as for the linear model in panel A. Initially, A is very small such that the background rotation for which waves are resonant is $\Omega = \Omega_0$. As $\Omega_b = \Omega_0 = \Omega$, the first harmonic is therefore in perfect resonance and the amplitude of A increases. However, as the resonance increases the amplitude A, the value of Ω also increases as can be seen in (5.66). Therefore, the initially resonant system has now evolved to a state in which $\Omega > \Omega_0$ and is further from resonance than before. Therefore, the amplitude A decreases until it is small enough that $\Omega \approx \Omega_0$ once again, at which point the process repeats. From this it is seen that a system

predicting resonant growth in the linear model will produce periodic non-resonant behaviour in the weakly nonlinear system.

Panel C gives a weakly nonlinear model representation background angular velocity $\Omega_b > \Omega_0$ close to the linear resonance Ω_0. Initially A will be small and $\Omega \approx \Omega_0$. As the system is close to resonance, the amplitude A will increase giving a more disturbed vortex. As A increases, the angular velocity Ω for which disturbances appear stationary is now $\Omega_0 < \Omega < \Omega_b$. As Ω is now closer to Ω_b the system is closer to resonance. Therefore the amplitude A continues to increase, with an associated increase in Ω such that $\Omega \to \Omega_b$. Therefore, an initially off-resonant disturbance with $\Omega_b > \Omega_0$, close enough to the linear resonance Ω_0, will move into resonance by self-tuning the value of Ω with which the wave is associated towards Ω_b.

In contrast, panel D shows a system that can never achieve resonance. Initially, $\Omega_b < \Omega_0$ is relatively close to resonance so the first harmonic amplitude A is at first small and $\Omega = \Omega_0$. Any amplification of A will act to increase Ω, such that Ω is now even further from resonance and further growth of A is suppressed. Therefore, the amplitude A will remain relatively small for the entirety of the vortex motion. This suppression of the growth in A occurs for all $\Omega_b < \Omega_0$, with the case already discussed in panel B when $\Omega_b = \Omega_0$ acting as a limiting case. In Fig. 5.9 the inability of these states to achieve resonance is equivalent to the absence of a discontinuity in A_{max} in the region $\Omega_b < \Omega_0$.

5.6.2 Hierarchy of models

The results of section 5.5 show that the Kida vortex model acts as an excellent approximation to the weakly nonlinear topographically forced model, which in turn accurately predicts the location of self-tuning resonance in a fully nonlinear topographically forced model. Strictly speaking, the Kida vortex only approximates the behaviour of the weakly nonlinear model when $k = 2$ and $\mathcal{B} = 0$. However, Fig. 5.14 suggests that as long as $k = 2$, vortex behaviour classification in the Kida vortex model still gives a reasonable prediction of classification in the fully nonlinear model when $\mathcal{B} = 1.44$ if one replaces the curved Kida regime boundary with the line of discontinuity in A_{max}. This is particularly true for the stable anticlockwise rotation regime.

As previously discussed in chapter 2, the behaviour of the stratospheric polar vortex during vortex-splitting SSWs is primarily barotropic. That is, the vortex splits near simultaneously at all heights. Furthermore, the way in which the vortex splits, with pinching at the vortex centroid before complete splitting into two almost equally sized daughter vortices, is similar to the vortex split observed during Love type instabilities of an elliptical vortex. Therefore, as already discussed, a simple single layer quasi-geostrophic f-plane model may offer significant insight into the underlying dynamics of vortex-splitting SSWs. When interpreting the work of this chapter in this context, our results have two main implications for the study of vortex-splitting SSWs.

Firstly, it has been shown that from the perspective of self-tuning resonance, an increase in physical forcing amplitude in the system is not strictly necessary in order

to generate vortex-splitting behaviour. Rather, for time-independent forcing which is of sufficient amplitude, an arbitrarily small change in the background rotation can cause vortex-splitting behaviour. In terms of the stratospheric polar vortex, this implies that an increase in wave forcing at the tropopause may not necessarily be required for the onset of vortex-splitting SSWs. That is, given a constant forcing at the tropopause which is of sufficient amplitude, a periodic stable polar vortex evolution may make the transition to an unstable vortex-splitting evolution with only a very small change in the background flowin the stratosphere.

The second implication comes from the relevance of the Kida vortex model in identifying the regions in parameter space at which these transitions in vortex behaviour occur. In particular, should an observational study of the polar vortex prior to a vortex-splitting SSW reveal anticlockwise rotation, one could confidently locate the vortex in the stable regime of the Kida vortex model, and hence the low amplitude disturbance region of the weakly nonlinear theory. With continued observations of the vortex leading up to the onset of the SSW, it may be possible to track the location of the vortex in the parameter space of the weakly nonlinear theory. In this framework, it is expected that the onset of vortex-splitting SSWs would be accompanied by a self-tuning of the vortex towards resonance in the weakly nonlinear theory, with a transition from a low amplitude stable vortex state to a high amplitude unstable vortex state by crossing the line of discontinuity in the parameter space of the weakly nonlinear theory.

Chapter 6

Resonant Response in an Idealized Three-Dimensional Model: Vortex-Displacement

6.1 Introduction

In this chapter we will study the behaviour of an idealized polar vortex using the three-dimensional quasi-geostrophic (QG) polar f-plane model already discussed in chapter 3.

The anelastic three-dimensional QG polar f-plane model has been used extensively in existing studies of the stratospheric polar vortex. One such study is that of Dritschel and Saravanan (1994), who investigated the nature of wave breaking on the polar vortex edge in the presence of topographic forcing at the lower boundary of the model. Dritschel and Saravanan (1994) considered a stratosphere of finite height, in

which the polar vortex was represented by a column of anomalously high PV, with the PV and cross-sectional area of the vortex being the same at all heights of the model (i.e. a "barotropic" vortex). Two types of behaviour were observed in the vortex model depending on the amplitude of the underlying topographic forcing. For large topographic forcing amplitudes, "local" wave breaking was observed, in which greatest disturbance to the vortex was observed in the lower vortex. For smaller topographic forcing amplitudes, "remote" wave breaking was observed, whereby wave-like disturbances to the vortex edge propagated up the vortex before breaking in the upper vortex region.

In a following study by Waugh and Dritschel (1999), the cross sectional area and PV of the vortex were varied with height to investigate the impact of the vortex structure on the occurrence of wave breaking. Using a fully nonlinear numerical model, it was found that if the PV of the vortex increased with height, wave breaking at the vortex edge was less likely to be observed than for other vortex structures and was restricted to the lower vortex. That is, vortex structures in which PV increased with height were more resilient to topographic forcing than barotropic vortices.

A similar model was used by Esler and Scott (2005), who proposed that under certain conditions, resonance of a barotropic polar vortex with topographic forcing at the lower boundary (similar to that discussed in the β-plane models of Plumb 1981; Tung and Lindzen 1979a, b, and the single layer model of chapter 5 in this thesis) can lead to behaviour similar to that observed during vortex-splitting SSWs. By considering a semi-infinite stratosphere model, Esler and Scott (2005) developed a linear theory predicting that for small amplitude topographic forcing, a resonance

of barotropic disturbances to the vortex edge, that is disturbances which are the same at all heights of the vortex, dominates the vortex behaviour. For relatively large amplitude topographic forcing with azimuthal wavenumber-2 structure, fully nonlinear numerical experiments showed that a resonance of the barotropic mode of the vortex lead to an almost simultaneous split of the vortex at all heights. These vortex splits exhibited a similar vertical structure and timescale to those seen in the observations, for example Fig. 2.3 of chapter 2. As the forcing amplitude in the nonlinear numerical model increased, Esler and Scott (2005) observed a migration of the flow conditions leading to the greatest resonant excitation away from that predicted by linear theory. Qualitatively similar behaviour had been previously found by Plumb (1981) for a β-plane model.

In this chapter, using topographic forcing with azimuthal wavenumber-1 structure, the linear theory of Esler and Scott (2005) is adapted to investigate resonant excitation of disturbances on the polar vortex leading to behaviour similar to that seen during vortex-displacement SSWs. Although it was shown by Esler and Scott (2005) that a resonance of the barotropic mode of the vortex dominates the behaviour of the vortex in their model, such a resonance is unlikely to be responsible for behaviour observed during vortex-displacement SSW, due to the strong dependence on height which characterizes the polar vortex during these events (see chapter 2). Therefore, in an attempt to simulate polar vortex behaviour during vortex-displacement SSWs, this chapter we will investigate the role of resonant excitation of the baroclinic modes of the vortex, that is modes which are height dependent, as a possible mechanism for vortex breakdown.

The structure of this chapter is as follows. We will start with a thorough review of the linear theory presented in Esler and Scott (2005), which predicts the background flow conditions leading to greatest resonant excitation of the vortex in the presence of topographic forcing. Using a fully nonlinear numerical model, these predictions are then tested for topographic forcing with azimuthal wavenumber-1 structure, focusing on resonant excitation of baroclinic modes of the vortex, leading to height-dependent vortex behaviour. Three-dimensional evolution of the vortex during the fully nonlinear model experiments is then discussed and a "Baroclinic Sudden Warming" event (BCSW) is described, in which a resonance of the baroclinic modes of the vortex leads to vortex breakdown similar to that observed during vortex-displacement SSWs. Finally, the conclusions of the chapter are presented.

6.2 Resonant excitation of normal modes: linear theory

This section will present a thorough review of the work of Esler and Scott (2005) giving linear predictions of the relative importance of barotropic and baroclinic disturbances to an initially barotropic vortex in the presence of topographic forcing.

6.2.1 Quasi-geostrophic compressible atmosphere model

The theoretical model used in this formulation is the anelastic three-dimensional quasi-geostrophic f-plane model in Cartesian coordinates introduced in chapter 3.

Following the discussion surrounding equation (3.30), the quasi-geostrophic PV

is conserved following the geostrophic flow which has horizontal streamfunction $\mathbf{u} = -\nabla \times \psi\mathbf{k}$. The streamfunction ψ and PV q are related by

$$f_0 + \nabla_H^2\psi + \frac{1}{\rho(z)}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\left(\rho(z)\frac{\partial\psi}{\partial z}\right) \;=\; q(\mathbf{x}, z, t), \tag{6.1}$$

where $\mathbf{x}$ is the horizontal position vector (x, y), z is log-pressure height, ∇_H is the horizontal Laplacian, the density $\rho(z) = \rho_s \exp\{-z/H\}$, with H a reference height and ρ_s a reference density, and f_0 is the Coriolis parameter evaluated at the pole. N is the Brunt-Väisälä buoyancy frequency, which is taken to be constant.

The model is semi-infinite in the vertical, and here we take the lower boundary condition to be the ersatz boundary condition

$$\frac{\partial\psi}{\partial z} + \frac{N^2}{f_0}h_T \;=\; 0 \quad \text{on} \quad z = 0. \tag{6.2}$$

The ersatz boundary condition is used here for analytical convenience, and a discussion of the associated impact on the streamfunction of the flow is given in chapter 3. By inverting the elliptic operator in equation (6.1) subject to condition (6.2) on the lower boundary, the velocity $\mathbf{u} = -\nabla \times \psi\mathbf{k}$ can be determined at any point in the domain from the PV field $q(\mathbf{x}, z, t)$.

6.2.2 Polar vortex representation

In the most general case, the vortex is defined at all heights of the model by its local cross sectional area $D(z)$, with bounding contour $\partial D(z)$. The PV inside this bounding contour is single valued and higher than that of the background PV outside

the vortex such that

$$q(\mathbf{x}, z) = \begin{cases} f + \Delta(z) + \Delta_b & \mathbf{x} \in D(z) \\ f + \Delta_b & \mathbf{x} \notin D(z). \end{cases} \tag{6.3}$$

The parameter Δ_b controls the value of the ambient PV and corresponds to a solid body rotation of angular velocity $\Omega_b = \Delta_b/2$. Δ_b acts as a tunable parameter in the model.

For the idealized polar vortex structure considered in the remainder of this chapter, the interior PV anomaly Δ is taken to be independent of height, and the undisturbed vortex structure is columnar with radius R at all heights, i.e. the vortex is barotropic. The analysis which follows will investigate the propagation of linear Rossby waves on the edge of this barotropic vortex structure.

6.2.3 Dispersion relation

In the absence of topographic forcing, the streamfunction for the flow is partitioned in a basic state streamfunction component Ψ satisfying

$$\nabla_H^2 \Psi + \frac{f_0^2}{N^2}\left(\frac{\partial^2 \Psi}{\partial z^2} - \frac{1}{H}\frac{\partial \Psi}{\partial z}\right) = \begin{cases} \Delta + \Delta_b & r < R \\ \Delta_b & r > R, \end{cases} \tag{6.4}$$

$$\frac{\partial \Psi}{\partial z} = 0 \quad \text{on} \quad z = 0, \tag{6.5}$$

and a component ψ, corresponding to infinitesimal disturbances of the vortex edge η, satisfying

$$\nabla_H^2 \psi + \frac{f_0^2}{N^2}\left(\frac{\partial^2 \psi}{\partial z^2} - \frac{1}{H}\frac{\partial \psi}{\partial z}\right) = 0, \tag{6.6}$$

$$\frac{\partial \psi}{\partial z} = 0 \quad \text{on} \quad z = 0. \tag{6.7}$$

In cylindrical coordinates (r, ϕ, z), the disturbed vortex edge location is given by $r(\phi, z, t) = R + \eta(\phi, z, t)$, and the kinematic condition can be written as

$$\left(\frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_H\right)(\eta - r) = 0, \quad \text{on} \quad r = R + \eta, \tag{6.8}$$

where $\mathbf{u} = (u, v)$ is the horizontal velocity, with azimuthal component u and radial component v in the cylindrical coordinate system. Linearizing the kinematic condition gives

$$\left(\frac{\partial}{\partial t} + \Omega_e \frac{\partial}{\partial \phi}\right)\eta - v = 0 \quad \text{on} \quad r = R, \tag{6.9}$$

where $\Omega_e = (\Delta + \Delta_b)/2$ is the angular velocity of the the basic state and v is the $O(\eta)$ radial velocity due to the disturbance. By expressing the disturbance PV on the vortex edge in terms of the Dirac-delta function $q' = \Delta(z)\delta(r - R(z))\eta$, following Swanson et al. (1997), the $O(\eta)$ jump in azimuthal velocity at the vortex edge associated with the infinitesimal disturbance η satisfies

$$[\psi_r]_{r=R_-}^{r=R+} = -\Delta\eta. \tag{6.10}$$

Barotropic disturbance

Infinitesimal barotropic disturbances to the columnar vortex are given by

$$\eta_0(\phi, t) = \epsilon R \exp\left[ik\phi - i\omega_{0k}t\right], \tag{6.11}$$

where the real part is taken, with $k \geq 1$ an integer azimuthal wavenumber and ω_{0k} a frequency which is yet to be determined, with corresponding streamfunction

$$\psi_0(r, \phi, t) = \tilde{\psi}_0(r)\eta_0(\phi, t). \tag{6.12}$$

Substituting (6.12) into (6.6) yields an ODE for $\tilde{\psi}_0(r)$

$$r^2\tilde{\psi}_{0_{rr}} + r\tilde{\psi}_{0_r} - k^2\tilde{\psi}_0 = 0, \tag{6.13}$$

which by using the velocity jump condition (6.10) in conjunction with continuity of ψ on $r = R$ gives

$$\tilde{\psi}_0(r;k) = \frac{\Delta}{2k}\begin{cases} \left(\frac{r}{R}\right)^k & \text{for} \quad r < R \\ \left(\frac{R}{r}\right)^k & \text{for} \quad r > R. \end{cases} \tag{6.14}$$

In deriving (6.14), the streamfunction ψ is taken to be bounded everywhere.

The linearized radial velocity on the vortex edge corresponding to the barotropic disturbance, $v = (1/R)\psi_{0_\phi}$, can now be derived from (6.12) giving

$$v = \frac{\Delta}{2} i\eta_0,$$

which is substituted along with (6.11) into (6.9) to give the barotropic dispersion relation

$$\omega_{0k} = k\Omega_e - \frac{\Delta}{2}. \tag{6.15}$$

Should the disturbance (6.11) be stationary in the fixed frame, this dispersion relation can be rearranged to give

$$\Delta_0 = \frac{\Delta(1-k)}{2k}, \tag{6.16}$$

that is, the value of Δ_b which ensures that such disturbances are stationary. It is seen that Δ_0 is identical to that for wave-like disturbances of the form (6.11) in the single layer problem (Su 1979), as shown in chapter 5.

Baroclinic disturbance and the Charney-Drazin spectrum

For height-dependent baroclinic disturbances to the columnar vortex edge, we consider linear normal mode solutions of the form

$$\eta_m(\phi, z, t; m) \;=\; \epsilon R \exp\left[\frac{z}{2H} + imz + ik\phi - i\omega_k(m)t\right], \tag{6.17}$$

where $0 < m < \infty$, with a corresponding height-dependent disturbance streamfunction

$$\psi(r, \phi, z, t; m) \;=\; \tilde{\psi}(r; m)\eta_m(\phi, z, t; m). \tag{6.18}$$

Substituting (6.18) into (6.6) yields an ODE for $\psi(r)$

$$r^2\tilde{\psi}_{rr} + r\tilde{\psi}_r - \left(\gamma(m)^2 r^2 + k^2\right)\tilde{\psi} \;=\; 0, \tag{6.19}$$

where

$$\gamma(m) \;=\; \frac{f_0}{N}\left(m^2 + \frac{1}{4H^2}\right)^{1/2}. \tag{6.20}$$

Utilizing the velocity jump condition (6.10), solution of (6.19) with the lower boundary condition (6.7) gives

$$\tilde{\psi}(r; m, k) \;=\; \Delta R \begin{cases} K_k(\gamma(m)R)I_k(\gamma(m)r) & \text{for} \quad r < R \\ I_k(\gamma(m)R)K_k(\gamma(m)r) & \text{for} \quad r > R, \end{cases} \tag{6.21}$$

by demanding that $\psi(r, \phi, z, t; m)$ be continuous on the vortex edge and bounded everywhere. The functions I_k and K_k appearing in equation (6.21) are modified Bessel functions of the first and second kind respectively. Equations (6.18) and (6.21) give the $O(\eta)$ radial velocity $v = (1/R)\psi_\phi(r, \phi, z, t; m)$ on the vortex edge as

$$v \;=\; ik\Delta I_k\left(\gamma(m)R\right) K_k\left(\gamma(m)R\right)\eta_m,$$

which is substituted along with (6.17) into (6.9) to give the baroclinic dispersion relation

$$\omega_k(m) \;=\; k\Omega_e - \Delta k I_k\left(\gamma(m)R\right) K_k\left(\gamma(m)R\right). \tag{6.22}$$

The $\omega_k(m)$ correspond to a spectrum of frequencies in the range $\omega_s^- \leq \omega_k \leq \omega_s^+$ where

$$\omega_k^-(0) \;=\; k\Omega_e - \Delta k K_k\left(\frac{R}{2L_R}\right) I_k\left(\frac{R}{2L_R}\right),$$

$$\omega_k^+(\infty) \;=\; k\Omega_e,$$

where $L_R = NH/f_0$ is the Rossby radius of deformation. Following Charney and Drazin (1961) this 'Charney-Drazin spectrum' corresponds to the possible frequencies of Rossby wave which are permitted to propagate vertically on the vortex edge. Note that this propagation may engender both upward and downward propagation of the permitted Rossby waves.

The choice of Δ_b that renders the mth baroclinic disturbance stationary with respect to the fixed frame can be obtained by setting $\omega_k(m) = 0$ and is found to be

$$\Delta_b(m) \;=\; \Delta\left[K_k(\gamma(m)R)I_k(\gamma(m)R) - 1\right], \tag{6.23}$$

and the largest and smallest values of $\Delta_b(m)$ for which stationary forcing will excite upward propagating Rossby waves are given by

$$\Delta_- \;=\; \Delta\left[K_k\left(\frac{R}{2L_R}\right) I_k\left(\frac{R}{2L_R}\right) - 1\right],$$

$$\Delta_+ \;=\; -\Delta.$$

6.3 Vortex response to topographic forcing at the lower boundary

The barotropic and baroclinic disturbances $\{\eta_0,\eta(m)\}$ form a complete basis with respect to which all disturbances to the vortex edge can be expressed. Therefore, when a lower boundary topographic forcing is introduced, linear theory dictates that the resulting disturbance of the vortex edge will be some linear combination of η_0 and the spectrum $\eta(m)$. In this section, an analytic expression for the combined linear disturbance η_R is derived following Esler and Scott (2005).

The vortex is forced with a generalized time-dependent topographic forcing on the lower boundary of the form

$$h_T(r,\phi,t) = \frac{h}{2}T(t)J_k(lr)\exp\{ik\phi\}. \tag{6.24}$$

The general time-dependent function $T(t)$ will be chosen to represent an isolated pulse of forcing from the lower boundary. It can be expressed as a Fourier integral transform of the form

$$T(t) \;=\; \int_{-\infty}^{\infty}\widehat{T}(\omega)\exp\{i\omega t\}d\omega. \tag{6.25}$$

with the forward transform given by

$$\widehat{T}(\omega) \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}T(t)\exp\{-i\omega t\}dt.$$

The disturbance streamfunction of the resulting flow is partitioned into a topographic component ψ_T and a response component ψ_R such that $\psi = \psi_T + \psi_R$ with

$$\nabla_H^2\psi_T + \frac{f_0^2}{N^2}\left(\frac{\partial^2\psi_T}{\partial z^2} - \frac{1}{H}\frac{\partial\psi_T}{\partial z}\right) \;=\; 0, \tag{6.26}$$

$$N^2 h_T + f_0 \frac{\partial \psi_T}{\partial z} = 0 \quad \text{on} \quad z = 0, \tag{6.27}$$

and

$$\nabla_H^2 \psi_R + \frac{f_0^2}{N^2}\left(\frac{\partial^2 \psi_R}{\partial z^2} - \frac{1}{H}\frac{\partial \psi_R}{\partial z}\right) = q - f_0, \tag{6.28}$$

$$\frac{\partial \psi_R}{\partial z} = 0 \quad \text{on} \quad z = 0. \tag{6.29}$$

The linearized form of ψ_R is identical to that of the disturbance streamfunction in the absence of topographic forcing, which satisfies (6.6) and (6.7), as derived in the previous section . However it remains to solve (6.26) and (6.27) for ψ_T using (6.24). Using separation of variables, the solution is expressed in the form

$$\psi_T(r, \phi, t, z) = AT(t)F(z)J_k(lr)\exp\{ik\phi\}, \tag{6.30}$$

where $T(t)$ may be expressed as an integral transform following (6.25). Substituting (6.30) into (6.26) yields an ODE for $F(z)$,

$$\frac{d^2F}{dz^2} - \frac{1}{H}\frac{dF}{dz} + \frac{N^2 l^2}{f^2}F = 0, \tag{6.31}$$

which with the condition that $F \to 0$ as $z \to \infty$ gives

$$F(z) = e^{\frac{z}{2H}} e^{-\alpha z}, \tag{6.32}$$

where

$$\alpha^2 = \frac{1}{4H^2} + \frac{l^2 N^2}{f^2}, \tag{6.33}$$

and only the positive root of α is taken. Using the above formulation for $F(z)$, we now substitute (6.30) into (6.27) giving

$$A = \frac{f_0 L_R^2 h}{H(2\alpha H - 1)}.$$

Therefore, expressing the general time-dependent forcing as an integral transform, the topographic streamfunction can be written as

$$\psi_T = \frac{f_0 L_R^2 h}{H(2\alpha H - 1)} \int_{-\infty}^{\infty} \widehat{T}(\omega) J_k(lr) \exp\left\{\frac{z}{2H} - \alpha z + ik\phi - i\omega t\right\} d\omega. \quad (6.34)$$

It is now desirable to project ψ_T onto the vertical modes of the vortex defined in section 3.2.2 of chapter 3 as

$$\chi_0 = \sqrt{\frac{1}{H}}, \quad (6.35)$$

$$\chi(z;m) = \sqrt{\frac{2}{\pi}} e^{\frac{z}{2H}} \cos[mz + \epsilon(m)], \qquad \tan\epsilon(m) = \frac{1}{2mH}, \quad (6.36)$$

with corresponding associated eigenvalues

$$\gamma_0 = 0, \quad (6.37)$$

$$\gamma(m) = \frac{f}{N}\left(m^2 + \frac{1}{4H^2}\right)^{1/2}, \quad (6.38)$$

and $m > 0$. By construction, these vertical modes are orthonormal, satisfying the orthogonality conditions (3.38), (3.41) and (3.42) given in section 3.2.2 of chapter 3. We start by expressing $\exp\{z/2H - \alpha z\}$ as an integral transform of these vertical modes, that is

$$\exp\left\{\frac{z}{2H} - \alpha z\right\} = a_0\chi_0 + \int_0^{\infty} a(m)\chi(z;m)dm. \quad (6.39)$$

Multiplying by $\chi_0 \exp\{-z/H\}$ and making use of the orthogonality conditions for $\chi_0(z)$ and $\chi(z;m)$ gives

$$\begin{aligned} a_0, &= \int_0^\infty \chi_0 \exp\left\{ -\frac{z}{2H} - \alpha z \right\} dz, \\ &= \frac{2H^{1/2}}{1+2H\alpha}. \end{aligned} \tag{6.40}$$

Similarly, multiplying by $\chi(z;m')\exp\{-z/H\}$ and integrating with respect to z gives

$$\begin{aligned} \int_0^\infty \chi(z;m') \exp\left\{ -\frac{z}{2H} - \alpha z \right\} dz &= \int_0^\infty a(m) \times \\ &\qquad \left[\int_0^\infty \exp\left\{ -\frac{z}{H} \right\} \chi(z;m')\chi(z;m) dz \right] dm, \\ &= \int_0^\infty a(m)\delta(m-m')dm, \\ &= a(m'). \end{aligned} \tag{6.41}$$

Integrating the LHS by parts and using $\sin\epsilon = 1/(4m'^2H^2+1)^{1/2}$ and $\cos\epsilon = 2m'H/(4m'^2H^2+1)^{1/2}$ gives the $a(m')$ coefficient as

$$a(m') = \sqrt{\frac{2}{\pi}} \frac{m'(2\alpha H - 1)}{(4m'^2H^2+1)^{1/2}(\alpha^2 + m'^2)}. \tag{6.42}$$

Substituting the integral transform expression for $\exp\{z/2H - \alpha z\}$ into ψ_T yields an expression for ψ_T as an integral transform in terms of the vertical modes

$$\psi_T = \psi_{T_0} + \psi_{T_m}, \tag{6.43}$$

where

$$\psi_{T_0}(\phi,t) = \frac{f_0 h}{2l^2 H} J_k(lr) \int_{-\infty}^{\infty} \widehat{T}(\omega) \exp\{ik\phi - i\omega t\} d\omega, \tag{6.44}$$

$$\psi_{T_m}(\phi,z,t;m) = \frac{2f_0 L_R^2 h}{\pi H} J_k(lr) \int_{-\infty}^{\infty} \widehat{T}(\omega) \exp\left\{\frac{z}{2H} + ik\phi - i\omega t\right\} \times \left[\int_0^{\infty} \widehat{\psi}(m) \cos[mz + \epsilon(m)] dm\right] d\omega, \tag{6.45}$$

$$\widehat{\psi}_{T_m}(m) = \frac{m}{(1+4m^2H^2)^{1/2}(\alpha^2+m^2)}. \tag{6.46}$$

We now derive the disturbance to the vortex edge $\eta_R(\phi,z,t)$, which occurs in response to the topographic forcing, as a linear combination of the barotropic and baroclinic disturbances (6.11) and (6.17). As $\eta_R(\phi,z,t)$ is some linear combination of $\eta_0(\phi,t)$ and $\eta_m(\phi,z,t;m)$, it is convenient to treat the barotropic and baroclinic contributions separately. Therefore, using the linear kinematic condition on the vortex edge (6.9) we have

$$\left(\frac{\partial}{\partial t} + \Omega_e \frac{\partial}{\partial \phi}\right) \eta_{R_0} + \frac{1}{r}\frac{\partial}{\partial \phi}\left(\psi_{R_0} + \psi_{T_0}\right) = 0 \quad \text{on} \quad r = R, \tag{6.47}$$

and

$$\left(\frac{\partial}{\partial t} + \Omega_e \frac{\partial}{\partial \phi}\right) \eta_{R_m} + \frac{1}{r}\frac{\partial}{\partial \phi}\left(\psi_{R_m} + \psi_{T_m}\right) = 0 \quad \text{on} \quad r = R. \tag{6.48}$$

From the form of $h_T(r,\phi,t)$ it can be deduced that $\eta_t = -i\omega\eta$ and $\eta_\phi = ik\eta$ in both of the equations (6.47) and (6.48). Utilizing this property with (6.47) and (6.48) yields

$$\eta_{R_0} = \frac{k\psi_{T_0}}{R\left(\omega - k\Omega_e + k\tilde{\psi}_0(R;k)\right)}, \tag{6.49}$$

$$= \frac{kf_0 hJ_k(lR)}{2l^2RH}\int_{-\infty}^{\infty}\frac{\widehat{T}(\omega)\exp\{ik\phi - i\omega t\}}{\omega - k\Omega_e + \Delta/2}d\omega, \tag{6.50}$$

and

$$\eta_{R_m} = \frac{k\psi_{T_m}}{R\left(\omega - k\Omega_e + k\Delta\tilde{\psi}_m(R;m,k)\right)}, \tag{6.51}$$

$$= \frac{2kf_0L_R^2hJ_k(lR)}{\pi RH}\int_{-\infty}^{\infty}\widehat{T}(\omega)\exp\left\{\frac{z}{2H} + ik\phi - i\omega t\right\}\times$$

$$\left[\int_0^{\infty}\frac{\widehat{\psi}_{T_m}(m)\cos[mz + \epsilon(m)]}{\omega - k\Omega_e + \Delta I_k(\mathcal{B}(m))K_k(\mathcal{B}(m))}dm\right]d\omega, \tag{6.52}$$

where $\mathcal{B}(m) = \gamma(m)R$. The integration with respect to ω in (6.52) and (6.50) is now approached using complex techniques.

If ζ is a complex coordinate, then for a holomorphic function $f(\zeta)$, Cauchy's Integral Formula gives the integral around a closed contour Γ in the ζ-plane as

$$\oint_{\Gamma}\frac{f(\zeta)}{\zeta - \zeta'}d\zeta = 2\pi i f(\zeta').$$

Taking Γ to be a semicircular contour lying in the upper half of the ζ-plane, if $\max|f(\zeta)| \to 0$ as $|\zeta| \to \infty$, application of Jordan's Lemma leaves us with

$$\int_{-\infty}^{\infty}\frac{f(\zeta)}{\zeta - \zeta'}d\zeta = 2\pi i f(\zeta').$$

The correct causality can be enforced by taking $\zeta = \omega - i\varepsilon$ where ε is a small parameter, the sign of which will be determined, after taking the limit $\varepsilon \to 0$. The integrals with respect to ω in (6.52) and (6.50) transform to

$$\int_{-\infty}^{\infty}\frac{\widehat{T}(\omega)}{\omega - \omega'}\exp\{-i\omega t\}d\omega = \exp\{\varepsilon t\}\int_{-\infty}^{\infty}\frac{\widehat{T}(\zeta + i\varepsilon)}{\zeta - \zeta'}\exp\{-i\zeta t\}d\zeta,$$

where ω' can be either ω_{0k} or $\omega_k(m)$ and $\zeta' = \omega' - i\epsilon$. Demanding that the disturbance disappears as $t \to -\infty$ requires that $\varepsilon > 0$. Therefore, if $\widehat{T}$ is a holomorphic function satisfying $\max|\widehat{T}| \to 0$ as $|\zeta| \to \infty$, this gives

$$\lim_{\varepsilon \to 0} \int_{-\infty}^{\infty} \frac{\widehat{T}(\zeta + i\epsilon)}{\zeta - \zeta'} \exp\{-i\zeta t\} d\zeta \quad = 2\pi i \widehat{T}(\zeta' + i\epsilon) \quad = 2\pi i \widehat{T}(w'),$$

such that the long time behaviour of the disturbance can be expressed as

$$\eta_{R_0}(\phi) \quad \sim \quad \frac{\pi s i}{L_R^2 l^2} \widehat{T}(\omega_{0k}) \exp\{ik\phi - i\omega_{0k} t\}, \tag{6.53}$$

and

$$\begin{aligned} \eta_{R_m}(\phi, z) \quad \sim \quad & 4si \int_0^{\infty} \widehat{T}(\omega_k(m)) \widehat{\psi}_{T_m}(m) \times \\ & \cos[mz + \epsilon(m)] \exp\left\{ \frac{z}{2H} + ik\phi - i\omega_k(m) t \right\} dm, \end{aligned} \tag{6.54}$$

where

$$s \quad = \quad \frac{k f_0 L_R^2 h J_k(lR)}{RH}.$$

6.3.1 Wave-activity conservation and the Eliassen-Palm flux

Following Esler and Scott (2005) and Wang and Fyfe (2000), the vortex response to lower boundary forcing is now investigated from the perspective of wave-activity and Eliassen-Palm flux diagnostics. In this section, the radius of the vortex cross sections $R(z)$, vortex interior PV anomaly $\Delta(z)$, and fluid density $\rho(z)$, are left as general functions of z to demonstrate the applicability of these diagnostics to non-barotropic vortex structures, which will appear in chapter (7).

Similar to the angular impulse of a vortex patch in the single layer model, as discussed in section 4.5.4 of chapter 4, the total angular impulse of a vortex described at each height by a single circular vortex patch of uniform PV can be written

$$\begin{aligned} \mathcal{I}_0 &= \frac{1}{2}\int_0^\infty\int_0^{2\pi}\int_0^{R(z)} \rho(z)\Delta(z) r^3 dr d\phi dz, \\ &= \int_0^\infty \frac{\pi}{4}\Delta(z) R(z)^4 dz. \end{aligned}$$

By subtracting the angular impulse $\mathcal{I}_0$ of the undisturbed vortex from that of the disturbed vortex, the angular impulse associated with wave-like disturbances of the vortex edge $r = R(z) + \eta(\phi, z, t)$ can be written

$$\mathcal{I} = \int_0^\infty A(z,t) dz,$$

where $A(z,t)$ is the wave-activity density

$$A(z,t) = \frac{1}{8}\rho(z)\Delta(z)\int_0^{2\pi}\left[(R+\eta)^2 - R^2\right]^2 d\phi. \tag{6.55}$$

For linear disturbances, at leading order in η the wave-activity density is therefore

$$A(z,t) \approx \frac{1}{2}\rho(z)\Delta(z)R(z)^2\overline{\eta^2}, \tag{6.56}$$

where $\overline{\eta^2}$ is the azimuthal mean of the square of the disturbance amplitude, calculated at the vortex edge

$$\overline{\eta^2} = \frac{1}{2\pi}\int_0^{2\pi}\eta(\phi, z, t)^2 d\phi. \tag{6.57}$$

Taking the linearized kinematic condition on the vortex edge at each height

$$\frac{\partial\eta}{\partial t} + \frac{U}{r}\frac{\partial\eta}{\partial\phi} - \frac{1}{r}\frac{\partial\psi}{\partial\phi} = 0 \quad \text{on} \quad r = R(z) \tag{6.58}$$

where U is the azimuthal velocity of the basic state, multiplying by $\rho(z)R(z)^2\Delta(z)\eta$ and taking the azimuthal average as shown in (6.57) gives

$$\frac{\partial A}{\partial t} - \rho(z)\Delta(z)R(z)\eta\overline{\frac{\partial \psi}{\partial \phi}} = 0. \tag{6.59}$$

Next, the Eliassen-Palm flux $\mathbf{S} = [S^{(r)}, S^{(\phi)}, S^{(z)}]$ (Gill 1982, see also discussion in the Introduction 1.3) is introduced where

$$S^{(r)} = 0$$

$$S^{(\phi)} = -\rho(z)\overline{\frac{\partial \psi}{\partial r}\frac{\partial \psi}{\partial \phi}}$$

$$S^{(z)} = -\rho(z)\frac{f_0^2}{N^2}\overline{\frac{\partial \psi}{\partial \phi}\frac{\partial \psi}{\partial z}},$$

satisfying

$$\nabla \cdot \mathbf{S} = -\rho(z)\overline{q'\frac{\partial \psi}{\partial \phi}},$$

where q' is the PV associated with the disturbance streamfunction ψ. For a single wave-like disturbance, the $S^{(\phi)}$ component becomes zero, leaving

$$\frac{\partial S^{(z)}}{\partial z} = -\rho(z)\overline{q'\frac{\partial \psi}{\partial \phi}}. \tag{6.60}$$

By expressing the disturbance PV in terms of the Dirac-delta function $q' = \Delta(z)\delta(r - R(z))\eta$, multiplying both sides of (6.60) by r and then integrating with respect to r gives

$$\frac{\partial}{\partial z}\int_0^\infty \rho(z)\frac{f_0^2}{N^2}\overline{\frac{\partial \psi}{\partial \phi}\frac{\partial \psi}{\partial z}}r dr = -\rho(z)\Delta(z)R(z)\eta\overline{\frac{\partial \psi}{\partial \phi}}. \tag{6.61}$$

Then, by substituting (6.61) into (6.59), it is seen that

$$\frac{\partial A}{\partial t} + \frac{\partial F}{\partial z} = 0, \tag{6.62}$$

where A is the wave-activity density and $F(z,t)$ is the integral of the vertical component of the Eliassen-Palm flux over the horizontal plane

$$F(z,t) = \int_0^{2\pi}\int_0^{\infty} \rho(z)\frac{f_0^2}{N^2}\frac{\partial\psi}{\partial\phi}\frac{\partial\psi}{\partial z} r dr d\phi.$$

Equation (6.62) takes the form of a conservation relation for the wave-activity density $A(z,t)$. Integrating this conservation relation with respect to z gives

$$\frac{d\mathcal{A}}{dt} = F(0,t), \tag{6.63}$$

where

$$\mathcal{A} = \int_0^{\infty} A(z,t)dz. \tag{6.64}$$

Integrating (6.63) with respect to time, and demanding that disturbances decay to zero as $t \to -\infty$ gives

$$\mathcal{A}_\infty = \int_{-\infty}^{\infty} F(0,t)dt = \mathcal{F}. \tag{6.65}$$

The quantity $\mathcal{A}$ is the total wave-activity, or angular pseudomomentum, of the system (Dritschel and Saravanan 1994).

For disturbances to the semi-infinite barotropic vortex discussed in sections 6.2 and 6.3, the vortex edge disturbance at large times is given by taking the real part $\eta_R = \eta_{R0} + \eta_{Rm}$ from (6.54) and (6.53). Substituting this value of η_R into (6.56) and (6.64) enables $\mathcal{A}_\infty$ to be calculated as

$$\mathcal{A}_\infty = \mathcal{F}_0 + \mathcal{F}_{CDS}$$

$$= \pi^2 k^2 f_0^2 h^2 J_k(lR)^2 \Delta \Big[\frac{\pi |\widehat{T}(\omega_{0k})|^2}{2Hl^4}$$

$$\frac{4L_R^4}{H^2} \int_0^\infty |\widehat{T}(\omega_k(m))|^2 \widehat{\psi}_T(m)^2 dm \Big]. \tag{6.66}$$

In this formulation, $\mathcal{F}_0$ represents waves which are generated by an excitation of the barotropic mode of the vortex, and $\mathcal{F}_{CDS}$ represents excitation of the baroclinic modes corresponding to the Charney-Drazin spectrum previously discussed.

It was shown in Esler and Scott (2005) that the relative contributions of $\mathcal{F}_0$ and $\mathcal{F}_{CDS}$ are such that resonant excitation of the barotropic mode is dominant in terms of wave-activity, compared to the excitation of the baroclinic modes in the Charney-Drazin spectrum, i.e. $\mathcal{F}_0 \gg \mathcal{F}_{CDS}$. Note, however, that the exponential increase in wave amplitude associated with the Charney-Drazin modes means that they always dominate at very high altitudes. The result $\mathcal{F}_0 \gg \mathcal{F}_{CDS}$ also generally holds for the wave-1 case studied here, but in the wave-1 case the result is somewhat misleading, because relatively large values of $\mathcal{F}_0$ are generated by small displacements of the entire vortex relative to its initial centroid. The interesting dynamics in the wave-1 case are entirely due to the baroclinic modes of the Charney-Drazin spectrum.

6.4 Nonlinear response of the QG polar vortex to topographic forcing

6.4.1 Lower boundary forcing and vortex conditions

In this section, a fully nonlinear model is used to examine the evolution of a barotropic vortex when subject to topographic forcing at the lower boundary of the model. The focus of these experiments will be on the role of resonance excitation of baroclinic disturbances on the vortex edge leading to behaviour similar to that observed during vortex-displacement SSWs.

Experiments were performed using the CASL numerical model (see chapter 3) of the quasi-geostrophic f-plane in a cylindrical domain. Forcing of the form

$$h_T(r,\phi,t) = \frac{h}{2}T(t)J_1(lr)\exp\{i\phi\}, \tag{6.67}$$

is introduced at the lower boundary, corresponding to $z = 0$ in the model. As we are focusing on vortex-displacement type behaviour, the topography is specified with azimuthal wavenumber-1 structure. The time-dependent function $T(t)$ in the forcing is given by

$$T(t) \;=\; \exp\Big\{-\left(\frac{t}{\tau}\right)^2\Big\}, \tag{6.68}$$

with associated transform

$$\widehat{T}(\omega) \;=\; \frac{\tau}{2\pi}\exp\Big\{\left(\frac{\tau\omega}{2}\right)^2\Big\}.$$

The scale of the topographic forcing was taken to be $lR = 1.647$ (so that the fifth zero of the J_1 Bessel function coincides with the outer wall of the cylindrical model

domain). Using the terminology of Esler and Scott (2005), this scale of forcing will be referred to as 'vortex scale' (VS hereafter) as the peak in forcing occurs in the vicinity of the vortex edge. Experiments were also performed with two other scales of forcing, $lR = 0.702$ and $lR = 1.017$, (with the second and third zeroes of J_1 coincident with the outer wall of the domain), with the results being qualitatively the same as those shown here for the VS forcing scale topography.

Experiments were performed with $\tau = 2$ days, with the experiment time running from $t = -5\ \tau$ to $t = 5\ \tau$ days. Atmospheric parameters are taken to be the same as those in Esler and Scott (2005) and Waugh and Dritschel (1999), that is $H = 6.14$ km, $L_R = 900$ km, $R = 3L_R = 2700$ km. A rigid lid is imposed at a height $z = D$, where $D = 12H$ corresponding to a height of approximately 73 km above the lower boundary. In the Earth's atmosphere, the vortex only inhabits the lowest 4-5 scale heights of the model, but to minimize wave reflection from the upper rigid lid boundary, and to test the results for a semi-infinite vertical domain in the preceding sections, we consider a domain of twice this height.

For the PV distribution in the model (6.3), we set $\Delta = 0.4f_0$ and Δ_b acts as a tunable parameter of the system with which a solid body background rotation can be added to the flow. By varying Δ_b, it is possible to fix the angular speed of linear disturbances on the vortex edge to be zero, bringing such disturbances into resonance with the lower boundary forcing. A value of $\Delta_b = -0.1f_0$, which was used by Waugh and Dritschel (1999), corresponds to a background flow correction which is typical of that observed during the winter polar stratosphere. Therefore, varying Δ_b from this reference value of $-0.1f_0$ represents stratospheric zonal winds which tend to be

stronger, or weaker, than the climatological mean.

6.4.2 Model details and numerical parameters

The CASL model used for the fully nonlinear numerical experiments consists of a cylindrical domain with outer walls at a distance $30L_R$ from the origin. The domain is discretized into 120 vertical layers of uniform depth. For the spectral velocity calculations, contour node locations are interpolated onto a quadratically spaced radial grid with a resolution of 192 azimuthal grid points and 96 radial grid points. Quadratic grid spacing is chosen such that grid points are concentrated in the inner domain, the region in which vortex advection primarily occurs. On each model layer, the vortex initially has a single PV contour with 51 nodes, corresponding to a node resolution parameter of $\mu = 0.18$ in the contour dynamics component of the CASL algorithm.

6.4.3 Measures of finite amplitude vortex disturbances

As discussed in section 6.3.1, a useful measure of vortex disturbance is that of the vertically integrated wave-activity density $\mathcal{A}$. In terms of the discretized model vortex, the wave-activity density on layer l is calculated using the formula of Dritschel and Saravanan (1994), which generalizes (6.55) to the case of strongly deformed contours,

$$A_l \;=\; \frac{1}{8}\sum_{k=1}^{N_l}\Delta_l\rho_l\oint_{C_k}(x^2+y^2)(x\;dy - y\;dx) - \frac{\pi}{4}\Delta_l\rho_l R_l^4,$$

where N_l is the number of PV contours on layer l, ρ_l is the density on layer l, R_l is the initial radius of the vortex cross section on layer l (R_l is the same on all layers for

the barotropic vortex considered here) and x and y are the node positions on each PV contour. The diagnostic $\mathcal{A}$ is then calculated using

$$\mathcal{A} \;=\; \sum_{l=1}^{L} A_l.$$

For disturbances with azimuthal wavenumber-1 structure, large barotropic disturbances are observed as an almost uniform displacement of the vortex at all heights. That is, for an initially columnar vortex structure, the vortex remains columnar while being displaced from the pole. As we are interested in studying disturbances associated with the baroclinic modes of the vortex, it is desirable to find the wave-activity due to these modes. In order to do this, it is necessary to find the mass weighted centroid of the model vortex $(\bar{x}_0, \bar{y}_0)$ by calculating the first order moment of the PV distribution on each layer. Just as for the single layer model of (4.5.3) the mass weighted centroid can be calculated as a density-weighted average of the first moments of vorticity on each layer,

$$\begin{aligned}
\bar{x}_l &= \frac{1}{4}\sum_{k=1}^{N_l} \Delta_l \oint_{C_k} (x^2\, dy - 2xy\, dx), \\
\bar{y}_l &= \frac{1}{4}\sum_{k=1}^{N_l} \Delta_l \oint_{C_k} (2xy\, dy - y^2\, dx), \\
(\bar{x}_0, \bar{y}_0) &= \sum_{l=1}^{L} \rho_l(x_l, y_l). \qquad (6.69)
\end{aligned}$$

By transforming the x and y PV node locations to $x' = x - x_0$ and $y' = y - y_0$, the total baroclinic wave-activity density can then be measured relative to mass weighted

centroid giving

$$\mathcal{A}_{CDS} = \sum_{l=1}^{L} \frac{1}{8} \sum_{k=1}^{N_l} \Delta_l \rho_l \oint_{C_k} (x'^2 + y'^2)(x'\ dy' - y'\ dx') - \frac{\pi}{4} \Delta_l \rho_l R_l^4.$$

Barotropic wave-1 disturbances, which displace the vortex without changing its shape, do not contribute to $\mathcal{A}_{CDS}$, which consequently measures the wave-activity in the Charney-Drazin spectrum in which we are interested. In terms of the linear prediction for the total wave-activity at large times $\mathcal{A}_\infty$, given in equation (6.66), the vertical integral of the baroclinic wave-activity density satisfies

$$\mathcal{A}_{CDS} \quad = \quad \mathcal{F}_{CDS}.$$

From the expressions for $\mathcal{F}_0$ and $\mathcal{F}_{CDS}$, it is seen that both depend quadratically on the topographic height h. Therefore, to enable direct comparison of the vortex response at various topographic forcing heights in the fully nonlinear model with the linear predictions $\mathcal{F}_0$ and $\mathcal{F}_{CDS}$, the total wave-activity $\mathcal{A}$ is nondimensionalized by multiplying by a factor of $H^2/\mathcal{I}_0 h^2$.

In the finite model, each of the baroclinic modes which contributes to $\mathcal{A}_{\mathcal{CDS}}$ has an associated integer wavenumber $n \geq 1$, where n comes from a discretization of the continuous vertical wavenumber m in the baroclinic dispersion relation (6.22) using

$$m \quad = \quad \frac{n\pi}{D}, \quad 1 \leq n \leq 120, \quad n \in \mathbb{Z}.$$

6.4.4 Results from the three-dimensional numerical model

Wave-activity diagnostics

For the fully nonlinear numerical model results, we will start by looking at the total wave-activity of the vortex at the end of the experiments, i.e. after the pulse of

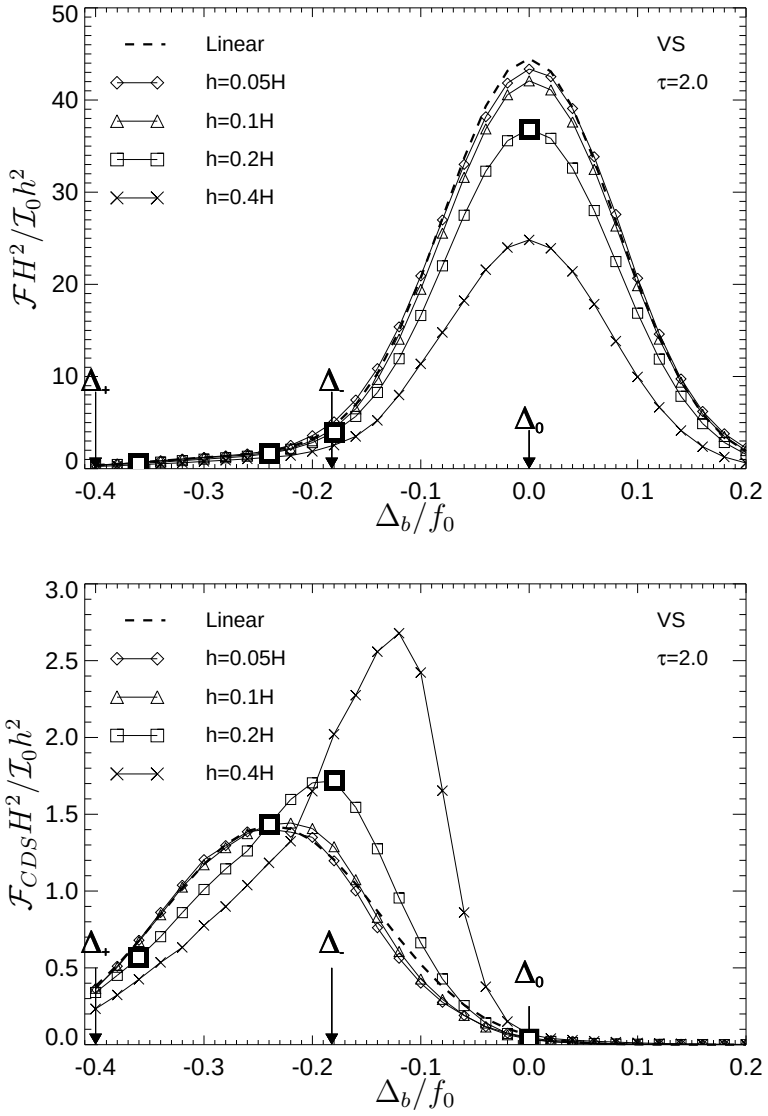

Figure 6.1: Top panel: the total nondimensional wave-activity $\mathcal{F}H^2/\mathcal{I}_0h^2$ as a function of the tuning parameter Δ_b (in units of f_0) for the experiments with $\tau = 2$ days and VS topographic forcing scale. The dashed line corresponds to the prediction of the linear model and the symbols represent results obtained from the fully nonlinear model for a selection of forcing amplitudes. Bottom panel: as top panel, but with total nondimensional wave-activity measured relative to the mass weighted vortex centroid $\mathcal{F}_{CDS}H^2/\mathcal{I}_0h^2$ as a function of the tuning parameter Δ_b. In both panels, Δ_0 denotes the value of Δ_b for which $\omega_{0k} = 0$. The range of Δ_b for which $\omega_k(m) = 0$ lies between Δ_+ and Δ_-. White large black bordered boxes denoted experiments which are discussed in connection with Fig. 6.2.

forcing at the lower boundary has decayed. Fig. 6.1 gives the nondimensional wave-activity as a function of tuning parameter Δ_b for VS topographic forcing scale with $\tau = 2.0$. There is excellent agreement between the predictions of the linear model and those of the nonlinear model with topographic forcing height $h = 0.05H$, in both the location and amplitude of the peak response. As the amplitude of the forcing increases however, the amplitude of the response decreases. It is emphasized, however, that the location of the peak response remains at $\Delta_b = \Delta_0$, at which barotropic disturbances of the vortex appear stationary with respect to the forcing in the linear theory, for all forcing amplitudes. This point is interesting in particular, as for topographic forcing with wavenumber-2 azimuthal structure, the Δ_b location at which the peak response occurs migrates from Δ_0 as the topographic forcing amplitude increases (see Esler and Scott 2005, see also the single layer model results in chapter 5 for azimuthal wavenumber-2 and 3 forcing structures). Also, the distribution of the total wave-activity around the peak at $\Delta_b = \Delta_0$ remains almost uniform at all forcing amplitudes.

We now investigate the wave-activity due to disturbances with baroclinic structure in the vertical. In the linear model this corresponds directly to the wave-activity generated by the $\mathcal{F}_{CDS}$ component in the total wave-activity equation (6.66). The bottom panel of Fig. 6.1 shows the nondimensional baroclinic wave-activity for the fully nonlinear model experiments discussed above. Once again, excellent agreement is observed between the linear prediction for $\mathcal{F}_{CDS}$, as shown by the dashed line, and the nondimensional baroclinic wave-activity for the fully nonlinear model experiments with $h = 0.05H$. In both cases, the peak response is located at $\Delta_b = -0.24 f_0$, cor-

responding to the resonant frequency of the second baroclinic mode, i.e. the vertical mode with wavenumber $n = 2$ ($m = 0.523$). For larger values of the topographic forcing amplitude, the peak response in wave-activity migrates to more positive values of the tuning parameter Δ_b, away from the region $[\Delta_+, \Delta_-]$ associated with the Charney-Drazin spectrum. For forcing amplitude $h = 0.2H$, it is seen that the peak response now occurs for $\Delta_b = -0.18f_0$ near the uppermost limit of the Charney-Drazin spectrum at Δ_-, and for $h = 0.4H$ the peak response occurs at $\Delta_b = -0.12f_0$ completely outside the Charney-Drazin spectrum. In addition, the normalized amplitude of the baroclinic wave-activity increases as the forcing amplitude increases, in contrast to the barotropic wave-activity which decreases as forcing amplitude increases.

As the forcing amplitude increases, the distribution of the baroclinic wave-activity around its peak becomes increasingly asymmetrical, with larger values seen to the left of the peak, in the Charney-Drazin region $[\Delta_+, \Delta_-]$, than to the right of the peak.

Three-dimensional evolution of the polar vortex

We now undertake a qualitative investigation of the three-dimensional structure of the vortex during the nonlinear experiments highlighted in Fig. 6.1. Snapshots of vortex evolution in the latter half of the numerical experiments with $h = 0.2H$ and $\Delta_b = 0, -0.18f_0, -0.24f_0, -0.36f_0$ are shown in Fig. 6.2. The three-dimensional representation is generated by shading the outer surface of the PV contours on each layer of the model. In all panels, the peak forcing occurs at $t = 10$ days, before decaying exponentially with time.

In panel A, the tuning parameter is $\Delta_b = 0$, at which the peak in total wave-

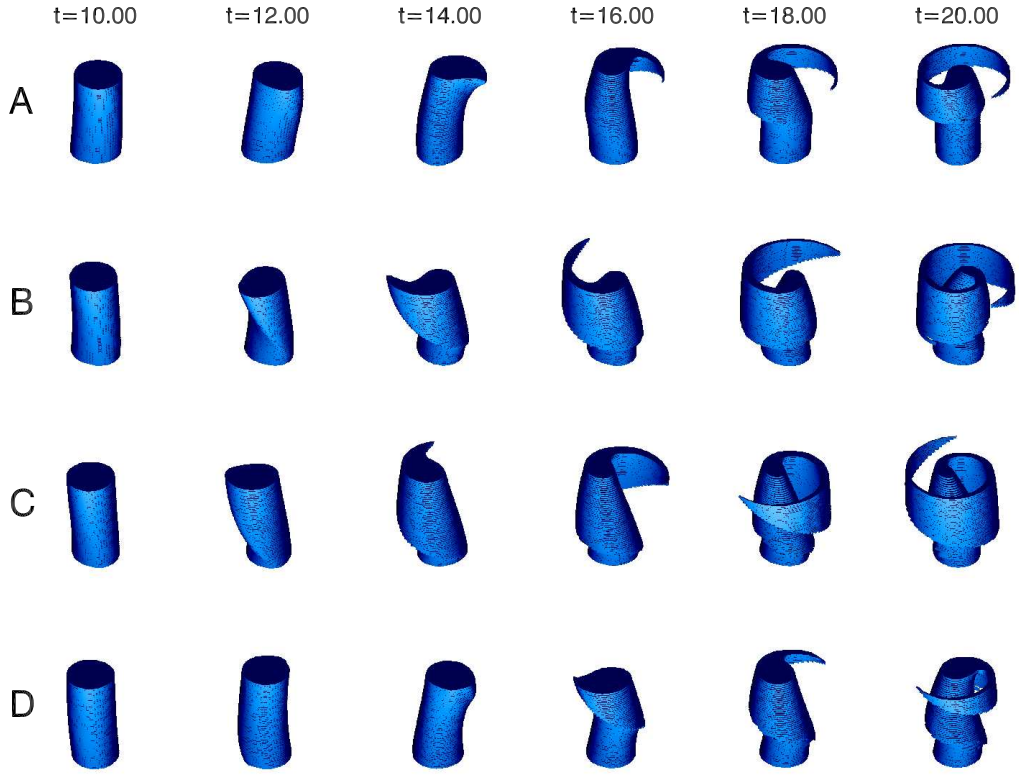

Figure 6.2: Three-dimensional snapshots of the lowest 6 scale heights of the model vortex for experiments with $\tau = 2$ days, topographic forcing scale VS and forcing amplitude $h = 0.2H$. The values of the tuning parameter shown are: (A) $\Delta_b = 0$ (B) $\Delta_b = -0.18f_0$ (C) $\Delta_b = -0.24f_0$ (D) $\Delta_b = -0.36f_0$. Each of these experiments corresponds to one of the highlighted model runs in Fig. 6.1. In all panels, the vortex has been centred around its mass-weighted centroid.

activity occurs for experiments with $h = 0.2H$. At $t = 12$ days, shortly after the peak in topographic forcing has occurred, the vortex is displaced from the pole at all heights, indicating a growth of azimuthal wavenumber-1 disturbances on the vortex edge. The direction of this displacement is approximately the same at all heights of the vortex, suggesting that the dominant mode of the disturbance is the barotropic mode, which is perhaps unsurprising seeing as the tuning parameter $\Delta_b = \Delta_0$ for this experiment. Over the next 8 days up to the end of the experiment at $t = 20$ days, the vortex exhibits moderate wave-breaking towards the top of the vortex, while retaining an almost barotropic structure.

The tuning parameter in panel B is $\Delta_b = -0.18 f_0$, corresponding to the upper limit of the region $[\Delta_+, \Delta_-]$ which marks the tuning parameter range for which baroclinic disturbances to the vortex edge can appear stationary in the linear theory. This choice of Δ_b also coincides with the peak in baroclinic wave-activity for experiments with forcing amplitude $h = 0.2H$. Soon after the peak in topographic forcing has occurred, the direction of the vortex displacement varies with height and the vortex appears slightly twisted, indicating an excitation of baroclinic disturbances with small vertical wavenumber. This is consistent with the choice of $\Delta_b \approx \Delta_-$, for which disturbances with vertical wavenumber $n = 1$ ($m = 0.262$) appear stationary with respect to the topography in the linear theory of section 6.2. As the experiment progresses, the vortex edge becomes increasingly distorted the further up the vortex one looks, with long tongues of PV being stripped away from the main vortex core in the upper half of the vortex.

In panel C, for $\Delta_b = -0.24 f_0$ lying within the Charney-Drazin spectrum, the

vortex behaviour at $t = 12$ days is similar to that observed in panel B, although the twisting of the vortex is more pronounced. By $t = 16$ days, the variation of wavenumber-1 disturbances to the vortex edge with height is such that 'corkscrew' like behaviour is observed. That is, the baroclinic disturbances are dominated by higher vertical wavenumbers than in panel B. Once again, this is consistent with the predictions of linear theory, for which the choice of $\Delta_b = -0.24 f_0$ is expected to produce a resonance of baroclinic disturbances with vertical wavenumber $n = 2$ ($m = 0.523$). In both panels B and C, disturbances in the upper half of the vortex lead to strong erosion of PV from the vortex via a process of filamentation, with wave-breaking being far more pronounced than that observed in panel A, despite the lower values of the total wave-activity for these experiments.

When $\Delta_b = -0.36$ in panel D, the behaviour of the vortex is qualitatively different to that seen in panels B and C. Variation of the displacement of the vortex with height results in the vortex structure becoming 'bowed', with very little sign of wave-breaking on the vortex edge. At later times, disturbances on the vortex edge exhibit higher vertical wavenumber than in any other panel. The vortex appears to wind around its centroid, with more than one vertical wavelength being observed in the disturbance. In the linear theory, this corresponds to the choice of $\Delta_b = -0.36$ leading to a resonance of baroclinic disturbances with vertical wavenumber $n = 13$ ($m = 3.403$).

One important observation from the vortex behaviour in these panels is that the large total wave-activity $\mathcal{F}$ and large disturbances of the vortex from its columnar structure are not necessarily equivalent. Instead, it is the *baroclinic* wave-activity which gives most insight into the occurrence of wave-breaking and overall vortex

disturbance in this instance.

6.4.5 A 'Baroclinic Sudden Warming Event'

Previous investigations of resonant wave growth on a cylindrical vortex (Esler and Scott 2005) have made attempts to simulate observed vortex splitting events. In particular, as part of that study, it was found that for a cylindrical vortex subject to lower boundary forcing of the form (6.24) with $k = 2$, resonant excitation of the barotropic mode of the vortex resulted in a vortex split similar to that observed during vortex-splitting SSWs (for example, see Fig. 2.3 of chapter 2). This simulated vortex-splitting SSW was labelled a "barotropic sudden warming".

In the spirit of the barotropic sudden warming of Esler and Scott (2005), we now investigate resonant excitation of the baroclinic modes of the vortex as a mechanism for a "baroclinic sudden warming" (BCSW hereafter). For the purposes of the current study, we define a BCSW to have occurred if a reversal of the zonal wind occurs at a radius of $r = 4L_R$ and a height $z = 4H$, approximately corresponding to the WMO criterion (1.2). The BCSW described below will feature a vortex breakdown similar to that observed during vortex-displacement SSWs (see, for example, the behaviour of the polar vortex during the composite and January 1987 vortex-displacement SSWs in Fig. 2.7 of chapter 2).

BCSW: experiment set-up

The fully nonlinear numerical model set-up is similar to that detailed in sections 6.4.1 and 6.4.2 with the only difference being that experiments are performed with

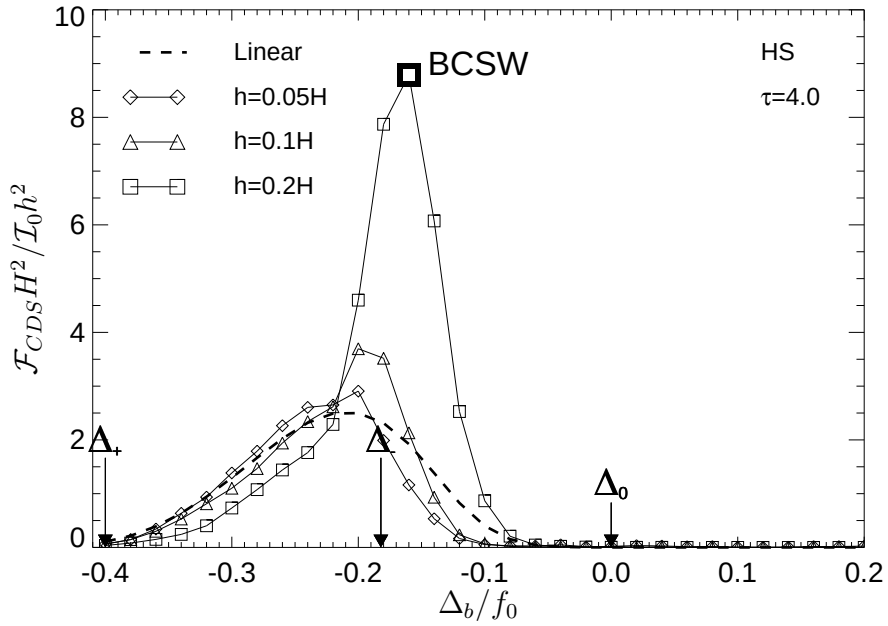

Figure 6.3: The nondimensional baroclinic wave-activity $\mathcal{F}_{CDS}H^2/\mathcal{I}_0h^2$ as a function of the tuning parameter Δ_b (in units of f_0) for the experiments with $\tau = 4$ days and HS topographic forcing scale. The dashed line corresponds to the prediction of the linear model and the symbols represent results obtained from the fully nonlinear model for a selection of forcing amplitudes. The symbol for Δ_0 denotes the value of Δ_b for which $\omega_{0k} = 0$. The range of Δ_b for which $\omega_k(m) = 0$ lies between Δ_+ and Δ_-. The large black bordered box denotes the BCSW experiment.

topographic forcing scale HS ($lR = 0.702$), and time decay parameter $\tau = 4$ in (6.68). The topographic forcing scale HS is chosen, rather than VS scale, as it has a horizontal scale comparable to the zonal wavenumber-1 contribution to the climatological geopotential height field on the tropopause (~ 350 K isentropic surface), which plays a similar role in the winter polar stratosphere to the lower boundary forcing in our model.

In the following sections we discuss the fully nonlinear model experiment with forcing amplitude $h = 0.2H$, and tuning parameter $\Delta_b = -0.16f_0$. This value of Δ_b

gives the peak in baroclinic wave-activity when $h = 0.2H$ (HS, $\tau = 4$), as can be seen in Fig. 6.3. For forcing amplitude $h = 0.2H$, several experiments in the vicinity of $\Delta_b = -0.16f_0$ also exhibited behaviour similar to that described here for the BCSW.

Three-dimensional structure and polar vortex life cycle

Fig. 6.4 shows the evolution of the model vortex during the BCSW. Panel A shows three-dimensional visualizations of the vortex for the lowest six scale heights of the model; panel B shows cross sections of the vortex on model heights $0.85H$, $2.65H$ and $4.45H$; panel C shows the equivalent ellipses corresponding to each patch of vorticity on a selection of model heights. The model layers in panel B correspond approximately to the isentropic surfaces 450 K, 750 K and 1200 K, representative of the lower, middle and upper vortex respectively. The underlying map in all three panels has been introduced so that comparisons can be drawn with the polar vortex behaviour during vortex-displacement SSWs, as seen in Figs. 2.7 and 2.8 (these figures are summarized here in Fig. 6.5 for the composite vortex-displacement case). It is noted that the map edge in panel B is the equator, whereas the map edge in panel C is 45° N. From qualitative observations of the vortex evolution, the BCSW 'onset day'[1], $\tau_0 = 0$, is taken to be at $t = 3$ days, such that the times shown correspond to lagged times $\tau_0 = -5$ days, $\tau_0 = -1$ days, $\tau_0 = +1$ days and $\tau_0 = +4$ days, similar to that shown for the observations in chapter 2.

At $t = -2$ days ($\tau_0 = -5$), the vortex is still relatively undisturbed, with only a slight displacement off the pole and southward along the 90° E meridian while

[1]The 'onset day' as defined in chapter 2, that is, the day on which the Charlton and Polvani (2007) SSW criteria are met in the case of observed SSWs.

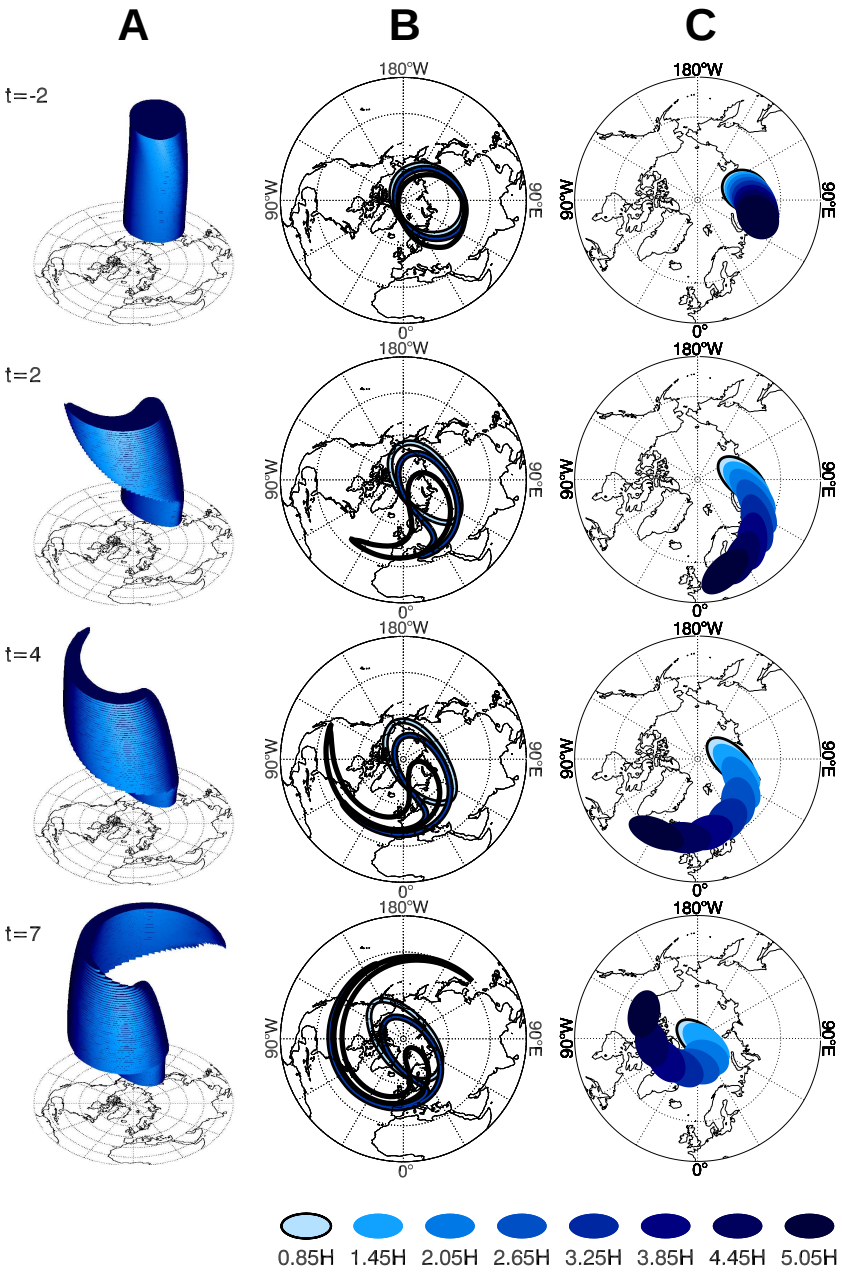

Figure 6.4: Evolution of the model vortex during a baroclinic sudden warming (BCSW) event with $\tau = 4$ days, topographic forcing scale HS and forcing amplitude $h = 0.2H$. Panel A: three-dimensional vortex structure for the lowest six scale heights. Panel B: contours of the vortex edge on model heights $z = 0.85H$, $z = 2.65H$ and $z = 4.45H$, approximately corresponding to the isentropic surfaces 450 K, 750 K and 1200 K in the winter polar stratosphere. Panel C: equivalent ellipses of the model vortex on a selection of model heights. Ellipse areas are scaled to be 0.1 times that of the actual vortex area.

maintaining an overall circular cross section at all heights. There is, however, a slight tilt of the vortex with height, with the centroid of the top of the vortex located at approximately (60° E, 70° N; $5.05H$) in comparison with the centroid of the bottom of the vortex at approximately (90° E, 80° N; $0.85H$).

By $t = 2$ days ($\tau_0 = -1$), the vortex is seen to be highly disturbed above $z = 1.5H$, with the amplitude of the disturbance increasing with height. From panels B and C, it is seen that the aspect ratio of the equivalent ellipse at each height of the vortex has significantly increased from that at $t = -2$ days. Similar to the polar vortex orientation seen in Fig. 2.8 for observed vortex-displacement events, the orientation of the model vortex is approximately north-south along the 45° E meridian at the vortex base, with the vortex orientation increasingly aligning itself with latitude circles as height increases. In addition to the change in orientation of the vortex, there is a displacement of the vortex centroid away from the pole. The centroid of the vortex base remains almost stationary at (90° E, 80° N; $0.85H$), while the vortex centroid is seen to migrate southwards and to the west the further up the vortex one looks, with the uppermost layer shown having approximate centroid (5° E, 55° N; $5.05H$).

By $t = 4$ days ($\tau_0 = +1$), deformation of the vortex has become even more pronounced at model heights above $z = 2.65H$. In comparison, at the lowest heights $z < 2.65H$, disturbances are largely unchanged from those observed at $t = 2$ days. At model heights $z = 4.45H$ and $z = 5.05H$, filamentation of PV from the disturbed vortex edge results in tongues of interior PV being drawn westward and away from the main vortex core, with the filamentation increasing with altitude. For heights above $z = 2.65H$, the equivalent ellipses indicate that the centroid of the vortex is

advected westward, with no further increase in southward displacement. The upper levels of the vortex continue to experience the greatest displacement, with the vortex centroid now located at approximately (40° E, 55° N) at $z = 5.05H$. In contrast, the base of the vortex ($z < 2.05H$) has remained almost stationary in relation to the underlying topography. When referring to the underlying map, the variation of the centroid displacement with height is so large that the centroids of the vortex on the highest and lowest surfaces inhabit completely different hemispheres, which is very similar to the observed displacement of the polar vortex centroid during typical vortex-displacement SSWs (see Fig. 2.8).

As the time increases to $\tau = 27$ days ($\tau_0 = 4$) it is seen that the BCSW event is essentially complete. Wave-breaking at the vortex edge has resulted in filamentation over the majority of the height of the vortex, with only the lowest heights ($z < 2.05H$) retaining a near elliptical cross sectional structure with no filamentation. The three-dimensional visualization and cross sections of the vortex indicate that at the highest levels of the vortex ($z > 3.85H$), the vortex core has been almost entirely eroded by this filamentation, leaving no coherent vortex structure at these heights. Upon completion of the BCSW event, the equivalent ellipses indicate that the overall centroid of the vortex at all heights is displaced back towards the pole. This is accompanied by a continuing westward displacement of the vortex centroid at all heights, with the centroid on levels $z = 3.85H$,$4.45H$,$5.05H$ being displaced by a further 90° W.

Overall, the life cycle of the vortex during the BCSW shares many similarities with that of the polar vortex during vortex-displacement SSWs. Comparison of the

three-dimensional vortex structure seen in panels A and B of Fig. 6.4 with that seen in Fig. 6.5 taken from chapter 2 shows that in both cases, the vortex is initially displaced from the pole before tilting westward with height, with the location of the lower vortex remaining largely unchanged over the vortex evolution.

Zonal winds

Zonal wind fields during the BCSW experiment described above are shown in Fig. 6.6. The SSW definition of Charlton and Polvani (2007), which is in itself partly derived from the major warming definition of the World Meteorological Organization (WMO), demands that the zonal mean zonal winds in the stratosphere ($\bar{u}$) reverse direction, becoming easterly at 60° N on the 10 hPa pressure level. For a BCSW event, a zonal wind reversal at a distance $4L_R$ from the pole at $z = 6H$ is taken to be equivalent to a reversal at 60° N on the 10 hPa surface in the real stratosphere.

Ten days prior to the BCSW onset, at $t = -7$ days, it is seen that the zonal mean velocity is independent of height and westerly, with peak velocities of approximately 40 m s^{-1} at the vortex edge, at a distance of $r = 3L_r$ (2700 km) from the pole. However, by $t = 3$ days, given as the BCSW onset day, there is a significant weakening of $\bar{u}$ at model distances less than $4L_R$ from the pole. The peak $\bar{u}$ in this region is now approximately 15 m s^{-1}, and at $r = 4L_R$ a complete reversal of $\bar{u}$ is evident at heights greater than $z = 3.5H$. At the highest altitudes of the vortex, this reversal extends almost to the pole. By the end of the experiment at $t = 20$ days, $\bar{u}$ values have increased once again in the region $r < 4L_R$, although they are still smaller in magnitude than pre-onset values. At $r = 4L_r$, easterlies are only observed at heights

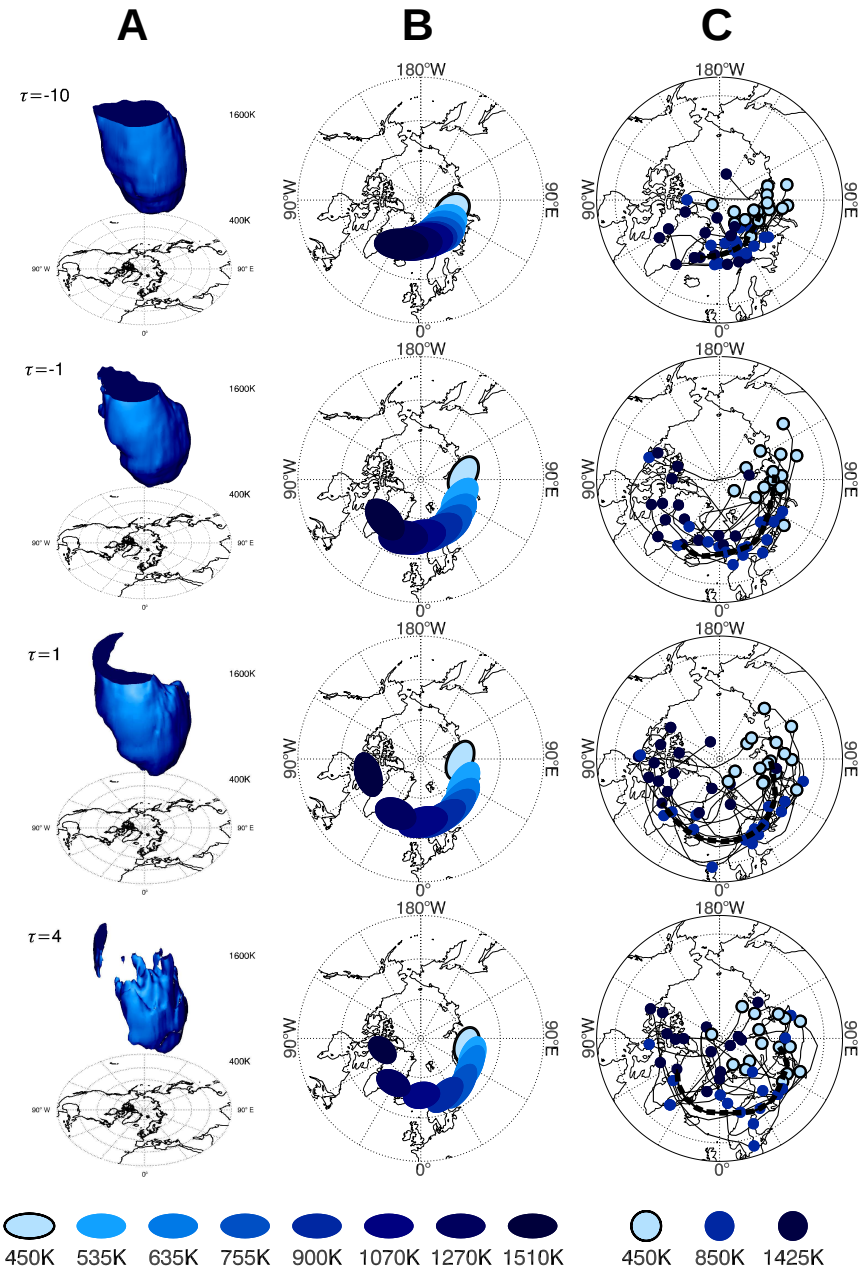

Figure 6.5: Summary of Figs. 2.7, 2.8 and 2.9 of chapter 2, detailing the evolution of the time-lagged composite polar vortex during vortex-displacement SSWs, given here for comparison with Fig. 6.4. Panel A: three-dimensional isosurface of the vertically scaled PV of the composite vortex edge, $Q^* = 38$ PVU. Panel B: equivalent ellipses of the composite PV distribution shown on a selection of isentropic surfaces approximately corresponding to the model scale heights shown in panel C of Fig. 6.4. Panel C: variability of the vortex centroid location during individual vortex-displacement SSWs, for the isentropic surfaces 450 K, 850 K and 1425 K. Times are shown relative to the SSW onset day $\tau = 0$. For further details, please see chapter 2.

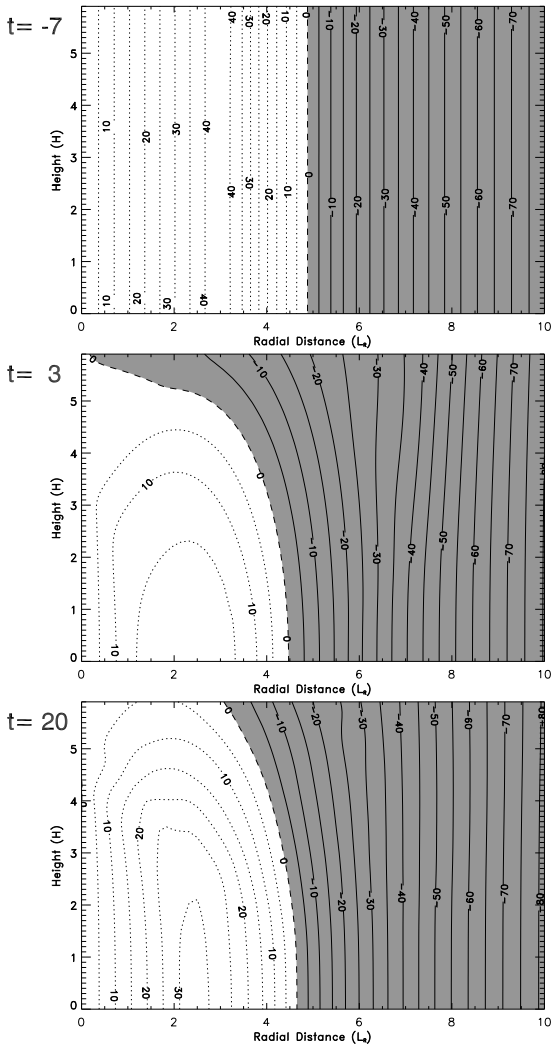

Figure 6.6: Contours of the zonal wind profile as a function of distance from the pole in Rossby radii (L_R) and height measured in scale heights (H). Zonal winds are shown at $t = -7, 3, 20$ days. The topographic forcing has horizontal scale HS and forcing amplitude $h = 0.2H$. Contour intervals are 5 m s^{-1}, with solid lines corresponding to negative (easterly) velocities and dotted lines corresponding to positive (westerly) velocities. Regions where the zonal wind is negative (easterly) are shaded.

above $z = 4L_R$. This persistent decrease in $\bar{u}$ at the highest altitudes may in some part be attributed to the destruction of the coherent vortex structure in this region during the BCSW.

6.5 Conclusions

In this chapter, the response of a columnar vortex structure to a pulse of lower boundary forcing with azimuthal wavenumber-1 structure in a rotating background flow has been investigated using the linear theory of Esler and Scott (2005) along with a fully nonlinear numerical model. Whereas Esler and Scott (2005) looked at forcing with azimuthal wavenumber-2, here we have concentrated on azimuthal wavenumber-1.

The main difference between the azimuthal wavenumber-1 and azimuthal wavenumber-2 behaviours was found to follow from the significance of the barotropic mode in the dynamical behaviour of the vortex. In Esler and Scott (2005), when the barotropic mode of the vortex was resonantly excited by wavenumber-2 forcing of sufficient amplitude a barotropic (columnar) vortex split, strongly resembling an observed vortex split SSW, was found to occur. Here, excitation of the barotropic mode by azimuthal wavenumber-1 forcing leads to a uniform displacement of the vortex from the pole at all heights. It is therefore the baroclinic component of the excitation to the vortex which is of dynamical interest. It has been shown that greatest disturbance of the vortex from its initially columnar state, and associated wave-breaking at the vortex edge, do not coincide with the peak in total wave-activity. Rather, in determining the background conditions leading to such disturbances, it is necessary to find the

peak in the baroclinic wave-activity, which typically occurs as a result of a resonance of baroclinic modes with the smallest vertical wavenumbers $n = 1, 2$.

For experiments in which resonant excitation of baroclinic modes with the smallest vertical wavenumbers $n = 1, 2$ occurs, it has been shown that the three-dimensional vortex behaviour shares many similarities with that seen in the polar vortex breakdown observed during vortex-displacement SSWs.

In these baroclinic sudden warming experiments (BCSW), the centroid of the vortex on each layer was seen to vary strongly with height in a similar way to that observed during vortex-splitting SSWs. Furthermore, by the end of the experiment, wave-breaking at the upper vortex has destroyed the coherent vortex structure in this region, which is once again similar to the destruction of the upper polar vortex during vortex-displacement events, with the vortex base remaining relatively undisturbed throughout the BCSW.

In light of these similarities in behaviour, it is tentatively proposed that a possible mechanism for the onset of vortex-displacement SSWs may be that of a resonance of baroclinic modes of the polar vortex with tropospheric forcing in the presence of favourable stratospheric conditions (similar to the background flow conditions in our idealized model).

Chapter 7

Resonant Responses of Three-Dimensional Model Vortices with Realistic Structure: Vortex-Splitting

7.1 Introduction

In the final chapter of this thesis, self-tuning resonance as a mechanism for SSW-like behaviour is investigated for the most realistic member of our hierarchy of models. As already discussed in chapters 3 and 6, the three-dimensional vortex models of Dritschel and Saravanan (1994), Waugh and Dritschel (1999), Esler and Scott (2005) and Esler et al. (2006) used an ersatz boundary condition on the model lower boundary. By utilizing this ersatz boundary condition, the zero-order Lamb mode

of the system became barotropic, that is independent of height. This property of the zero-order Lamb mode was of particular importance to the analysis of Esler and Scott (2005), as it enables the flow to be separated into a barotropic and baroclinic (dependent on height) component.

Using a similar model and boundary condition, Esler et al. (2006) used the findings of Esler and Scott (2005) to propose that self-tuning resonance was a plausible mechanism for the onset of the southern hemisphere vortex-splitting SSW in September 2002. It was found that resonant behaviour leading to a partial split of the vortex, similar to the observed behaviour during the SSW, occurred in a very narrow band of parameter space in the three-dimensional model.

One complementary study using a more complex dynamical model was that of Polvani and Saravanan (2000). By investigating the behaviour of a topographically forced columnar vortex in a primitive equation model, Polvani and Saravanan found that the vertical structure of the vortex was an important factor controlling wave breaking on the vortex edge. In particular, their findings echoed those of Waugh and Dritschel (1999), in that vortex structures with area increasing with height were resilient to wave breaking at the vortex edge. Vortices with area decreasing with height were, on the other hand, far less resilient to wave breaking at the vortex edge.

The purpose of this chapter is to conduct a study similar to that of Esler et al. (2006) but using greater variety of realistic vortex structures along with the correct lower boundary condition. We restrict our investigations to resonance of azimuthal wavenumber-2 disturbances, in the interests of producing vortex-splitting behaviour. When the correct lower boundary condition (3.29) is used, it was shown in chapter

3 that the associated vertical modes do not include a barotropic mode. That is, the external mode of the vortex has an exponential dependence on height when using the correct lower boundary condition, meaning that in order to study resonance in this more realistic system a modification of the analysis of Esler and Scott (2005) is not straightforward. However, by using a three-dimensional fully nonlinear numerical model in which the correct lower boundary condition is used, this chapter will seek to answer the following questions:

Q1. Does the model reproduce observed vortex splitting behaviour?

Q2. Can vortex-splitting in the model be associated with resonance of a vertical mode?

Q3. To what extent do realistic variations in vortex structure affect the onset and development of vortex-splitting behaviour?

The content of the chapter is summarized as follows. The representation of realistic polar vortex structures is discussed within the framework of the three-dimensional quasi-geostrophic (QG) model with correct lower boundary condition (see chapter 3). A model calculating the wave-speeds of linear disturbances to these structures is then briefly discussed. Using the Contour Advective Semi-Lagrangian numerical model (CASL hereafter, see section 3.3.3 of chapter 3), a selection of realistic polar vortex structures are then chosen to test the hypothesis of Esler and Scott (2005) that, for small topographic forcing heights, the greatest disturbance to the vortex occurs due to a resonance of the external mode of the vortex.

For one of these vortex structures, a more thorough investigation of the vortex

behaviour gives a qualitative view of the role of nonlinearity in self-tuning of the vortex towards, or away from, resonance. The same vortex structure is also investigated for signs of vortex-splitting behaviour similar to that observed during vortex-splitting SSWs. Vortex-splitting behaviour is then shown for a selection of other realistic vortex structures. The conclusions of the chapter are then given.

7.2 Model framework and polar vortex representation

7.2.1 Quasi-geostrophic compressible atmosphere model

The dynamical model used in this chapter is that of the three-dimensional QG compressible atmosphere f-plane subject to the anelastic and hydrostatic approximations. As already shown in equation 3.30 of chapter 3, in this model, by assuming a stratosphere in which the temperature T, buoyancy frequency N, and Coriolis parameter f_0 are constant, the streamfunction at any point in the flow $\psi(r, \phi, z, t)$ is related to the potential vorticity q (PV hereafter) by

$$\nabla_H^2 \psi + \frac{1}{\rho}\frac{\partial}{\partial z}\left(\rho \frac{f_0^2}{N^2}\frac{\partial \psi}{\partial z}\right) + f_0 \;=\; q(\mathbf{x}, t),$$

where $\mathbf{x}$ is a position in cylindrical polar coordinates, ∇_H^2 is the horizontal Laplacian, f_0 is evaluated at the pole and the density is $\rho(z) = \rho_s \exp\{-z/H\}$ where H is a reference log-pressure scale height and ρ_s a reference density. In this chapter, ψ is required to satisfy the 'correct' lower boundary condition (3.29)

$$\frac{\partial \psi}{\partial z} - \frac{\kappa}{H}\psi + \frac{N^2}{f_0} h_T \;=\; 0 \quad \text{on} \quad z = 0, \tag{7.1}$$

where h_T represents topographic forcing, equivalent to forcing from the tropopause. In order to utilize the CASL numerical algorithm, it is convenient to impose an upper lid in the model at $z = D$, with corresponding boundary condition

$$\frac{\partial \psi}{\partial z} - \frac{\kappa}{H}\psi = 0, \quad \text{on} \quad z = D. \tag{7.2}$$

This boundary condition is chosen to match the form of the lower boundary condition (7.1). It is noted however that the exponential decay of density with height means that if D is large enough, the choice of boundary condition at $z = D$ has very little impact on the overall dynamics of the system.

7.2.2 Model representation of the polar vortex

In chapter 6, the polar vortex in the three-dimensional QG f-plane model was represented by a semi-infinite columnar vortex structure. However, in the northern hemisphere there is significant variability in the observed structure of the polar vortex, meaning that the idealized structure in chapter 6 might not necessarily offer the most realistic depiction.

The vertical structure of the polar vortex, as shown in the vertically weighted PV distribution as a function of equivalent latitude and height, is shown for a selection of polar vortex structures prior to observed northern hemisphere SSWs.[1] Panel A closely corresponds to what has so far been referred to as a barotropic vortex, with the location of the vortex edge and the rescaled vortex PV remaining approximately the same at all heights. However, panels B and C show that in some winters, the

[1]The vertically weighted PV here was generated using the methodology presented in chapter 2 using European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-40 data.

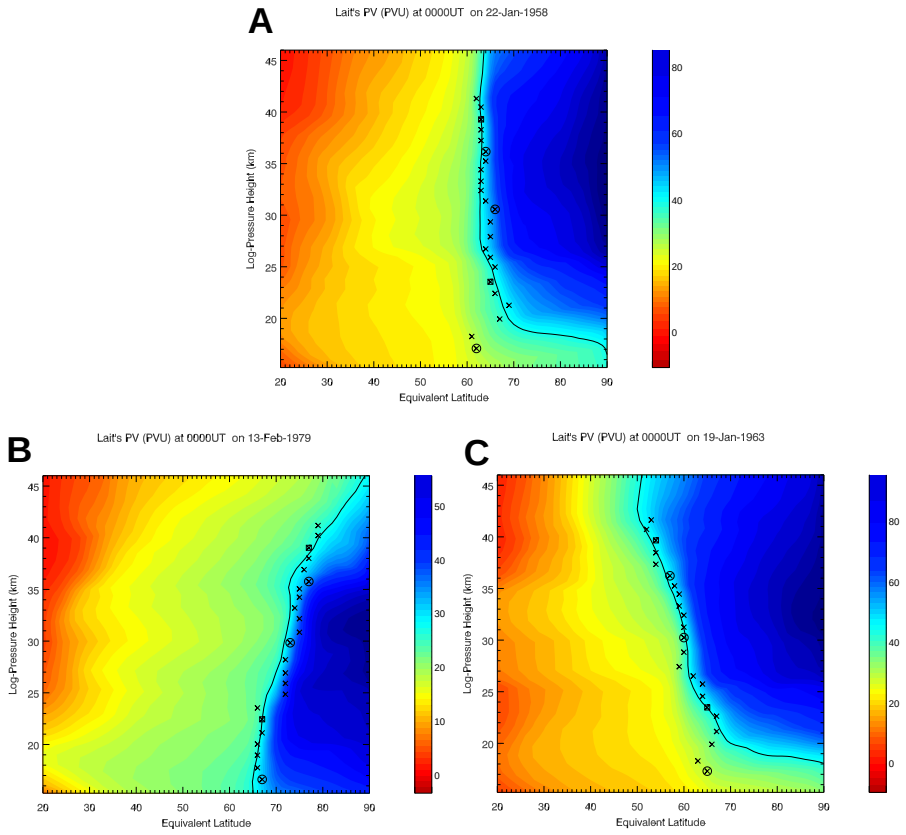

Figure 7.1: Vertically weighted PV Q as a function of equivalent latitude and altitude. A: 0000UTC 22nd January 1958, prior to the January 1958 vortex-splitting SSW, B: 0000UTC 13th February 1979, prior to the 1958 vortex-splitting SSW, C: 0000UTC 19th January 1963, prior to the January 1963 vortex-splitting SSW. Crosses show the location of greatest latitudinal (equivalent) PV gradient on each isentropic surface. Circled crosses represent the locations of the 450 K, 840 K and 1200 K isentropic surfaces. Boxed crosses correspond to isentropic surfaces θ_1 and θ_2 used in the vertical scaling of PV, as given in equation (2.2). The dark lines show the scaled PV isosurface Q^* corresponding to the polar vortex edge calculated using scaling index α where (A) $Q^* = 3.96438e-05$, $\alpha = -3.91947$, (B) $Q^* = 2.74528e-05$, $\alpha = -4.46688$, (C) $Q^* = 4.43907e-05$, $\alpha = -3.58890$.

polar vortex edge location varies considerably with height, with the location moving both poleward (B) and equatorward (C) as altitude increases, when compared with the barotropic vortex of panel A. Observations also show that the rescaled vortex PV can increase or decrease in magnitude with height, while also being generally stronger during some winters and weaker during others. To reflect this observed variability in polar vortex structure prior to SSWs, the shape and strength of the polar vortex which is initialized in our fully nonlinear numerical model will be allowed to vary in a similar way.

Another observation from Fig. 7.1 is that in many cases, the region of high PV corresponding to the polar vortex does not reach down to the polar tropopause, which resides at an altitude of approximately 10 km. Furthermore, in the observed winter polar stratosphere, the strong westerly winds which characterize the polar vortex do not reach far into the mesosphere, being mostly confined to the region between the lower stratosphere and the stratopause. Therefore, consistent with these observations, the polar vortex used in the fully nonlinear numerical model will be of finite height, with the bottom of the vortex starting a short distance above the tropopause and the top of the vortex occurring at the approximate height of the stratopause in the model.

The resulting 'realistic' polar vortex structure in the three-dimensional numerical model, with shape and interior PV varying with height between the lower stratosphere and the model stratospause, is similar to that used by Esler et al. (2006). For each

experiment, the PV q in the model is expressed as

$$q(r,z) = \begin{cases} f_0 + \Delta_b + \Delta(z) & r < R(z) \\ f_0 + \Delta_b & r > R(z), \end{cases} \tag{7.3}$$

where the polar vortex is initialized with radius $R(z)$ and PV anomaly $\Delta(z)$ given by

$$R(z) = \begin{cases} 0 & z < H \\ R\left[1 + s_r\left(z - 3.5H\right)/7.5H\right] & H < z < 6H \\ 0 & 6H < z < D, \end{cases} \tag{7.4}$$

and

$$\Delta(z) = \begin{cases} 0 & z < H \\ \Delta_0(1 + 0.2s_p)\left[1 + s_q 0.75\left(z - 3.5H\right)/2.5H\right] & H < z < 6H \\ 0 & 6H < z < D. \end{cases} \tag{7.5}$$

In these expressions, R_0 and Δ_0 are reference values of the polar vortex radius and PV anomaly respectively, and s_r, s_p, and s_q are parameters with which the shape and strength of the initial vortex structure is varied to reflect the various polar vortex structures observed in the winter polar stratosphere in the northern hemisphere.

In this preliminary study, we investigate vortex structures corresponding to 7 choices of the parameters s_r, s_p, and s_q:

Vortex name	s_r	s_p	s_q	Description
B	0	0	0	Barotropic (control) vortex
U-S	0	0	1	PV increasing with height (Up-Stronger)
U-W	0	0	-1	PV decreasing with height (Up-Weaker)
U-O	1	0	0	Area increasing with height (Up-Outward)
U-I	-1	0	0	Area decreasing with height (Up-Inward)
B-S	0	1	0	Barotropic Strong
B-W	0	-1	0	Barotropic Weak

7.3 Stationary disturbances: linear calculations

Resonance of linear disturbances on the vortex edge occurs when these disturbances appear stationary with respect to the underlying forcing. In order to investigate the conditions leading to such stationary disturbances, it is now useful to review the work presented in section 4 of Waugh and Dritschel (1999), in which a linear dispersion relation was derived for a spectrum of vertically propagating waves on the edge of a three-dimensional vortex in a QG f-plane model.

Discretizing the three-dimensional model into L vertical layers, by defining R_j to be the reference vortex radius on layer j, linear wave-like disturbances to the vortex edge can be expressed as $r_j = R_j + \eta_j(\phi, t)$, where

$$\eta_j(\phi, t) \quad = \quad A_j(t) \exp\{ik(\phi - \omega)t\}.$$

It is now convenient to express the azimuthal and radial velocities on each layer as discrete transforms using the orthogonal eigenfunctions χ_{ln} as a basis, where χ_{ln}

correspond to the discrete forms of the normal modes $\chi_0(z)$ and $\chi_n(z)$ given in chapter 3, with $\chi_{ln} = \chi_n(z_l)$. By first defining the function

$$X^k_{jl}(r_1, r_2) \;=\; \rho_l \Delta_l \frac{r_1}{r_2} \sum_{n=0}^{L-1} \chi_{ln}\chi_{jn} dz_l \begin{cases} I_k(\gamma_n r_1)K_k(\gamma_n r_2) & r_1 < r_2 \\ K_k(\gamma_n r_1)I_k(\gamma_n r_2) & r_1 > r_2, \end{cases}$$

at leading order, the azimuthal and radial velocities on the jth layer, $u_j = \bar{u}_j + O(\eta)$ and $v_j = v'_j \exp\{ik(\phi - \omega t)\} + O(\eta^2)$ are

$$\bar{u}_j \;=\; \sum_{l=1}^{L} r X^1_{jl}(R_l, r),$$

$$v'_j \;=\; ik \sum_{l=1}^{L} A_l X^k_{jl}(R_l, r).$$

The background PV parameter Δ_b is now added to the PV in the system, equivalent to adding a solid body rotation to the flow of angular velocity $\Omega_b = \Delta_b/2$. This additional PV gives an azimuthal velocity contribution u_{bj} when calculating the flow at the vortex edge on layer j.

On each layer, the linearized kinematic condition on the vortex edge is then written

$$\left(\frac{\partial}{\partial t} + \frac{(\bar{u}_j + u_{bj})}{r}\frac{\partial}{\partial \phi}\right)\eta_j - v_{1_j}\frac{\eta_j}{A_j} \;=\; 0 \quad \text{on} \quad r = R_j,$$

which becomes

$$\left[\sum_{l=1}^{L} X^1_{jl}(R_l, R_j) + \frac{u_{bj}}{R_j} - \omega\right] A_j - \sum_{l=1}^{L} A_l X^k_{jl}(R_l, R_j) \;=\; 0. \tag{7.6}$$

Equation (7.6) corresponds to the jth row of an $L \times L$ matrix equation, the solution of which yields L eigenvalues ω_l where $1 \leq l \leq L$. These ω_l are the propagation frequencies of disturbances with vertical structure corresponding to each of the L vertical modes of the vortex.

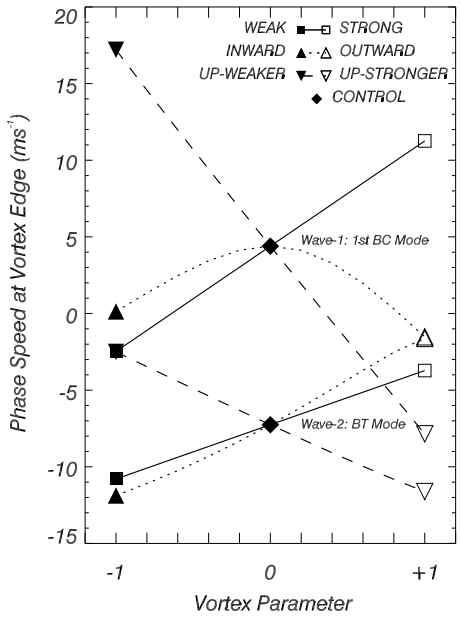

Figure 7.2: Phase speed of linear disturbances on the vortex edge in the three-dimensional QG model with correct lower boundary condition. In the case of weak/strong (B-W/B-S) vortex configurations, the vortex parameter corresponds to s_p, in inward/outward (U-I/U-O) s_r, and in up-weaker/up-stronger (U-W/U-S) s_q. The 'BT mode' represents the phase speed of the wave associated with the external mode of the vortex when $k = 2$, and the '1st BC mode' the phase speed associated with the first vertical mode of the Charney-Drazin spectrum when $k = 1$.

As seen in chapters 4 and 5, in order for one of these disturbances to be stationary with respect to the fixed frame, the corresponding frequency of the free-travelling wave, ω_i, must be zero. Therefore, similar to the approach already seen in section 5.3 of chapter 5, by setting ω_i to zero, the matrix equation yields a corresponding value of Δ_b for which this wave is stationary. This value of Δ_b is the value at which resonant excitation of disturbances associated with the ith vertical mode is expected to occur.

When $\Delta_b = -0.1 f_0$, the phase speed associated with the frequency ω_i is shown in

figure 7.2.[2] For the various vortex structures given in table 7.2.2, the phase speeds are seen to vary significantly for both the external mode when $k = 2$ and and the first mode of the Charney-Drazin spectrum when $k = 1$. In the $k = 1$ case, a strengthening of the vortex PV with height reverses the direction in which disturbances travel around the vortex boundary, when compared to a vortex in which the PV weakens with height.

7.4 Nonlinear response of QG polar vortices to topographic forcing

7.4.1 Lower boundary forcing and vortex conditions

Fully nonlinear numerical experiments were performed using the CASL algorithm. The reference values for the vortex structure parameters in section 7.2.2 were $R_0 = 3L_R$ and $\Delta_0 = 0.7f_0$. The lid of the model is imposed at height $D = 12H$, chosen to be large enough that wave reflection and other unrealistic effects due to the upper boundary are minimized.

The stratosphere in the model is initialized with parameters $H = 6.14$ km, $N = 0.02$ s^{-1}, $L_R = 900$ km, $f_0 = 4\pi$ day^{-1}, with $\kappa = 2/7$ used in the correct lower boundary condition (3.29). The maximum time in all model runs is $t = 40$ model time units.

Forcing from the tropopause is represented in the model by a constant topographic

[2]Figure kindly provided by Dr J. G. Esler, and used with permission.

forcing of the form

$$h_T(r,\phi) \;=\; \frac{h_0}{2} H J_2(lr) \exp\{2i\phi\}, \tag{7.7}$$

where, as already stated, only azimuthal wavenumber-2 is used for the topographic forcing in an attempt to excite vortex-splitting behaviour. The radial scaling parameter l is given by $lR_0 = 1.162$, for which the peak of the topographic forcing occurs at a distance of $r = 7214$ km from the pole.

The background PV Δ_b acts as tunable parameter with which the vortex can be brought into and out of resonance with the topographic forcing.

7.4.2 Model details and numerical parameters

The CASL model used for the fully nonlinear numerical experiments consists of a cylindrical domain with outer walls at a distance $30L_R$ from the origin. The domain is discretized into 120 vertical layers of uniform depth. For the spectral velocity calculations, contour node locations are interpolated onto a quadratically spaced radial grid with a resolution of 192 azimuthal grid points and 96 radial grid points. Quadratic grid spacing is chosen such that grid points are concentrated in the inner domain, the region in which vortex advection primarily occurs. The node resolution parameter in the contour dynamics algorithm is $\mu = 0.18$, such that each PV contour in the "B" set of experiments has initially 51 nodes.

7.4.3 Measures of wave-activity and vortex moment diagnostics

The vertically integrated wave-activity $\mathcal{A}$ is calculated as described in section 6.4.3 of chapter 6.

As the forcing amplitude in equation (7.7) remains constant for each experiment, the physical interpretation of $\mathcal{A}_\infty$ is not the same as seen in chapter 6, in which a pulse of topographic forcing was considered. Therefore, in an approach which is similar to the maximum amplitude diagnostic of chapters 4 and 5, it is enlightening to interpret the wave-activity of the vortex using $\mathcal{A}_{max} = \max[\mathcal{A}]$, which gives the maximum amplitude of disturbances to the vortex over the experiment.

Vortex-splitting is diagnosed using the parameter κ_4 as derived in section 2.2.4, with vortex moments in the model being calculated as described in section 6.4.3. For each height of the vortex, vortex-splitting is diagnosed if $\min[\kappa_4] = \kappa_{min} < -0.6$. Experiments are classified as "vortex-splitting experiments" if $\kappa_{min} < -0.6$ on model layer $z = 2.5H$, which would approximately correspond to a split of the vortex on the 840 K isentropic surface in the stratosphere which is used in the SSW definition of Charlton and Polvani (2007).

7.4.4 Results from the three-dimensional numerical model

Wave-activity and vortex splitting

Fig. 7.3 shows the maximum vertically integrated wave-activity $\mathcal{A}_{max}$ for fully nonlinear numerical model experiments initialized with the vortex structures in table

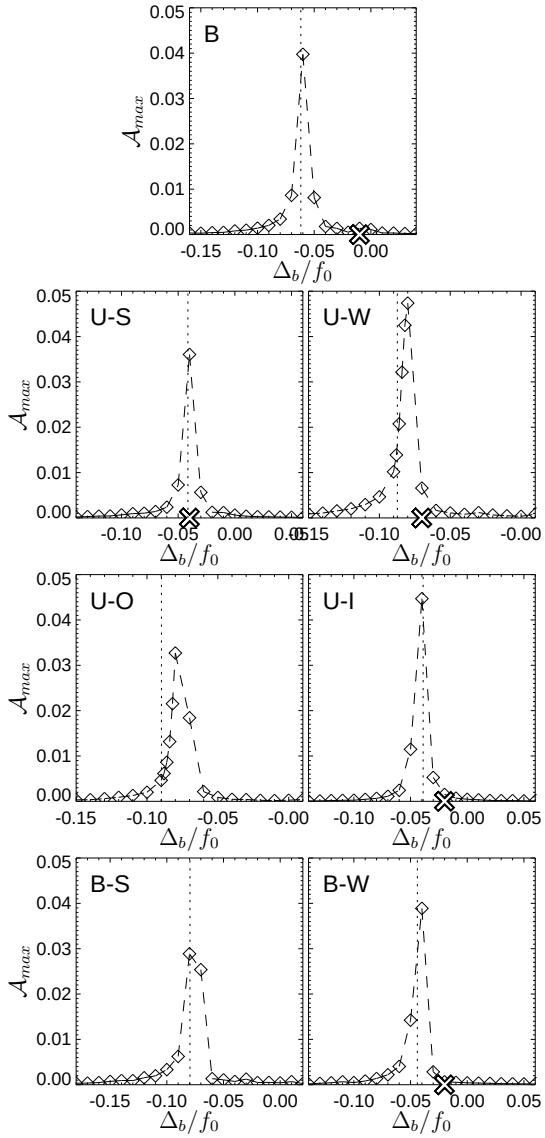

Figure 7.3: The maximum vertically integrated wave-activity $\mathcal{A}_{max}$ (dashed lines and diamonds) as a function of tuning parameter Δ_b in the numerical model when $h_0 = 0.01$. Panels correspond to the vortex structures given in table 7.2.2. The Δ_b values for which disturbances associated with the external mode of the vortex become linearly resonant with the topographic forcing are marked by dotted lines. Experiments shown in Fig. 7.6 and panel B of Fig. 7.5 are marked on the Δ_b-axis by crosses.

7.2.2. For each set of experiments, the dotted line shows the tuning parameter Δ_b at which linear resonance of the external mode of the vortex with the topographic forcing occurs.

In most cases, the location of the peak in $\mathcal{A}_{max}$ coincides with the dotted line, indicating a resonant excitation of disturbances associated with the external mode of the vortex by the topographic forcing. The response in the vicinity of the dotted line is far larger than at other values of Ω_b, implying that resonant excitation of the external mode dominates over the excitation of disturbances associated with modes in the Charney-Drazin spectrum.

For experiments with initial vortex structure "B", contours of the maximum vertically integrated wave-activity A_{max} are shown in Fig. 7.4 as a function of tuning parameter Δ_b and topographic forcing height h_0. The overall appearance of the figure is highly reminiscent of the contours of maximum disturbance amplitude seen in chapter 5 (see, for example, Fig. 5.10). For small forcing amplitudes, the location of the peak response in $\mathcal{A}_{max}$ occurs at $\Delta_b \approx -0.062 f_0$ as already shown in Fig. 7.3. As the forcing amplitude increases, the increasing nonlinearity of the disturbance is observed as a migration of the peak in A_{max} to the right.

Vortex-splitting experiments are denoted by the shaded region. As seen in the single layer experiments of chapters 4 and 5 (see Figs. 4.7, 5.10 and 5.12), the vortex-splitting sub-regime has a distinctive wedge shape, and coincides with the region in which $\mathcal{A}_{\max}$ is greatest. The left-hand boundary of the sub-regime occurs at $\Omega_b = 0.03$ at all forcing heights, whereas the right-hand boundary follows the the steep gradients in $\mathcal{A}_\infty$, sloping to the right. Also similar to the single layer models is the fact that

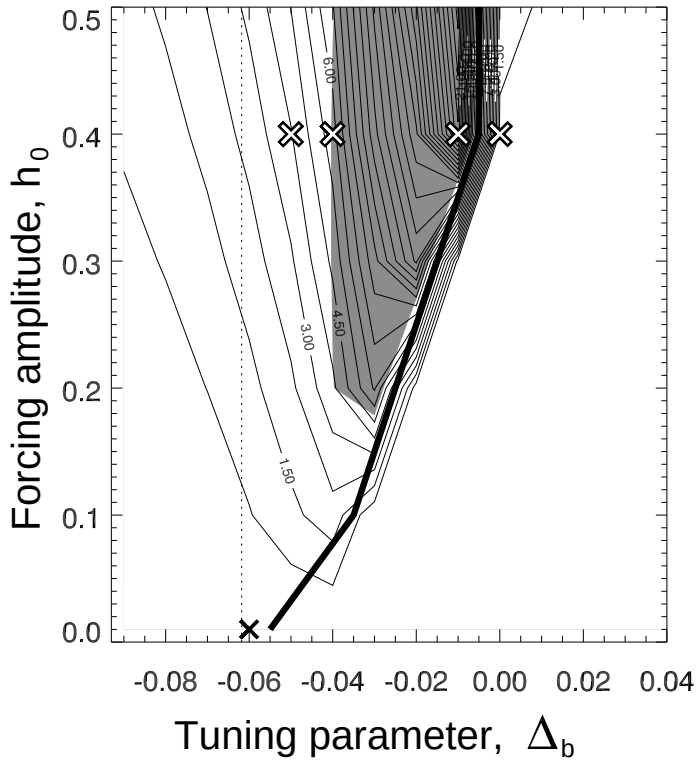

Figure 7.4: Contours of the maximum in vertically integrated wave-activity $\mathcal{A}_{max}$ as a function of tuning parameter Δ_b and forcing height h_0 for numerical experiments initialized with the vortex configuration "B" (barotropic vortex). Outlined white crosses correspond to experiments which are shown in Fig. 7.5. The black cross shows the location of the peak in A_{max} when $h_0 = 0.01$, and the vertical dotted line gives the location of $\Delta_b = -0.062 f_0$ at which disturbances associated with the external mode of the vortex are in linear resonance with the topographic forcing. The heavy solid line marks the steepest gradient in $\mathcal{A}_{max}$ with respect to Δ_b. The shaded region denotes the vortex-splitting sub-regime in which $\min[\kappa_4] < -0.6$ on model height $z = 2.5H$. Contours are shown at intervals of 0.1.

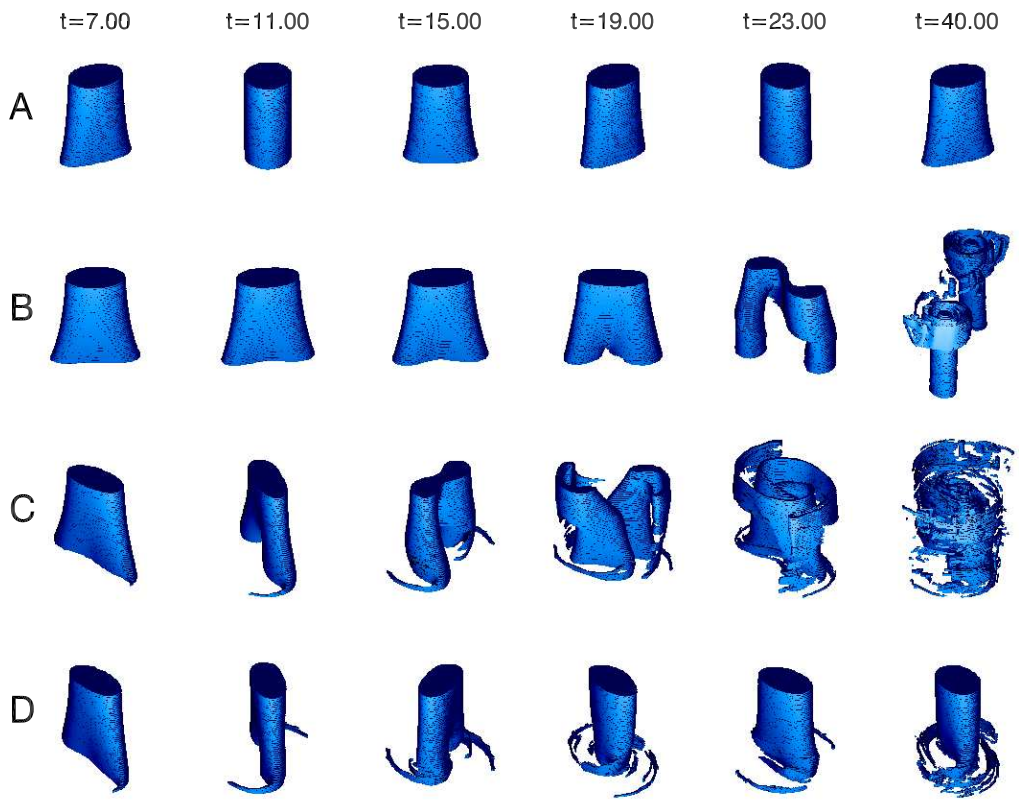

Figure 7.5: Snapshots of vortex evolution at model times $t = 7, 11, 15, 19, 23, 40$ during numerical experiments initialized with the 'control' vortex configuration ($s_r = s_p = s_q = 0$). The forcing amplitude is $h_0 = 0.4$ in all panels. Panel A: $\Delta_b = 0.0$. Panel B: $\Delta_b = -0.01$. Panel C: $\Delta_b = -0.04$. Panel D: $\Delta_b = -0.05$.

splits only occur for sufficiently large forcing amplitudes, in this case when $h_0 \gtrsim 0.2$.

The remarkable similarities between Fig. 7.4 and Figs. 5.10 and 5.12 show that the barotropic single layer model offers excellent insight into vortex-splitting dynamics in an idealized three-dimensional stratosphere.

Vortex structure and evolution

Three-dimensional views of evolution of the polar vortex boundary during experiments initialized with vortex configuration "B" are shown in Fig. 7.5. Each panel

(top-bottom) corresponds to one of the outlined white crosses in Fig. 7.4 (right-left). In panel A, disturbances to the vortex edge result in periodic, anticlockwise rotating vortex evolution, with disturbances giving the vortex cross sections an approximately elliptical shape. The aspect ratio of these cross sections varies periodically at all heights of the vortex. Aspect ratios of the quasi-elliptical cross sections are generally larger at the base of the vortex and smaller in the upper half of the vortex. In addition to this periodic evolution, there is no evidence of either filamentation or pinching type instabilities on the vortex edge at any height.

A small change in the tuning parameter, moving into the vortex-splitting sub-regime shown in Fig. 7.4, results in a dramatic change in vortex evolution as shown in panel B. Between $t = 7$ and $t = 15$, disturbances to the vortex edge appear almost stationary, with deformation of the vortex resulting in almost elliptical vortex cross sections, the aspect ratio of which increases over time. The greatest aspect ratios of quasi-elliptical cross sections are once again observed in the lower vortex. By $t = 19$ the vortex is showing signs of splitting behaviour in the lower vortex, which by $t = 23$ has resulted in a split of the vortex at all heights into two approximately columnar daughter vortices. The daughter vortices remain coherent to much later times, such that by $t = 40$, both vortex columns are relatively unchanged from their state immediately after the split, with only mild erosion of PV at high altitudes.

Another example of vortex behaviour lying in the vortex-splitting sub-regime is observed in panel C. For $\Delta_b = -0.04$ lying on the left hand boundary of the vortex-splitting sub-regime, vortex-splitting occurs at earlier times than for $\Delta_b = -0.01$, which was shown in panel B. The orientation of the vortex with respect to the under-

lying topography is not fixed between $t = 7$ and $t = 15$, with the vortex exhibiting clockwise rotation. In contrast to the coherence of the daughter vortices observed in panel B, at $t = 23$, soon after the split has occurred, interaction of the daughter vortices with the background flow and one another quickly acts to destroy both structures. At $t = 40$, long after the vortex split, the remnants of the daughter vortices are seen to have merged again, reforming into a single columnar vortex structure, as can be seen by examining PV cross sections on each model level. Further investigation of vortex cross-sections during the split in both panels B and C reveals that the orientation or the vortex with respect to the underlying topography, and location of each of the daughter vortex centroids, is the same at all heights of the vortex.

For Δ_b just to the left of the vortex-splitting sub-regime as shown in Fig. 7.4, the vortex evolution does not exhibit vortex-splitting behaviour at all. Instead, the orientation of the vortex oscillates with respect to the underlying topography, while instability of disturbances leads to filamentation of PV from the lower vortex.

From these results it is clear that the stability and qualitative behaviour of these vortex structures, when compared to the location of each experiment in parameter space as shown in Fig. 7.4, is remarkably similar to that observed in the single layer models (see, for example, Figs. 4.2, 4.6 and 5.14). Three-dimensional experiments with Δ_b to the right of the steepest gradients in $\mathcal{A}_{max}$, as marked by the heavy solid line in figure 7.4, are characterized by anti-clockwise vortex rotation, with very little indication of instability. In comparison, for model experiments with Δ_b lying to the left of the steepest gradient in $\mathcal{A}_{max}$, vortex evolution is either oscillatory or clockwise rotating and disturbances to the vortex structure are far larger. Also, experiments

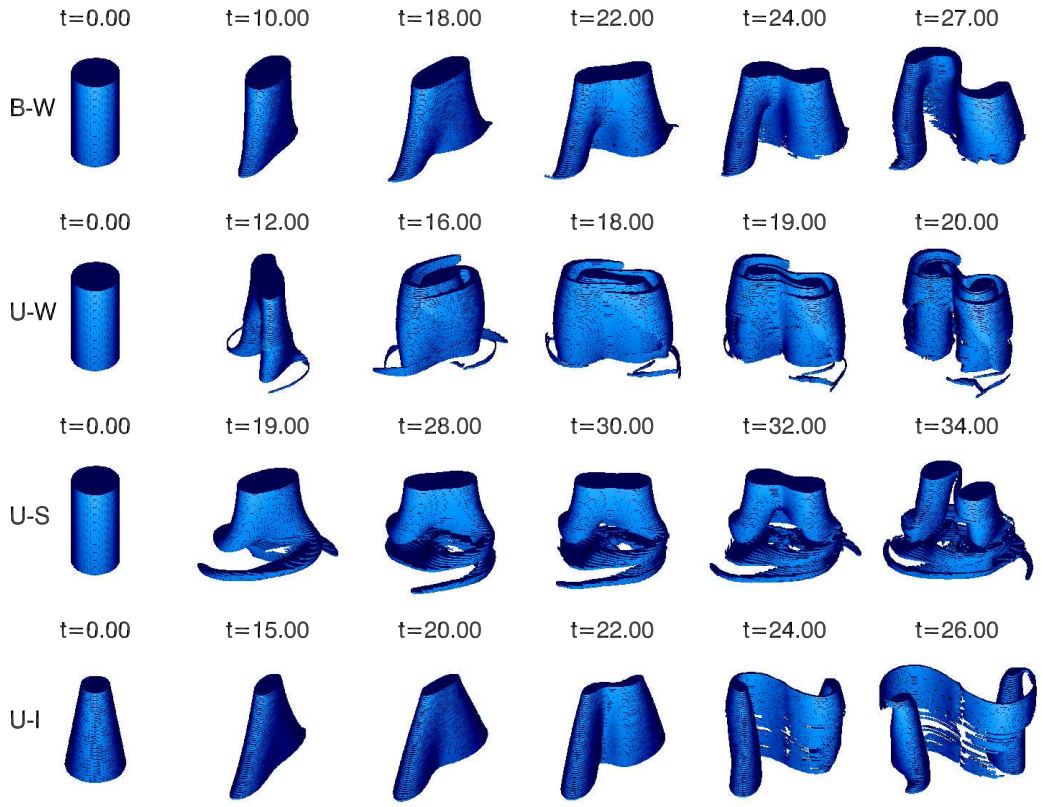

Figure 7.6: Snapshots of vortex evolution during vortex-splitting experiments:

B-W: $h_0 = 0.2$, $\Delta_b = -0.02$
U-W: $h_0 = 0.4$, $\Delta_b = -0.07$
U-S: $h_0 = 0.4$, $\Delta_b = -0.04$
U-I: $h_0 = 0.2$, $\Delta_b = -0.02$

Times are shown as model times.

in this region of parameter space are unstable. This is similar to the findings for the single layer topographically forced model of chapters 4 and 5, the stability of which was in turn predicted by the stability of the Kida vortex model in a purely two-dimensional system (e.g. Dritschel 1990).

We now demonstrate that vortex-splitting behaviour, similar to that observed in panels B and C of figure 7.5, also occurs for a range of non-barotropic vortex structures.

Fig. 7.6 shows evolution of the model vortex boundary for a choice of experiments with different initial vortex structures (see table 7.2.2). For all vortex structures, splitting of the vortex can occur when Δ_b is greater than the Δ_b at which linear resonance occurs, as shown by the relative locations of the crosses and dotted lines in Fig. 7.3. In all experiments shown in Fig. 7.6, the vortex undergoes a process of elongation of its quasi-elliptical cross sections, followed by a pinching of the vortex at its centroid. This pinching then leads to a complete split of the vortex at all heights. It is pointed out that in all cases, the narrow filaments of PV which connect the columnar daughter vortices have very little impact on the subsequent dynamics, which are dominated by the daughter vortex structures.

While vortex-splitting was observed for all of the vortex structures in table 7.2.2, the minimum forcing amplitude necessary for vortex-splitting behaviour to be observed varied between model runs. That is, to generate vortex-splitting experiments, model runs U-S and U-W required forcing amplitudes $h_0 = 0.4$, approximately double that of model runs B-W and U-I in which splitting occurred for $h_0 = 0.2$. The most resilient vortex structures were B-S and U-O, which required the largest forcing amplitudes ($h_0 = 0.6$) for vortex-splitting to be observed. This result for U-O experiments is in agreement with Waugh and Dritschel (1999) and Polvani and Saravanan (2000), who found that vortex structures with area increasing with height were most resilient to lower boundary forcing.

Comparisons with the January 1958 SSW

Although it must be stressed that this work is only a preliminary step in explaining the role of a resonant excitation mechanism in the generation of realistic SSWs, it is nevertheless of great interest to compare the idealized model results with observed vortex-splitting events.

The barotropic structure of the control vortex in the fully nonlinear model is qualitatively similar to that observed in the polar vortex prior to the January 1958 vortex-splitting SSW, as shown by the vertical structure of the vertically weighted PV for this event (see panel A of Fig. 7.1). Therefore, simple comparisons are now made between vortex-splitting behaviour in the fully nonlinear numerical model and vortex-splitting behaviour observed during the January 1958 vortex splitting SSW.

Left-hand panels of Fig. 7.7 show cross sections of the mid-vortex during the numerical model experiment shown in panel A of Fig. 7.5. Right-hand panels show vertically weighted PV cross sections on the 840 K isentropic surface during the January 1958 vortex-splitting SSW. The 840 K isentropic surface is chosen due to its role as the reference level on which SSWs are identified in both the World Meteorological Organization and Charlton and Polvani (2007) SSW definitions. It is pointed out that the map has been added to the model results in the left-hand panels for reference purposes only.

Vortex behaviour during the numerical experiment mirrors that of the observed SSW remarkably well. In both panels, the vortex is initially seen to be approximately elliptical, with slight deformation over the pole. Over the next two days, pinching of the vortex over the pole becomes more extreme, such that by $t = 20$ in the model (28

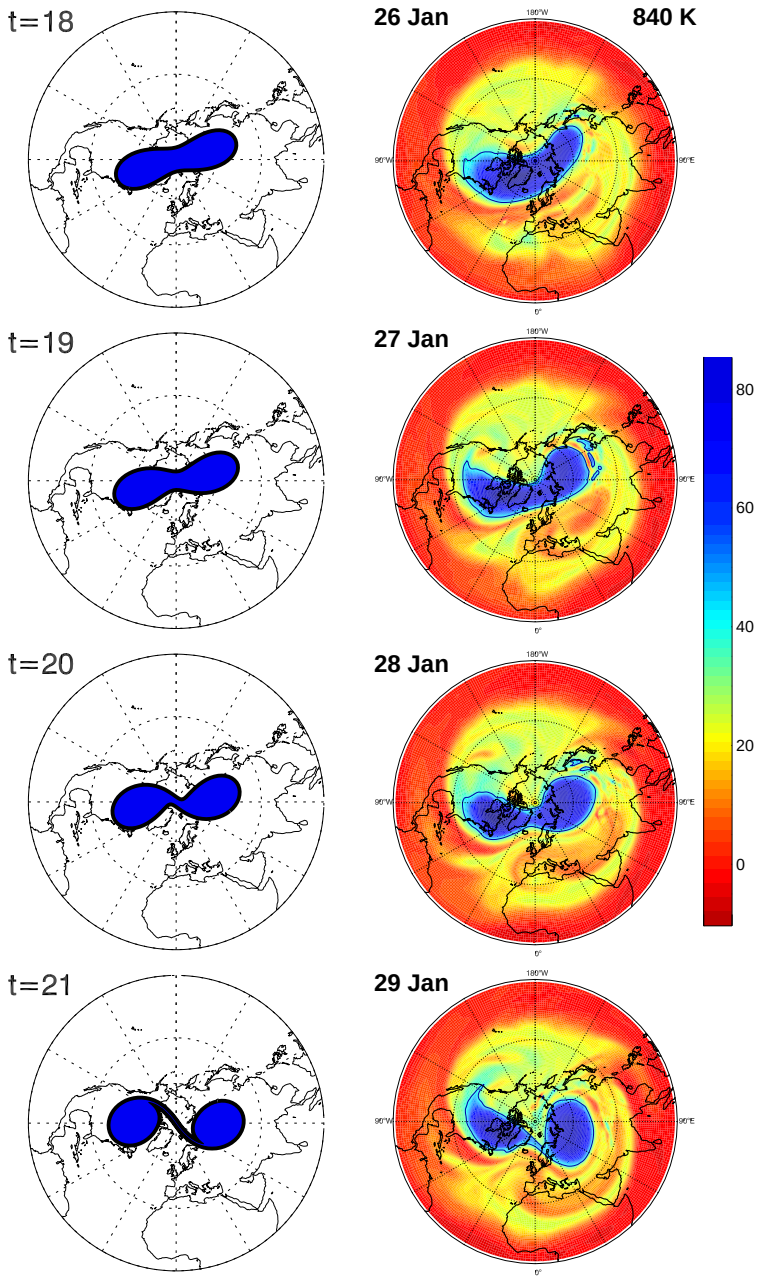

Figure 7.7: Left panels: filled contours of the vortex boundary at model times $t = 18, 19, 20, 21$ days for the numerical model experiment shown in panel B of figure 7.5 with $(\Omega_b, h_0) = (0.09, 0.4)$. The model height shown is $z = 2.5H$. Right panels: contours of vertically weighted PV (PVU) on the 840 K isentropic surface, corresponding approximately to model height $z = 2.5H$, during the January 1958 vortex-splitting SSW.

January in the observations), the vortex consists of two diametrically opposite lobes of PV (located over Siberia and Canada in the observations) connected by a narrow filament of PV. After one more day has passed, in both the numerical model and the observations, the vortex has split into two smaller daughter vortices of comparable magnitude, with only a narrow filament of PV connecting the two. In both cases, there is very little change in the orientation of the vortex over the duration of the vortex split, implying that disturbances leading to elliptical deformation of the vortex are stationary with respect to the topography.

7.5 Conclusions

Below, the three questions posed in the introduction to the chapter are addressed.

Does the model reproduce observed vortex-splitting behaviour

Using a fully nonlinear numerical model of the stratosphere in which a realistic lower boundary condition is used, it has been shown that vortex-splitting behaviour is observed for a range of vortex structures. Comparison of one vortex-splitting experiment with the vortex-splitting SSW occurring in January 1958 showed that the qualitative nature of the split, and the time scale on which the split occurs, are similar between the two.

The qualitative features of vortex-splitting in the three-dimensional model are also highly reminiscent of vortex-splitting behaviour seen in the single layer model in chapter 5. That is, growth of disturbances on the vortex edge leads to vortex cross sections taking on an elongated elliptical shape. For model parameters leading to

large aspect ratios in these elliptical cross sections, unstable growth of disturbances is observed on the vortex edge. For the greatest aspect ratios, this unstable growth is dominated by elliptic azimuthal wavenumber-2 disturbances, leading to vortex-splitting (see section 4.2 of chapter 4). The results of this preliminary study suggest that, in most cases, vortex-splitting behaviour occurs almost simultaneously at all vortex heights.

Can vortex-splitting in the model be associated with resonance of a vertical mode?

By investigating wave-activity response to small amplitude forcing in the three-dimensional model, it has been shown that for a range of vortex structures, the location of greatest response in the model parameter space is close to that predicted by a resonance of the external mode of the vortex. That is, disturbance of the vortex resulting from resonant excitation of the vertical modes lying in the Charney-Drazin spectrum is far smaller than that of the external mode.

When considering larger amplitudes of topographic forcing, resonant excitation of disturbances associated with the external mode of the vortex lead to vortex-splitting. In contrast, resonant excitation of disturbances associated with the vertical modes of the Charney-Drazin spectrum seem to lead to filamentation of the vortex.

This suggests that the analytical predictions of Esler and Scott (2005), that for a semi-infinite barotropic vortex and ersatz lower boundary condition vortex-splitting is generated by resonant excitation of the barotropic mode, are directly relevant when considering finite non-barotropic vortex structures with realistic lower boundary

condition.

A thorough investigation into the peak response of a barotropic vortex to topographic forcing has shown that nonlinearity acts to tune the vortex either towards, or away from, resonance. As forcing amplitude increases, the location of the tuning parameter Δ_b at which the greatest response occurs migrates away from that predicted by linear theory. This migration is remarkably similar to that observed in the single layer model in chapter 5 when $\mathcal{B} = 1.44$.[3]

Furthermore, in the parameter space, the regime boundary between stable, periodic vortex motion and vortex-splitting events tracks the migration of the location at which resonance occurs. This, once again, is remarkably similar to the results for the single layer model of chapter 5, in which similar behaviour was observed.

To what extent do realistic variations in vortex structure affect the onset and development of vortex-splitting behaviour?

Resonant excitation of the external mode of the vortex has been shown to lead to vortex-splitting behaviour for a wide range of realistic vortex structures. That is, the initial structure of the vortex seems to have little to no effect on the potential for vortex-splitting behaviour. However, some vortex structures are more resistant to splitting than others. In general, vortex structures which have anomalously weak PV, or which have area decreasing with height, require smaller amplitude forcing for splits to be observed, when compared to vortex structures which have anomalously

[3]This choice of $\mathcal{B} = 1.44$ corresponds approximately to the Rossby radius of the external mode when using the correct lower boundary condition.

strong PV or which have area increasing with height. In contrast, variation of the vortex PV with height seems to have little impact on the forcing amplitudes required for vortex-splitting.

Chapter 8

Conclusions

The objective of this dissertation has been to establish a greater understanding of the fundamental dynamics responsible for polar vortex breakdown during major midwinter stratospheric sudden warmings (SSWs). The primary focus has been on a hierarchy of models approach using idealized dynamical models of the winter polar stratosphere, with a secondary focus on the observed behaviour of the polar vortex during all Northern Hemisphere SSWs since 1958. Due to the pivotal role of the potential vorticity (PV) in determining the dynamical behaviour of the winter polar stratosphere, a vortex dynamics approach has been adopted in both the choice of dynamical models and the techniques used in our observational data analysis.

Conclusions relating to the observational and modelling aspects of this work will now be summarized, with a final comment on future avenues of research which stem from these conclusions.

8.1 Observational study

An observational study, using European Centre for Medium Range Weather Forecasts (ECMRWF) ERA-40 re-analysis data, focused on the evolution of the Arctic polar vortex during SSWs. Time-lag composites and PV moment diagnostics were used to investigate the distinctive behaviour of the polar vortex during both displacement and splitting SSW events.

The motivating questions and their respective conclusions, as provided by our results, are now summarized.

1. What is the vertical structure of the Arctic polar vortex during vortex-displacement and vortex-splitting SSWs? Is there significant variation in vertical structure between individual SSWs within these classifications, and if not, does the typical vertical structure for displacement SSWs differ to that of splitting SSWs?

The Arctic polar vortex has a highly characteristic vertical structure during the two different classifications of SSW. Vortex-displacement events are highly baroclinic in nature, with vortex displacement off the pole and elongation of the vortex both increasing with altitude above 30 km. In contrast, during vortex-splitting events the vortex remains nearly barotropic, with the split occurring near-simultaneously over a large altitude range (at least 20-40 km).

2. Is there a preferred location and orientation of the Arctic polar vortex with respect to the Earth's surface during the development of each type of SSWs?

For both types of SSW, there is relatively little variation between individual events in the orientation of the vortex breakdown relative to the underlying topography, i.e. SSWs of each type are to a large extent fixed in relation to the Earth's surface.

For displacement events the vortex is located in the 90°E-0°E sector prior to the warming, and by the end of the warming spans the entire Atlantic hemisphere from 90°E (lower vortex) to 90°W (upper vortex). There is somewhat more variability in the case of vortex-splitting events, but the dominant direction for the splits is found to be parallel to the 60°E-120°W great circle, such that the vortex splits into two daughter vortices approximately located over Canada and Siberia.

3. Does the three-dimensional structure of the Arctic polar vortex exhibit a characteristic life cycle during each type of SSW?

Each type of SSW exhibits a distinct life cycle, in the sense that the centroid position, orientation and elongation of the Arctic polar vortex evolves in a similar fashion during the onset and occurrence of individual splitting events, with a separate behaviour during displacement events.

For vortex-displacement events, the separation in longitude between the lower vortex (450 K) and the upper vortex (1425 K) increases uniformly during the SSW, from 85° longitude 10 days prior to the SSW onset day to 150° longitude just after the onset day. By the end of the displacement SSW life cycle, the upper part of the

vortex is destroyed by deformation, and the vortex subsequently reforms around its remnants in the lower stratosphere.

Splitting events are characterized by a rapid increase of the aspect ratio of the vortex a few days prior to the SSW, followed by the roll-up of the elongated vortex into two distinct 'daughter' vortices by the onset day. The daughter vortices then propagate rapidly apart to a distance of up to 5000 km, after which they experience a retrograde rotation around their common centroid, usually leading to the destruction of the weaker Canadian vortex and reformation of the main vortex around the stronger Siberian vortex.

8.2 Hierarchy of models

The barotropic nature of vortex-splitting SSWs, as indicated in our observational study, suggests that single layer models of the stratosphere are suitable candidates for the most idealized members of our hierarchy, at least from the perspective of vortex-splitting SSWs. A single layer quasi-geostrophic f-plane shallow water model is therefore chosen for one of the least complex members of our hierarchy. Due to the large Rossby radius of deformation which is typical for flows in the stratosphere, a purely barotropic (two-dimensional) f-plane model is also chosen, as it is a limiting case of the shallow water model when the Rossby radius of deformation becomes infinite. The most complex member of our hierarchy is a three-dimensional quasi-geostrophic f-plane model.

The overarching objective of the hierarchy of models approach is to investigate whether vortex-splitting and vortex-displacement SSW-like behaviours are exhibited

in these relatively idealized systems. One of the benefits of using such idealized models is that their modest complexity, when compared to General Circulation Models (GCMs) for example, makes it more straightforward to apply analytical techniques to study the dynamical processes responsible for this behaviour. Therefore, in addition to performing fully nonlinear numerical experiments of vortex behaviour in these systems, a strong emphasis is on developing an analytical theory explaining vortex breakdown in these models. In particular, in all hierarchy members the primary objective of the mathematical analysis is to seek SSW-like behaviour resulting from resonant excitation of waves on the edge of an idealized vortex. Within this remit, a particular interest is on the role of nonlinearity in 'self-tuning' initially off-resonant vortex states towards resonance.

8.2.1 Single layer models

The single layer members of our hierarchy of models investigate vortex-splitting behaviour with an idealized vortex patch. In all of these models, the polar vortex is represented by an initially circular vortex of uniform interior PV which is subject to an external forcing in a solid body rotating flow. The strength of the solid body rotation is represented by a rotation parameter Ω_b.

There are two choices of experimental set-up. The first set-up is an idealized vortex patch in a barotropic f-plane framework incorporating a straining velocity external forcing, known as the Kida vortex (Kida 1981). A particularly attractive feature of this model is that the vortex behaviour can be described using fully nonlinear analytical techniques. The second set-up is an idealized vortex patch in a shallow water

f-plane framework, in which external forcing is provided by topography (hereafter referred to as the 'shallow water model with topography', or SWT model). Due to its association with vortex-splitting SSWs, the topographic forcing in the SWT model is taken to have azimuthal wavenumber-2 structure when looking for vortex-splitting behaviour.

The objectives of the single layer model study, and our findings in respect to these, are summarized below.

1. Is vortex-splitting behaviour, similar to that observed during vortex-splitting SSWs, observed in the SWT model?

By varying the strength of the background rotation parameter Ω_b and the topographic forcing height $\mathcal{H}$, a variety of vortex behaviours are observed in fully nonlinear numerical simulations of the SWT model. Such behaviour includes anti-clockwise periodic rotation, clockwise periodic rotation and periodic oscillation of approximately elliptical vortex states. In addition to these periodic vortex states, for sufficiently large forcing heights vortex-splitting behaviour occurs in a narrow band of Ω_b values, with the width of this band increasing as the forcing height increases (see panel A of Fig. 8.1).

The qualitative features of these vortex-splitting experiments are very similar to those observed during vortex-splitting SSWs. That is, growth of wavenumber-2 disturbances on the vortex edge result in the vortex becoming almost elliptical in shape, and once the aspect ratio of the ellipse becomes sufficiently large, unstable disturbances act to 'pinch' the vortex around its centroid, eventually leading to a

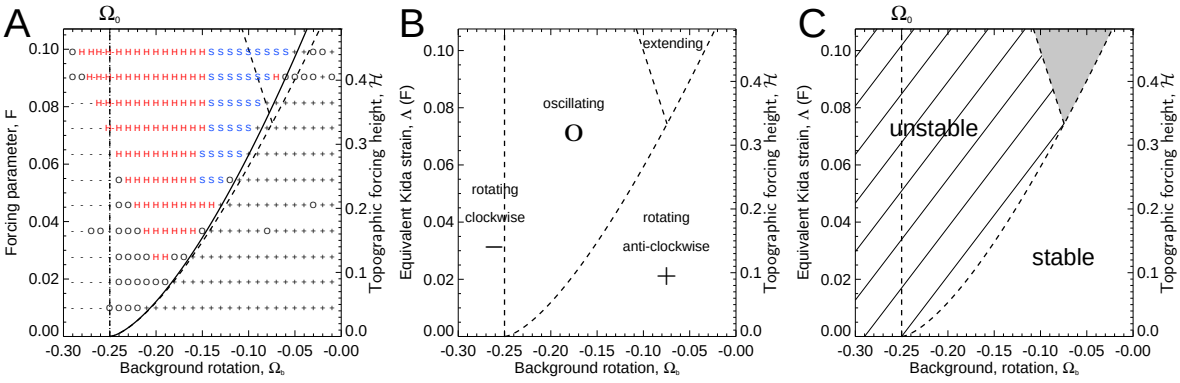

Figure 8.1: Panel A: Classification of vortex behaviour in the SWT fully nonlinear numerical model experiments as a function of rotation parameter Ω_b and nondimensional topographic forcing height $\mathcal{H}$ (see chapter 4). The Rossby radius of deformation in this case is taken to be infinite. Classification definitions are: $-$ clockwise rotation, $+$ anticlockwise rotation, **H** highly disturbed, **S** split, **O** oscillating/other. Panel B: Regimes of behaviour in the Kida model. Panel C: Stability regimes in the Kida model, with the shaded region showing the extending regime not considered by the stability analysis. For each $\mathcal{H}$ in panel A, the heavy solid line shows the Ω_b resulting in maximum excitation of the vortex, as predicted by weakly nonlinear analysis, and the dotted line shows the location of linear resonance at $\Omega_0 = -0.25$. In all panels, dashed lines give the location of regime boundaries in the Kida model.

complete split.

2. Under what circumstances, and to what extent, can the behaviour of the vortex in the SWT model be understood by reference to the Kida vortex model? Can a mapping between the two models be established?

If the Rossby radius of deformation in the SWT model is infinite, and if certain conditions are imposed on the topographic forcing structure, the SWT model and Kida model are shown to be identical. Even when these conditions are not satisfied, results from fully nonlinear numerical experiments show that there is still a close relationship between the SWT and Kida models. That is, as long as the topographic

forcing height is not too large, anti-clockwise rotation, clockwise rotation and oscillating regimes in the SWT model are accurately predicted by their counterpart regimes in the Kida model (see panel B of Fig. 8.1). When the topographic forcing height exceeds a certain threshold, the behaviour in the SWT model differs from that of the Kida model as a vortex-splitting regime emerges, lying wholly within the oscillating Kida regime.

Results from numerical experiments also indicate that stable and unstable regimes in the Kida model have analogous stability regimes in the SWT model. For example, highly disturbed and splitting behaviour in the SWT model only occur in the analogous unstable Kida regime (see panels A and C of Fig. 8.1).

Although the Kida model does not predict the existence of the vortex-splitting regime, it is seen that the regime boundary between unstable oscillating and stable anti-clockwise rotating states in the Kida model separates unstable vortex-splitting and stable anti-clockwise rotating states in the SWT model.

3. In the SWT model, does the excitation of waves on the vortex edge display features of a resonance with the background flow? If so, what is the role of nonlinearity in self-tuning the vortex towards resonance?

For small topographic forcing heights in the SWT model, numerical experiments show that the location in $(\Omega_b, \mathcal{H})$ parameter space of the most excited vortex states is exactly as predicted by a linear resonance of waves on the vortex edge. As the topographic forcing height increases, this location migrates from that predicted by a linear resonance. These findings are shown to hold for the SWT model with both

finite and infinite Rossby radii of deformation, and for several choices of topographic forcing structure.

The role of nonlinearity in this migration is approached using two analytical reductions. The first analytical reduction uses the SWT model with infinite Rossby radius of deformation in which the topographic forcing has wavenumber-2 azimuthal structure. Under these conditions, the Kida model acts as a good proxy when determining the behavioural characteristics of the system, as previously discussed. In the Kida model, the boundary between the oscillating and anti-clockwise regimes corresponds to vortex states which, although initially in motion, eventually become stationary with respect to the external straining field. From the perspective of wave-like disturbances in the SWT model, such solutions correspond to nonlinear corrections to a linear resonance, that is a self-tuning resonance. Therefore, the importance of this regime boundary in separating stable anti-clockwise vortex states from unstable vortex-splitting states in the SWT model strongly suggests that the underlying mechanism which causes such splits is that of a self-tuning resonance.

The second analytical reduction applies a weakly nonlinear asymptotic theory to the SWT model with arbitrary Rossby radius. For small topographic forcing heights, the weakly nonlinear theory gives quantitative predictions for self-tuning of wave-like disturbances towards, or away, from linear resonance. Comparing the weakly nonlinear predictions with results from fully nonlinear numerical experiments shows excellent agreement between the two, even when the topographic forcing height 'outgrows' the weakly nonlinear assumption of small forcing amplitude. Also, for experiments with azimuthal wavenumber-2 topographic forcing, the location of peak

disturbance amplitudes predicted by the weakly nonlinear theory separates stable anti-clockwise vortex states from unstable vortex-splitting states, indicating that the process of self-tuning resonance is responsible for the transition to splitting behaviour (see panel A of Fig. 8.1).

In summary, both analytical reductions show that as topographic forcing height increases, greatest disturbance amplitudes occur for vortex states which are initially offset from linear resonance, with the vortex self-tuning back towards resonance. Accurate predictions of the offset resulting in the greatest disturbance amplitude are given by each analytical reduction.

8.2.2 Three-dimensional models

The more complex members of our hierarchy of models use a stratified three-dimensional quasi-geostrophic f-plane framework. Forcing in the model is introduced in the form of an idealized lower boundary topography. When investigating vortex-splitting behaviour the forcing has wavenumber-2 azimuthal structure, whereas for vortex-displacement behaviour the forcing has wavenumber-1 azimuthal structure. At each height, the vortex is represented by an initially circular vortex patch of uniform interior PV. To investigate the impact of the vertical structure of the vortex on its resonant behaviour, the vortex area and interior PV are allowed to vary with height. To bring the vortex into or out of resonance with the forcing, a tuning parameter in the form of a uniform PV distribution Δ_b is added throughout the domain, which is analogous to the solid body rotation in the single layer models.

When summarizing below, it is natural to address the conclusions for vortex-splits

(S) and vortex-displacements (D) separately.

S-1. Is vortex-splitting behaviour, similar to that observed during vortex-splitting SSWs, reproduced in the three-dimensional model?

When topographic forcing with azimuthal wavenumber-2 structure is used in the idealized model, numerical experiments show that vortex-splitting behaviour is observed for a range of vortex structures. However, for each choice of vortex structure the topographic forcing height $\mathcal{H}$ must be sufficiently large for vortex splits to be observed at all. For an initially barotropic vortex structure, a thorough investigation of the (Δ_b,$\mathcal{H}$) parameter space shows that vortex-splitting only occurs in a narrow band of Δ_b values, with the width of this band increasing the forcing height increases (see Fig. 8.2). This behaviour agrees with that seen in the single layer model discussed previously.

Comparison of vortex-splitting experiments with the vortex-splitting SSW occurring in January 1958 shows that qualitative features of the split, and the time scale on which the split occurs, are similar between the two. The qualitative features of vortex-splitting in the three-dimensional model are also highly reminiscent of vortex-splitting behaviour in the single layer model.

S-2. Does the vertical structure of the vortex have a significant impact on the occurrence of vortex-splitting behaviour?

In the fully nonlinear numerical model experiments, vortex-splitting behaviour is observed for a wide range of realistic vortex structures. That is, the initial vortex structure in the idealized model has little to no effect on the potential for vortex-

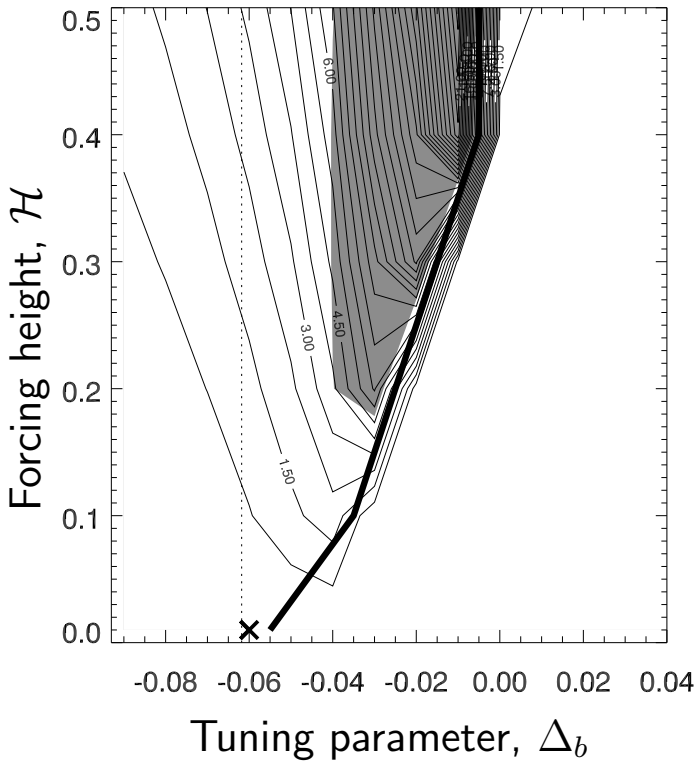

Figure 8.2: Contours of maximum disturbance amplitude as a function of tuning parameter Δ_b and topographic forcing height $\mathcal{H}$ for numerical experiments using an initially barotropic vortex structure. The shaded region denotes experiments in which the vortex is diagnosed as having split. The cross $\times$ shows the peak in maximum disturbance amplitude when $\mathcal{H} = 0.01$. The dashed line shows the Δ_b at which linear resonance occurs, and the heavy solid line traces the Δ_b location of the peak maximum disturbance amplitude at each forcing height.

splitting behaviour. However, some vortex structures are more resistant to splitting than others. In general, vortex structures which have anomalously weak PV, or which have area decreasing with height, require smaller forcing heights for splits to be observed, when compared to vortex structures which have anomalously strong PV or which have area increasing with height. In contrast, variation of the vortex PV with height seems to have little impact on the forcing heights required for vortex-splitting.

S-3. Can vortex-splitting be associated with a resonance of a particular vertical mode?

By investigating vortex excitation in response to small forcing heights in the idealized model, fully nonlinear numerical experiments indicate that greatest excitation occurs due to a linear resonance of the external, or barotropic, mode of the vortex. That is, vortex disturbances associated with resonant excitation of baroclinic vertical modes are far smaller than disturbances associated with a resonance of the barotropic mode.

Numerical experiments investigating the resonant response of an initially barotropic vortex show that nonlinearity results in the vortex self-tuning either towards, or away from, resonance. This self-tuning shares many similarities with the self-tuning exhibited by single layer members of our hierarchy of models. This suggests that the self-tuning mechanism responsible for vortex-splitting in our single layer models may also be responsible for vortex-splitting in the more complex models in our hierarchy.

D-1. Is vortex-displacement behaviour, similar to that observed during vortex-displacement SSWs, reproduced in an idealized model?

When topographic forcing with azimuthal wavenumber-1 structure is used in the idealized model, fully nonlinear numerical experiments show that sufficiently large forcing heights result in a vortex breakdown similar to that observed during vortex-displacement SSWs.

D-2. Can this vortex-displacement behaviour be explained in terms of an excitation of one of the baroclinic modes of the vortex, in a similar fashion to a linear resonance of the barotropic mode leading to vortex-splitting behaviour?

For wavenumber-1 topographic forcing, excitation of the barotropic mode of the vortex corresponds to an almost uniform displacement of the vortex from the pole at all heights. Therefore, by measuring the disturbance of the vortex relative to this displacement, it is possible to isolate the behaviour of disturbances to the vortex which are baroclinic in nature.

Calculations of the corresponding baroclinic wave-activity, that is the wave-activity measured relative to the displaced vortex centroid, show that the vortex breakdown discussed above is accompanied by a peak in baroclinic wave-activity associated with the smallest vertical wavenumbers of the baroclinic spectrum. This suggests that a resonant excitation of these baroclinic modes, rather than the barotropic mode, may be responsible for vortex breakdown similar to that observed during vortex-displacement SSWs.

8.3 Future work

The resonant excitation mechanism investigated here goes some way towards explaining why SSWs do not always occur in response to large amplitude forcing from the troposphere, and why their onset is so sudden. That is, for SSW-like behaviour to develop in our idealized models, the resonant mechanism requires not only large amplitude forcing but also favourable stratospheric conditions, and there exists a point at which an arbitrarily small change in these stratospheric conditions can result in a stable vortex state making an abrupt transition to an unstable vortex-splitting state.

While the conceptual models considered here are relatively idealized, it is proposed that progression of this work may consider a similar experimental framework in more complex primitive equation models. However, the nature of such models makes the derivation of analytical results far more complicated. Therefore, it is suggested that the first approach in such a study would be to compare patterns of resonant behaviour in more complex models with those of the three-dimensional models of our hierarchy, which in turn can be related back to the relatively well understood single layer models.

The ideal goal of this work is to find a quantifiable self-tuning mechanism leading to the onset of SSWs in the observed winter stratosphere, though a significant challenge remains in relating the controlling parameters of our idealized models to tangible quantities in the stratosphere. However, possible candidates may be the geopotential height anomaly on the 350 K isentropic surface as a proxy for topographic forcing height in our idealized models, or polar jet strength as a proxy for the background PV tuning parameter.

At present, efforts are continuing along the research paths outlined above with the

intention of developing the hierarchy of models approach towards more sophisticated, and state of the art, GCMs.

Appendix A

Weakly nonlinear dispersion relation: numerical verification

In this appendix, the weakly nonlinear dispersion relations given in equations (5.37) and (5.38) of chapter 5 are verified in the absence of rotating background flow, i.e. $\Omega_b = 0$.

Figures A.1 and A.2 show the nonlinear correction ω_2 to the linear wave frequency as a function of $\mathcal{B}$ for a fully nonlinear numerical model (diamonds) compared with that predicted by the weakly nonlinear theory (solid lines). The vortex in the numerical model is taken to have $\alpha_1 = 1$, hence ϵ acts as the sole parameter governing linear wave amplitude, and comprises of a weakly perturbed vortex of the form (5.18) with α_2 and α_3 given by (5.36). The ϵ^3 weakly nonlinear wave amplitude of the perturbation, α_3, was derived in a similar way to α_2 in (5.36), although the exact expression is not given here for brevity.

When calculating the linear wave frequency in the fully nonlinear numerical model,

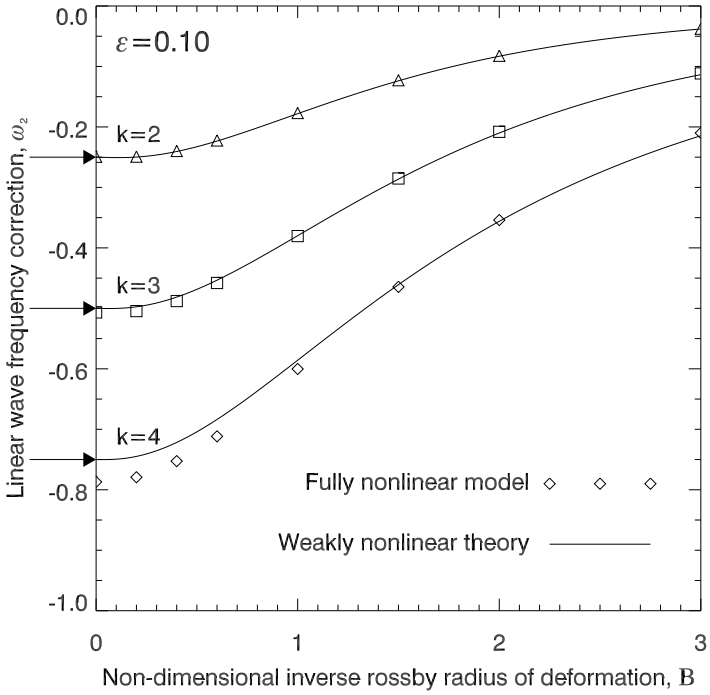

Figure A.1: The linear wave frequency correction ω_2 as a function of $\mathcal{B}$ when $\epsilon = 0.10$ for weakly nonlinear wave-like disturbances with azimuthal wavenumber $k = 2, 3, 4$.

a discrete Fourier transform was performed on the perturbed vortex edge as it evolved in time. From this transform, the frequency of the first harmonic was calculated and an average taken over one period of the linear wave's evolution T_0. To calculate the nonlinear correction ω_2 in the numerical model, the contribution of ω_0 given by (5.37) was then subtracted from the total linear wave frequency.

Figure A.1 shows excellent agreement for ω_2 between the numerical model and the weakly nonlinear theory, with a small exception in the case of $k = 4$ and $\mathcal{B} < 1$ for which there is slight disagreement between the two.

When the linear disturbance amplitude is marginally reduced to $\epsilon = 0.05$, as shown in figure A.2, it is seen that this disagreement is dramatically reduced such that the

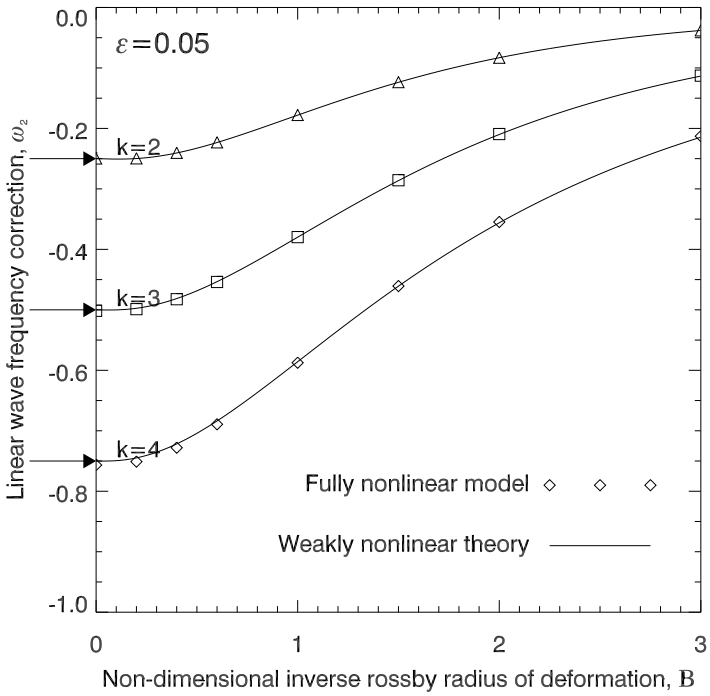

Figure A.2: The linear wave frequency correction ω_2 as a function of $\mathcal{B}$ when $\epsilon = 0.05$ for weakly nonlinear wave-like disturbances with azimuthal wavenumber $k = 2, 3, 4$.

numerical model and weakly nonlinear theory agree for all shown values of k and $\mathcal{B}$. As ϵ increases, instabilities imply that the linear frequency correction is unlikely to remain as predicted over one period of time T_0. This is most readily apparent by observing the value of ω_2 for a single time increment at $t = 0$. At his time, before instabilities can evolve, the ω_2 observed does indeed agree with that predicted by the weakly nonlinear theory.

In both A.1 and A.2 the arrows point to the frequency correction ω_2 derived in Su (1979), corresponding to taking $\mathcal{B} = 0$ in the shallow water calculations for ω_0 and ω_2. As expected, it is seen that the weakly nonlinear shallow water frequencies tend towards their purely two-dimensional counterparts as $\mathcal{B} \to 0$ (as given in Su 1979).

Figure A.3 shows the total linear wave frequency $\omega = \omega_0 + \epsilon^2\omega_2$ as a function of linear wave amplitude ϵ when $\mathcal{B} = 3.0$ for the fully nonlinear numerical model (symbols) and weakly nonlinear model (solid lines). Once again, there is excellent agreement between the two up until $\epsilon \approx 0.3$. At this point, the condition on the validity of the velocity fields $\epsilon\mathcal{B} \ll 1$ is violated, and the weakly nonlinear theory ceases to give a good approximation. Even so, it is encouraging that there is such good agreement even for these large values of $\epsilon\mathcal{B}$, and perhaps even more encouraging that the agreement in the $k = 2$ case, which is of primary importance when investigating vortex splitting events, remains very good for even higher values of $\epsilon\mathcal{B}$.

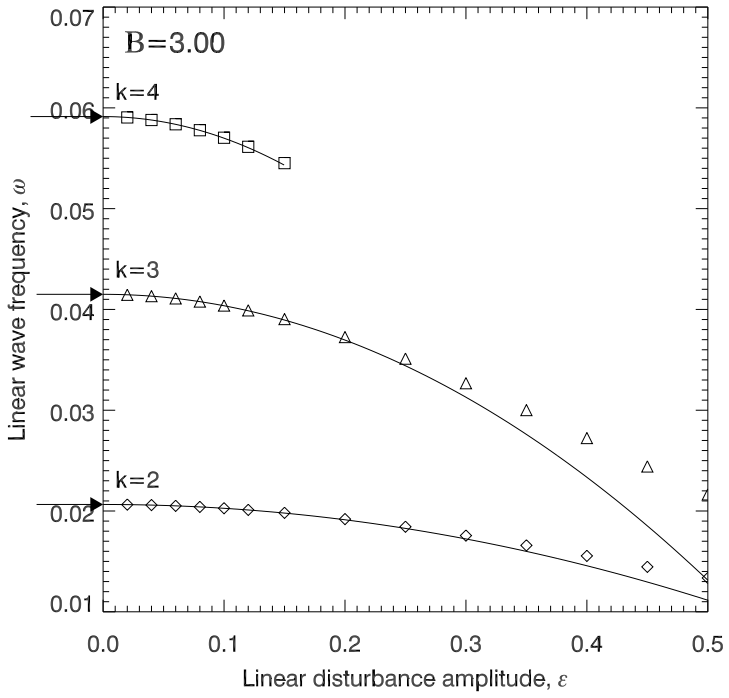

Figure A.3: The linear wave frequency ω as a function of linear disturbance amplitude ϵ when $\mathcal{B} = 3.0$ for weakly nonlinear wave-like disturbances with azimuthal wavenumber $k = 2, 3, 4$. Symbols correspond to numerical model runs, and solid lines to the weakly nonlinear theory.

Bibliography

Andrews, D. G., J. R. Holton, and B. Leovy, 1985: *Middle Atmosphere Dynamics*. Academic Press.

Andrews, D. G. and M. E. McIntyre, 1976: Planetary waves in horizontal and vertical shear: the generalized Eliassen-Palm relation and the mean zonal acceleration. *J. Atmos. Sci.*, **33**, 2031–2048.

— 1978: Generalized Eliassen-Palm and Charney-Drazin theorems for waves on axisymmetric mean flows in compressible atmospheres. *J. Atmos. Sci.*, **35**, 175–185.

Baldwin, M. P. and T. J. Dunkerton, 1999: Propagation of the arctic oscillation from the stratosphere to the troposphere. *J. Geophys. Res.*, **104**, 937–946.

— 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294**, 581–584.

Barnett, J. J., 1975: Large sudden warming in the Southern Hemisphere. *Nature*, **255**, 387–389.

Black, R. X. and B. A. McDaniel, 2007: The dynamics of Northern Hemisphere stratospheric final warming events. *J. Atmos. Sci.*, **64**, 2932–2946.

Butchart, N., S. A. Clough, T. N. Palmer, and P. J. Trevelyan, 1982: Simulations of an observed stratospheric warming with quasigeostrophic refractive index as a model diagnostic. *Q. J. Roy. Met. Soc.*, **108**, 475–502.

Butchart, N. and E. E. Remsberg, 1986: The area of the stratospheric polar vortex as a diagnostic for tracer transport on an isentropic surface. *J. Atmos. Sci.*, **43**, 1319–1339.

Charlton, A. J. and L. M. Polvani, 2007: A new look at stratospheric sudden warmings I: climatology and modeling benchmarks. *J. Clim.*, **20**, 449–469.

Charlton, A. J., L. M. Polvani, J. Perlwitz, F. Sassi, E. Manzini, K. Shibata, S. Pawson, J. E. Nielsen, and D. Rind, 2007: A new look at stratospheric sudden warmings II: Evaulation of Numerical Model Simulations. *J. Clim.*, **20**, 470–488.

Charney, J. G. and J. G. DeVore, 1979: Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.*, **36**, 1205–1216.

Charney, J. G. and P. G. Drazin, 1961: Propagation of planetary scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–109.

Charney, J. G. and M. E. Stern, 1962: On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19**, 159–172.

Clark, J. H. E., 1974: Atmospheric response to the quasi-resonant growth of forced planetary waves. *J. Met. Soc. Jap.*, **52**, 143–162.

Clough, S. A., N. S. Grahame, and A. O'Neill, 1985: Potential vorticity in the stratosphere derived using data from satellites. *Q. J. Roy. Met. Soc.*, **111**, 335–358.

Davis, P. J., 1974: *The Schwarz function and its applications*. The Mathematical Association of America.

Dritschel, D. G., 1986: The nonlinear evolution of rotating configurations of uniform vorticity. *J. Fluid Mech.*, **172**, 157–182.

— 1988: Contour surgery: a topological reconnection scheme for extended integrations using contour dynamics. *J. Comput. Phys.*, **77**, 240–266.

— 1990: The stability of elliptical vortices in an external straining flow. *J. Fluid Mech.*, **210**, 223–261.

— 1993: A fast contour dynamics method for many-vortex calculations in two-dimensional flows. *Phys. Fluids A*, **25**, 173–186.

— 1997: Introduction to "contour dynamics for the euler equations in two dimensions". *J. Comput. Phys.*, **135**, 217–219.

Dritschel, D. G. and M. H. P. Ambaum, 1997: A contour-advective semi-lagrangian algorithm for the simulation of fine-scale conservative fields. *Q. J. Roy. Met. Soc.*, **123**, 1097–1130.

Dritschel, D. G., L. M. Polvani, and A. R. Mohebalhojeh, 1999: The contour-advective semi-lagrangian algorithm for the shallow water equations. *Monthly Weather Review*, **127**, 1551–1565.

Dritschel, D. G. and R. Saravanan, 1994: Three-dimensional quasi-geostrophic contour dynamics, with an application to stratospheric vortex dynamics. *Q. J. Roy. Met. Soc.*, **120**, 1267–1297.

Dunkerton, T. J. and D. P. Delisi, 1986: Evolution of potential vorticity in the winter stratosphere of January-February 1979. *J. Geophys. Res.*, **91**, 1199–1208.

Edmon, H. J., B. J. Hoskins, and M. E. McIntyre, 1980: Eliassen-Palm cross sections for the troposphere. *J. Atmos. Sci.*, **37**, 2600–2616.

Eliassen, A. and E. Palm, 1961: On the transfer of energy in stationary mountain waves. *Geofys. Publikasjoner*, **12**, 1–23.

Esler, J. G., L. M. Polvani, and R. K. Scott, 2006: The antarctic sudden stratospheric warming of 2002: A self-tuned resonance? *Geophysical Research Letters*, **33**, doi:10.1029/2006GL026034.

Esler, J. G. and R. K. Scott, 2005: Excitation of transient Rossby waves on the stratospheric polar vortex and the barotropic sudden warming. *J. Atmos. Sci.*, **62**, 3661–3682.

Fahey, D. W., S. R. Kawa, and K. R. Chan, 1990: Nitric oxide measurements in the arctic winter stratosphere. *Geophysical Research Letters*, **17**, 489–492.

Fleagle, R. G., 1958: Inferences concerning the dynamics of the mesosphere. *J. Geophys. Res.*, **63**, 137–145.

Gill, A. E., 1982: *Atmosphere-Ocean Dynamics*. Academic Press.

Gray, L. J., S. Crooks, C. Pascoe, S. Sparrow, and M. Palmer, 2004: Solar and QBO influences on the timing of stratospheric sudden warmings. *J. Atmos. Sci.*, **61**, 2777–2796.

Gray, L. J., S. Sparrow, M. Juckes, A. O'Neill, and D. G. Andrews, 2003: Flow regimes in the winter stratosphere of the northern hemisphere. *Q. J. Roy. Met. Soc.*, **129**, 925–945.

Guo, Y., C. Hallstrom, and D. Spirn, 2004: The dynamics near an unstable Kirchhoff ellipse. *Comm. Math. Phys.*, **245**, 297–354.

Hartmann, D. L., L. E. Heidt, M. Loewenstein, J. R. Podloske, J. Vedder, W. L. Starr, and S. E. Strahan, 1989: Transport into the south polar vortex in early spring. *J. Geophys. Res.*, **94**, 16779–16795.

Holton, J. R., 1976: A semi-spectral numerical model for wave-mean flow interactions in the stratosphere: application to sudden stratospheric warmings. *J. Atmos. Sci.*, **33**, 1639–1649.

Holton, J. R. and H.-C. Tan, 1980: The influence of the equatorial Quasi-Biennial Oscillation on the global circulation at 50 mb. *J. Atmos. Sci.*, **37**, 2200–2208.

Jung, J. H., C. S. Konor, C. R. Mechoso, and A. Arakawa, 2001: A study of the stratospheric major warming and subsequent flow recovery during the winter of 1979 with an isentropic vertical coordinate model. *J. Atmos. Sci.*, **58**, 2630–2649.

Kida, S., 1981: Motion of an elliptic vortex in a uniform shear flow. *J. Phys. Soc. Jap.*, **50**, 3517–3520.

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and M. Fior-

ino, 2001: The NCEP-NCAR 50-Year reanalysis: monthly means CD-ROM and documentation. *Bull. Am. Met. Soc.*, **82**, 247–267.

Labitzke, K., 1977: Interannual variability of the winter stratosphere in the Northern Hemisphere. *Monthly Weather Review*, **105**, 762–770.

— 1981: The amplification of height wave 1 in January 1979: a characteristic precondition for the major warming in February. *Monthly Weather Review*, **109**, 983–989.

Labitzke, K. and B. Naujokat, 2000: The lower arctic stratosphere in winter since 1952. *SPARC Newsletter*, **15**.

Lahoz, W. A., A. O'Neill, A. Heaps, V. D. Pope, R. Swinbank, R. S. Harwood, L. Froidevaux, W. G. Read, J. W. Waters, and G. E. Peckham, 1996: Vortex dynamics and the evolution of water vapour in the stratosphere of the Southern Hemisphere. *Q. J. Roy. Met. Soc.*, **122**, 423–450.

Lait, L. R., 1994: An alternative for potential vorticity. *J. Atmos. Sci.*, **51**, 1754–1759.

Legras, B. and D. G. Dritschel, 1993: A comparison of the contour surgery and pseudo-spectral methods. *J. Comput. Phys.*, **104**, 287–302.

Liberato, M. L. R., J. M. Castanheira, L. de la Torre, C. C. DaCamara, and L. Gimeno, 2007: Wave energy associated with the variability of the stratospheric polar vortex. *J. Atmos. Sci.*, **64**, 2683–2694.

Limpasuvan, V., D. W. J. Thompson, and D. L. Hartmann, 2004: The life cycle of the Northern Hemisphere sudden stratospheric warmings. *J. Clim.*, **17**, 2584–2596.

Lindzen, R. S., 1966: Radiative and photochemical processes in mesospheric dynamics. Part IV: stability of a zonal vortex at middle latitudes to baroclinc waves. *J. Atmos. Sci.*, **23**, 350–359.

Loewenstein, M., J. R. Podolske, and K. R. Chan, 1990: N_2O as a dynamical tracer in the arctic vortex. *Geophysical Research Letters*, **17**, 477–480.

Loewenstein, M., J. R. Podolske, K. R. Chan, and S. E. Strahan, 1989: Nitrous oxide as a dynamical tracer in the 1987 Airborne Antarctic Ozone Experiment. *J. Geophys. Res.*, **94**, 11589–11598.

Love, A. E. H., 1893: On the stability of certain vortex motions. *Proc. Lond. Math. Soc.*, **25**, 18–43.

Macaskill, C., W. E. P. Padden, and D. G. Dritschel, 2003: The CASL algorithm for quasi-geostrophic flow in a cylinder. *J. Comput. Phys.*, **188**, 232–251.

Manney, G. L., D. R. Allen, K. K. uger, B. Naujokat, M. L. Santee, J. L. Sabutis, S. Pawson, R. Swinbank, C. E. Randall, A. J. Simmons, and C. Long, 2005a: Diagnostic comparison of meteorological analyses during the 2002 antarctic winter. *Monthly Weather Review*, **133**, 1261–1278.

Manney, G. L., J. D. Farrara, and C. R. Mechoso, 1994: Simulations of the February 1979 stratospheric sudden warming: model comparisons and three-dimensional evolution. *Monthly Weather Review*, **122**, 1115–1140.

Manney, G. L., K. Krüger, J. L. Sabutis, S. A. Sena, and S. Pawson, 2005b: The re-

markable 2003-2004 winter and other recent warm winters in the arctic stratosphere since the late 1990s. *J. Geophys. Res.*, **110**, doi:10.1029/2004JD005367.

Manney, G. L., W. A. Lahoz, R. Swinbank, A. O'Neill, P. M. Connew, and R. W. Zurek, 1999: Simulation of the December 1998 stratospheric major warming. *J. Geophys. Res.*, **26**, 2733–2736.

Manney, G. L., J. L. Sabutis, D. R. Allen, W. A. Lahoz, A. Scaife, C. E. Randall, S. Pawson, B. Naujokat, and R. Swinbank, 2005c: Simulations of dynamics and transport during the September 2002 antarctic major warming. *J. Atmos. Sci.*, **62**, 690–707.

Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, **28**, 1479–1494.

Matthewman, N. J., J. G. Esler, A. J. Charlton-Perez, and L. M. Polvani, 2009: A New Look at Stratospheric Sudden Warmings. Part III. Polar Vortex Evolution and Vertical Structure. *J. Clim.*, **22**, 1566–1585.

McIntyre, M. E., 1972: Baroclinic instability of an idealized model of the polar night jet. *Q. J. Roy. Met. Soc.*, **98**, 165–174.

— 1982: How well do we understand the dynamics of stratospheric warmings? *J. Met. Soc. Jap.*, **60**, 37–65.

McIntyre, M. E. and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature*, **305**, 593–600.

— 1984: The 'surf zone' in the stratosphere. *J. Atmos. Terr. Phys.*, **46**, 825–849.

Melander, M. V., N. J. Zabusky, and A. S. Styczek, 1986: A moment model for vortex interactions of the two-dimensional Euler equations. Part 1. Computational validation of a Hamiltonian elliptical representation. *J. Fluid Mech.*, **167**, 95–115.

Müller, R. and G. Günther, 2003: A generalized form of Lait's modified potential vorticity. *J. Atmos. Sci.*, **60**, 2229–2237.

Murray, F. W., 1960: Dynamic stability in the stratosphere. *J. Geophys. Res.*, **65**, 3273–3305.

Nash, E. R., P. A. Newman, J. E. Rosenfield, and M. R. Schoeberl, 1996: An objective determination of the polar vortex using Ertel's potential vorticity. *J. Geophys. Res.*, **101**, 9471–9478.

Naujokat, B., K. Krüger, K. Matthes, J. Hoffmann, M. Kunze, and K. Labitzke, 2002: The early major warming in December 2001 - exceptional? *Geophysical Research Letters*, **29**.

Nishii, K. and H. Nakamura, 2004: Tropospheric influence on the diminished Antarctic ozone hole in September 2002. *Geophysical Research Letters*, **31**.

O'Neill, A., 2003: Stratospheric sudden warmings. *Encyclopedia of Atmospheric Sciences*, J. R. Holton, J. A. Pyle, and J. A. Curry, eds., Elsevier, 1342–1353.

O'Neill, A., W. L. Grose, V. D. Pope, H. Maclean, and R. Swinbank, 1994: Evolution of the stratosphere during northern winter 1991/92 as diagnosed from U.K. Meteorological Office analyses. *J. Atmos. Sci.*, **51**, 2800–2817.

O'Neill, A. and V. D. Pope, 1988: Simulations of linear and non-linear disturbances in the polar stratosphere. *Q. J. Roy. Met. Soc.*, **114**, 1063–1110.

Palmer, T. N., 1981: Diagnostic study of a wavenumber-2 stratospheric sudden warming in a transformed Eulerian-mean formalism. *J. Atmos. Sci.*, **38**, 844–855.

Peters, D., P. Vargin, and H. Körnich, 2007: A study of the zonally asymmetric tropospheric forcing of the austral vortex splitting during September 2002. *Tellus*, **59A**, 384–394.

Plumb, R. A., 1981: Instability of the distorted polar night vortex: a theory of stratospheric warmings. *J. Atmos. Sci.*, **38**, 2514–2531.

Plumb, R. A., D. W. Waugh, R. J. Atkinson, P. A. Newman, L. R. Lait, M. R. Schoeberl, E. V. Browell, A. J. Simmons, and M. Loewenstein, 1994: Intrusions into the lower stratospheric arctic vortex during the winter of 1991-1992. *J. Geophys. Res.*, **99**, 1089–1105.

Polvani, L. M. and R. Saravanan, 2000: The three-dimensional structure of breaking Rossby waves in the polar wintertime stratosphere. *J. Atmos. Sci.*, **57**, 3663–3685.

Proffitt, M. H., D. W. Fahey, K. K. Kelly, and A. F. Tuck, 1989: A chemical definition of the boundary of the antarctic ozone hole. *J. Geophys. Res.*, **94**, 437–448.

Quiroz, R. S., 1974: Stratospheric warmings in the Southern Hemisphere deduced from satellite radiation data, 1969-1973. *Proceedings of the International Conference on Structure, Composition and General Circulation of the Upper and Lower*

Atmospheres and Possible Anthropogenic Perturbations, IAMAP/IAPSO First Special Assembly, 525–539.

Saffman, P. G., 1992: *Vortex Dynamics*. Cambridge University Press.

Schoeberl, M. R. and D. F. Strobel, 1978: The zonal averaged circulation of the middle atmosphere. *J. Atmos. Sci.*, **35**, 577–591.

Scott, R. K. and D. G. Dritschel, 2005: Quasi-geostrophic vortices in compressible atmospheres. *J. Fluid Mech.*, **530**, 305–325.

— 2006: Vortex-vortex interactions in the winter stratosphere. *J. Atmos. Sci.*, **63**, 726–740.

Scott, R. K., D. G. Dritschel, L. M. Polvani, and D. W. Waugh, 2004: Enhancement of Rossby wave breaking by steep potential vorticity gradients in the winter stratosphere. *J. Atmos. Sci.*, **61**, 904–918.

Scott, R. K. and L. M. Polvani, 2004: Stratospheric control of upward wave flux near the tropopause. *Geophysical Research Letters*, **31**.

Simmons, A., M. Hortal, G. Kelly, A. McNally, A. Untch, and S. Uppala, 2005: ECMWF analyses and forecasts of stratospheric winter polar vortex breakup: September 2002 in the Southern Hemisphere and related events. *J. Atmos. Sci.*, **62**, 668–689.

Smith, A. K., 1989: An investigation of resonant waves in a numerical model of an observed sudden stratospheric warming. *J. Atmos. Sci.*, **46**, 3038–3054.

Su, C. H., 1979: Motion of fluid with constant vorticity in a singly-connected region. *Phys. Fluids*, **22**.

Swanson, K. L., 2000: Stationary wave accumulation and the generation of low-frequency variability on zonally varying flows. *J. Atmos. Sci.*, **57**, 2262–2280.

Swanson, K. L., P. J. Kushner, and I. M. Held, 1997: Dynamics of barotropic storm tracks. *J. Atmos. Sci.*, **54**, 791–810.

Tang, Y., 1987: Nonlinear stability of vortex patches. *Trans. Am. Math. Soc.*, **304**, 617–637.

Tung, K. K. and R. S. Lindzen, 1979a: A theory of stationary long waves. Part I: a simple theory of blocking. *Monthly Weather Review*, **107**, 714–734.

— 1979b: A theory of stationary long waves. Part II: resonant Rossby waves in the presence of realistic vertical shears. *Monthly Weather Review*, **107**, 735–750.

Uppala, S. M., P. W. Kallberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. V. D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Holm, B. J. Hoskins, L. Isaksen, P. A. E. Janssen, R. Jenne, A. P. McNally, J.-F. Mahfouf, J.-J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen, 2005: The ERA-40 re-analysis. *Q. J. Roy. Met. Soc.*, **131**, 2961–3012.

Vallis, G. K., 2006: *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press.

Wang, X. and J. Fyfe, 2000: Onset of edge wave breaking in an idealized model of the polar stratospheric vortex. *J. Atmos. Sci.*, **57**, 956–966.

Watson, G. N., 1944: *A Treatise on the theory of Bessel functions*. Cambridge University Press.

Waugh, D. W., 1997: Elliptical diagnostics of stratospheric polar vortices. *Q. J. Roy. Met. Soc.*, **123**, 1725–1748.

Waugh, D. W. and D. G. Dritschel, 1999: The dependence of Rossby wave breaking on the vertical structure of the polar vortex. *J. Atmos. Sci.*, **56**, 2359–2375.

Waugh, D. W., R. A. Plumb, R. J. Atkinson, M. R. Schoeberl, L. R. Lait, P. A. Newman, M. Loewenstein, D. W. Toohey, L. M. Availone, C. R. Webster, and R. D. May, 1994: Transport out of the lower stratospheric Arctic vortex by Rossby wave breaking. *J. Geophys. Res.*, **99**, 1071–1088.

Waugh, D. W. and W. J. Randel, 1999: Climatology of arctic and antarctic polar vortices using elliptic diagnostics. *J. Atmos. Sci.*, **56**, 1594–1613.