

AN ABSTRACT OF THE DISSERTATION OF

Ivana Cerovečki for the degree of Doctor of Philosophy in Oceanography presented on October 15, 2002. Title: Long Baroclinic Planetary Waves in a Nonzonal Vertically Sheared Mean Flow.

Redacted for privacy

Abstract approved: _____

Roland A. de Szoeke

We consider long planetary waves of annual period linearized about a steady, wind-driven subtropical (ST) gyre circulation. The circulation divides the gyre into an eastern zone (EZ) where only the surface layer is in steady motion and a western zone (WZ) where both the surface and the middle layer are in steady motion because the potential vorticity of the middle layer is assumed to be homogenized. In both zones the upper two layers participate in the wave motion while the deepest third layer is supposed infinitely deep and therefore quiescent.

In the EZ the solution consists of two sets of baroclinic waves, each a combination of the classical first and the second baroclinic Rossby wave coupled by the background flow. In the WZ the propagation of information from the western boundary of the ocean is neglected; therefore the WZ solution consists of only the long set of baroclinic waves.

For sufficiently strong background flow the annual period perturbation solution in the southwestern part of the ST gyre is baroclinically unstable and grows westward. The instability may be generated by local wind forcing or energy incident from the east.

We study the first possibility by analysing perturbations that are generated by a specified initial perturbation or by annual period forcing started at some initial time. Even very gentle nonzonal background flow dramatically changes the waves away from the simple β solutions. In the first case a Gaussian shaped initial disturbance evolves in a highly anisotropic manner. Growth of the amplitude of the transient solution depends strongly on the central wavenumber and spectral width of the initial disturbance. In the latter case the solution consists of a forced part varying at the annual period plus a transient part.

Long Baroclinic Planetary Waves in a Nonzonal Vertically Sheared Mean Flow

by

Ivana Cerovečki

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed October 15, 2002

Commencement June 2003

Doctor of Philosophy dissertation of Ivana Cerovečki presented on October 15, 2002

APPROVED:

Redacted for privacy

Major Professor, representing Oceanography

Redacted for privacy

Dean of the College of Oceanic and Atmospheric Sciences

Redacted for privacy

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Redacted for privacy

Ivana Cerovečki, Author

ACKNOWLEDGMENTS

I would like to thank my advisor, Roland de Szoeko, for trusting me with an interesting and rich problem that both interested me and helped further my education by introducing me to new and fascinating areas of geophysical fluid dynamics and applied mathematics. He helped me learn how to present my work and helped me to get to know researchers outside OSU. As I move into new research areas, I look forward to continuing my scientific relationship with him.

I asked Roger Samelson to be on my committee before I had met him. I could not have made a better choice; sometimes the best things happen just by luck. I would like to express many thanks to Ronald Guenther for his steady interest in my work and his kindest support. I am very glad that I asked John Allen to be on my committee. His insisting on the highest standards has been the most useful preparation for the "real world".

Many thanks to Myrl Hendershott who showed continual interest in my work. His enthusiasm for science and his kind nature make him a truly outstanding person. Rick Salmon has always been available to talk about science with motivating enthusiasm, positive feedback and patience to read my manuscript in its very early stage. I would like to thank Clint Winant who has provided me encouragement when I needed it most and who has always been available to help.

Joe Pedlosky most generously found time to talk to me about my research on many occasions, answering patiently (too) numerous questions in the most motivating manner, for which I am truly grateful.

Many thanks to Irma Delson who has always been available for help and support. I am truly grateful to Scott Springer who has been the best friend when I needed him most, and who has shown endless patience with my compiling problems.

Julie Pullen has been the nicest friend one can wish to have for all these years.
Ricardo Matano was a kind friend in difficult moments.

Finally, I would like to thank my B.Sc. and M. Sc. thesis adviser Mirko Orlic who had introduced me to the world of physical oceanography and encouraged me to do my Ph.D. in the USA, for both of which I am truly grateful.

This work was funded by the National Science Foundation through Grants OCE-9907008 and 0220471.

TABLE OF CONTENTS

	<u>Page</u>
1 INTRODUCTION	1
2 LONG PLANETARY WAVES IN A TWO-AND-A-HALF LAYER MODEL OCEAN WITH A WIND-DRIVEN STEADY CIRCULATION	5
2.1 Introduction	8
2.2 The Model	11
2.2.1 Steady Model	13
2.2.2 East zone steady solution	14
2.2.3 West zone steady solution	16
2.2.4 An illustration of the steady solution	18
2.3 The perturbation equations	18
2.3.1 Perturbation equations in the EZ	20
2.3.2 Perturbation equations in the WZ	20
2.3.3 Sea surface displacement	21
2.4 Characteristics of the governing equations for $H_1^{(1)}$ and $H_2^{(1)}$:	21
2.4.1 In the EZ	22
2.4.2 In the WZ	26
2.5 Numerical integration	27
2.5.1 In the EZ	27
2.5.2 In the WZ	28
2.6 A preliminary numerical solution	29
2.7 WKB solution	32
2.7.1 In the EZ	32
2.7.2 In the WZ	35

TABLE OF CONTENTS (Continued)

	<u>Page</u>
2.7.3 An illustration of a WKB solution	35
2.8 Stability analysis of the numerical scheme used in the EZ	36
2.8.1 Numerical dispersion relation	41
2.8.2 Scale selective damping: 3 point averaging	46
2.8.3 Application of the stability analysis of the numerical scheme used in the EZ	47
2.8.3.1 Resolution of long waves in the numerical solution	50
2.8.3.2 Numerical instability with no analytical counterpart	52
2.8.3.3 Numerical instability due to resolution induced meridional wrap-around in the wavenumber plane	52
2.8.3.4 Effects of averaging	54
2.8.3.5 Numerical instability induced by northward propagating short waves being aliased into southward propagating un- stable waves	56
2.9 Discussion	60
2.9.1 The possibility of true baroclinic instability in the southwest- ern part of the subtropical EZ	60
2.9.1.1 Satisfaction of the Charney-Stern criterion	60
2.9.1.2 Meridional flow effects	61
2.9.1.3 Local forcing	65
2.9.1.4 Refraction	65
2.9.1.5 Relative vorticity	68
2.9.2 Numerical solution for the parameters representative of North Pacific subtropical gyre	69
2.9.3 Comparison with the observations and with other solutions	70
2.10 Summary	77
2.11 Appendix	82
2.11.1 Perturbation equations in the EZ	82
2.11.2 Perturbation equations in the WZ	83

TABLE OF CONTENTS (Continued)

	<u>Page</u>
3 THE INITIAL VALUE PROBLEM FOR LONG PLANETARY WAVES IN A TWO-AND-A-HALF LAYER MODEL IN THE PRESENCE OF NON- ZONAL MEAN FLOW	87
3.1 Introduction	88
3.2 The model	90
3.3 The perturbation equations	93
3.4 Plane wave solutions	94
3.5 The initial value problem: a transient solution	97
3.5.1 The transient solution obtained by numerically summing Fourier components	98
3.5.1.1 Gaussian disturbance centered at $k = 0, l = 0$; no background flow	98
3.5.1.2 Gaussian disturbance centered at $k = 0, l = 0$; spatially homogeneous background flow towards SW	100
3.5.1.3 Gaussian disturbance centered at k_0, l_0 ; spatially homogeneous background flow towards SW	105
3.6 Transient solution obtained by the method of steepest descents with modifications	107
3.6.1 Introduction	107
3.6.2 Transient solution for the Gaussian shaped initial disturbance	110
3.6.3 Front	116
3.6.4 Caustics	118
3.6.5 Double Caustic	119
3.6.6 Approximate solutions for stationary phase points far from the center of the disturbance	120

TABLE OF CONTENTS (Continued)

	<u>Page</u>
3.7 Solutions for the stationary phase points uniformly valid across the caustics	122
3.7.1 Solutions for stationary phase points uniformly valid across the caustic approximated by $\beta = \beta_0 = \arctan(-\frac{b}{v})$	122
3.7.2 Solution for stationary phase points uniformly valid across the caustic approximated by $\beta = \frac{\pi}{2}$	124
3.7.3 Discussion of the solutions of the quartic equation	127
3.8 Nature of the instability	130
3.9 Transient solution for a localised wavemaker switched on at the initial time $t = 0$	130
3.10 The sympathetic solution obtained by numerically summing Fourier components	135
3.11 Conclusions	138
4 SUMMARY	144
BIBLIOGRAPHY	149

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1 Time-longitude sections of annually filtered sea level for the midlatitude North Pacific (from Chelton and Schlax, 1996, Figure 2).	9
2.2 Three-dimensional perspective view of the layer base depth displacements due to the steady wind forcing looking into the basin from the southwest corner of the subtropical gyre.	12
2.3 Characteristics of the equations governing annual perturbations about the mean flow (left panel) and the spatial structure of both the steady and the time-dependent Ekman pumping (right panel). Dash-dot line denoted by x_B is the boundary between the EZ and the WZ. The star denotes the Rossby Attractor point (RA).	17
2.4 Detail of the two sets of the characteristics of the EZ solution. Dot-dash line shows the boundary between the EZ and the WZ.	25
2.5 A snapshot of the time-dependent interface displacement of the base of the middle layer (m) obtained with very weak steady wind forcing of the background flow; a case which emphasizes the waves.	30
2.6 A snapshot of the annual period time-dependent middle layer interface displacement (m) for the long wave part of the WKB solution for the same set of parameters as in Figure 2.5. Contour levels are 0.25 m ranging from -1 m to 1 m.	37
2.7 Periodically repeating pattern of identical cells on which von Neumann type of analysis was performed	39

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>	
2.8	Numerical dispersion relation (2.74) and the corresponding local analytical WKB dispersion relation (2.49) in χ, y coordinates (units: radians/km) for the parameters of Figure 2.5. We fix the frequency σ , vary the meridional wavenumber l (real) and solve the quadratic dispersion relation for the zonal wavenumbers k (complex in general). The real part of k is plotted as a light solid line for the analytical dispersion relation and as a dashed line for the numerical dispersion relation. The imaginary part of k , representing the spatial growth, is plotted as a dash-dot line for the analytical dispersion relation and as a dark solid line for the numerical dispersion relation. The spurious numerical instability occurs in the vicinity of the point in the wavenumber space where the aliased short wave branch crosses the long wave branch. Both plots are for the numerical solution of Figure 2.11, the upper panel is for a point $x = 6700$ km, $y = 3050$ km and the lower panel is for $x = 2000$ km, $y = 2000$ km.	44
2.9	Analytical and averaged numerical dispersion relation for the two points shown in Figure 2.8. The latter has been obtained by using the three point average with the central weight 0.4	48
2.10	Local analytical and numerical dispersion relation for a point A shown in Figure 2.12. The upper panel shows the analytical and numerical dispersion relation for the unaveraged solution shown in Figure 2.11 and the lower panel shows the analytical and numerical dispersion relations for the averaged numerical solution shown in Figure 2.12. The star indicates the wavelength estimate of the long wave part of the solution at $x = 9000$ km, $y = 3500$ km: $\lambda_z = 920$ km, $\lambda_m = 2180$ km.	51
2.11	A snapshot of the time-dependent interface displacement of the base of the middle layer (m) for the same set of parameters as in Figure 2.5 but stronger steady background flow.	53
2.12	A snapshot of the time-dependent interface displacement of the base of the middle layer (m) for the same set of parameters as in Figure 2.11; the numerical solution was obtained by averaging over three adjacent characteristics which introduces scale selective damping suppressing the numerical instabilities.	55

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
2.13 A snapshot of the time-dependent interface displacement of the base of the middle layer (m). The free annual period waves are forced by a wavemaker at the eastern boundary that primarily generates the short wave part of the solution. The parameters describing the background steady circulation are the same as in Figure 2.5.	58
2.14 A snapshot of the time-dependent interface displacement of the base of the middle layer (m). The free annual period waves are forced by a wavemaker at the eastern boundary that primarily generates the long wave part of the solution. The parameters describing the background steady circulation are the same as in Figure 2.5.	59
2.15 Meridional potential vorticity gradient of the steady flow in the upper layer ($10^{-13} \text{ m}^{-2} \text{ s}^{-1}$) for the solution shown in Figure 2.11.	62
2.16 Local analytical dispersion relation for four points in the southwestern part of the subtropical gyre with the fixed $x=2000$ km and varying y : 1500 km, 1502 km, 1510 km and 2500 km, for the numerical solution shown in Figure 2.11 illustrating the effect of the meridional background flow. Real zonal wavenumber k is plotted as a light solid line and the imaginary zonal wavenumber k is plotted as a dark solid line.	64
2.17 Local analytical dispersion relation for two points in the southernmost part of the EZ for the numerical solution shown and described in Figure 2.18 with coordinates $x=500$ km, $y=1550$ km (left panel) and $x=6000$ km, $y=1550$ km (right panel). The comparison of the two dispersion relations shows that the incident group of stable waves of different meridional wavelengths when arriving at the discontinuity will generate either stable or unstable waves, depending on their meridional wavelengths.	67
2.18 A snapshot of the time-dependent interface displacement of the base of the middle layer (m). The parameters used in this example are representative of a density profile in the North Pacific and the amplitude of the Ekman pumping is of realistic strength.	71
2.19 A snapshot of the sea surface elevation (cm) for the parameters as in Figure 2.18.	72

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
2.20	Time-longitude sections for the numerical solution of Figure 2.5. 74
3.1	The dispersion relation diagram obtained by fixing the frequency σ , varying the meridional wavenumber component l (real) and solving the quadratic equation (3.18) for the zonal wavenumber k (complex in general). The upper panel shows the two nondispersive long baroclinic planetary waves of the two-and-a-half layer system (dashed line) for the case of no background flow. The lower panel shows that the presence of even relatively weak background flow completely changes the dispersion relation: it couples the modes and introduces unstable solutions (the imaginary part of k is plotted as a solid line). 96
3.2	The dispersion relation (3.18): contours of frequency vs. k and l for the σ^- solution in the absence of the background flow. Dashed lines are contours of the amplitude of the initial condition in the wavenumber plane. The initial Gaussian disturbance is centered at $k = 0, l = 0$ and has the width of 94 km. 99
3.3	A snap-shot of the time-dependent interface displacement of the middle layer (z in m) for the σ^- and σ^+ part of the solution after 150 days (upper and lower panel, respectively). The σ^- solution is the classical first baroclinic Rossby wave and the σ^+ solution is the classical second baroclinic Rossby wave. Contour levels are from 0 to 1 m, with the interval 0.25 m. 101
3.4	The dispersion relation diagram for the σ^+ part of the solution in the presence of the SW background flow: contours of σ vs. k and l . Dashed lines show the contours of the amplitude of the initial Gaussian disturbance centered at $k = 0, l = 0$, with the width of 94 km. 102
3.5	A snap shot of the time-dependent interface displacement of the middle layer (z in m) as a function of x and y after 150 days in the presence of the SW background flow for σ^- (upper panel) and σ^+ (lower panel) parts of the solution. Positive (solid line) and negative (dashed line) displacement are largest in the central part of the disturbance. Contour levels are 0.5 cm, 1 cm, 2.5 cm, 5 cm, 0.25 m, 0.5 m, 0.75 m and 1m. 104

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>	
3.6	A zoom of the sum of the σ^+ and σ^- components of the time-dependent interface displacement of the middle layer (z in m) shown individually in Figure 3.5. Contour levels are the same as in Figure 3.5 with additional levels 0.25 cm and 0.125 cm.	106
3.7	The dispersion relation plot: contours of the real part of the frequency (left panel) and contours of the imaginary part of the frequency for σ^+ part of the solution (the right panel). Dashed lines show the contours of the amplitude of the initial disturbance which is a group with central wavelength 125 km modulated by a radially symmetric Gaussian of width of 94 km.	108
3.8	A zoom of the snapshot of the two separate components of the time-dependent interface displacement of the middle layer (z in m) as a function of x and y after 75 days for the parameters of Figure 3.7. . .	109
3.9	A snapshot of the time-dependent interface displacement of the middle layer of the transient solution obtained as the asymptotic solution for large time by the method of steepest descents, plotted in moving coordinates ξ and η . Caustics are shown as magenta stars. Two green stars show the double caustic.	117
3.10	Four roots H_i of the quartic equation (3.65) for $R=10$. Dotted lines show the approximate caustics; dashed lines show the more accurate expressions for the caustics. Dark solid line shows the real part of H and the light solid line shows the imaginary part of H	128
3.11	The roots of the quartic equation (3.65) as a function of β (upper panel), the roots of the cubic equations (3.75) (middle panel) and (3.84) (lower panel). Dotted and dashed lines indicate the position of the caustics as described in Figure 3.10	129
3.12	A time series of the logarithm of the middle layer interface displacement, moving with the disturbance (upper panel); it grows more rapidly than linearly as the time becomes large. The lower panel shows that the amplitude of the disturbance at the origin decays with time.	131

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
3.13 The dispersion relation for the annual frequency forcing: the real part of k is plotted as a solid line and the imaginary part of k (representing the spatial growth) is plotted as a dotted line. Dashed lines show the contours of absolute value of the initial Gaussian disturbance with the width of 32 km.	136
3.14 A snap shot of the time-dependent interface displacement of the middle layer as a function of x and y for the sympathetic solution forced with the wavemaker shown in Figure 3.13.	137
3.15 A snap shot of the time-dependent interface displacement of the middle layer as a function of x and y for the total (transient plus sympathetic) solution forced with the wavemaker shown in Figure 3.13.	139
3.16 The dispersion relation for the annual frequency forcing: the real part of k is plotted as a solid line and the imaginary part of k (representing the spatial growth) is plotted as a dotted line. Dashed lines show the contours of absolute value of the initial Gaussian disturbance with the width of 125 km.	140
3.17 A snap shot of the time-dependent interface displacement of the middle layer as a function of x and y for the sympathetic solution forced with the wavemaker shown in Figure 3.16.	141

LONG BAROCLINIC PLANETARY WAVES IN A NONZONAL VERTICALLY SHEARED MEAN FLOW

1. INTRODUCTION

The overall goal of this thesis is to study how long, baroclinic planetary waves are influenced by features of the underlying steady circulation. In the mid nineties satellite altimetry data first provided synoptic maps of the sea level for all the ocean basins of sufficient quality to permit analysis of the long Rossby wave signal. Among the first resulting papers was Chelton and Schlax (1996, hereafter CS), who examined sea level fluctuations in the annual-period band using TOPEX-Poseidon altimeter data. In both the Pacific and the Atlantic, at all latitudes analyzed by CS, each spatial extremum of the annually filtered signal propagates westward away from the eastern boundary of the ocean, its amplitude first decreasing and then increasing as the signal crosses into the western half of the basin. Comparison with the standard theory of free, linear planetary waves revealed significant discrepancies, the most important being observed phase speeds up to twice those of the standard theory (CS).

This result provided motivation for a number of theoretical studies. Killworth et al. (1997) showed that mean circulation could modify the planetary waves and cause enhanced zonal phase speeds. De Szoeke and Chelton (1999) examined the effect of constant potential vorticity regions in finite depth models of the large

scale circulation, showing that they were an important factor in the speedup. Dewar (1998) analyzed propagation of the annual frequency baroclinic planetary waves in the presence of the mean shear using a quasigeostrophic analysis with possible extensions to a layered planetary geostrophic system. Related studies of low frequency flow perturbations were carried out by Liu (1999), Huang and Pedlosky (1999), and Dewar and Huang (2001). Qiu et al. (1997) analyzed the propagation and decay of free and forced baroclinic Rossby waves in the off-equatorial oceans with emphasis on the influence of eddy dissipation on the sea surface height anomalies. White et al. (1998) considered biennial waves. Dewar and Morris (2000) analyzed the propagation of long, first mode baroclinic planetary waves in an eddy-resolving QG model.

The westward amplitude increase and the phase velocity discrepancy of the near-annual period Rossby waves was the initial motivation for this work. In the second chapter of this thesis we therefore consider annual period long Rossby waves. A steady, idealized anticyclonic Ekman pumping drives a steady circulation in a two-and-a-half layer planetary geostrophic model of the ocean, to be described in section 2.2. In this circulation an eastern and a western zone (EZ, WZ) can be distinguished. The EZ is a shadow zone of the eastern boundary, so that only the uppermost wind-driven layer is directly in motion. The WZ is a region of recirculation into and out of the western boundary in which potential vorticity is taken to be homogenized, motivated by a similar feature of the observed large-scale circulation (Keffer, 1985, Talley, 1988). This steady solution is linearly perturbed by an annually varying wind stress (section 2.3). The perturbation equations are set up and solved for the time-dependent interface displacements of the upper and the middle layer by the method of characteristics (section 2.4). The integration was performed westward numerically using a modified Runge-Kutta procedure (section 2.5). An example of

the numerical solutions is given and discussed in section 2.6. Approximate WKB solutions are worked out in section 2.7. The numerical scheme for solving across discrete coupled characteristics is examined using a method similar to the classical von Neumann analysis. This analysis reveals the possibility of numerical instabilities that may frustrate attempts to improve the solution by simply refining the numerical grid (section 2.8). In the EZ the perturbation solution consists of two sets of annual period baroclinic waves, each a combination of the classical first and the second baroclinic Rossby wave coupled by the background flow. In the WZ the propagation of information from the western boundary of the ocean is neglected; therefore the WZ solution consists of only the long set of baroclinic waves. For sufficiently strong background flow the annual period perturbation solution in the southwestern part of the ST gyre is baroclinically unstable and grows westward. These and other features of the annually varying solution are discussed in section 2.9. A summary (section 2.10) completes the chapter.

Much of the important classical work regarding baroclinic instability deals with purely zonal flow and seeks growth rates of spatially periodic disturbances. Pedlosky (1987, 2002) has pointed how nonzonal flow, such as that occurring in the present ST gyre, may enhance baroclinic instability. But what excites the baroclinic instability in the southwestern part of the ST gyre in the solutions of chapter two was not entirely clear. The numerical solutions indicated the existence of a region in the far southwest subtropical gyre where annual period perturbations grew zonally towards the west. Such spatially growing instabilities have been studied extensively in other contexts (Briggs 1964, Huerre and Monkewitz, 1990). The three possibilities for excitation of such disturbances in the present problem are that:

(i) the perturbation solution generated by local winds is indeed unstable and grows westward;

(ii) the perturbation solution has propagated westward through the inhomogeneous background flow from more easterly regions of wind stress and has been refracted into unstable wavenumbers that grow westward;

(iii) the perturbation solution, though physically stable, appears to grow westward because of numerical instability.

The last two possibilities are examined in the second chapter of this thesis. The third chapter of this thesis focuses on the first possibility. Since we analyse perturbations whose horizontal scales are much smaller than the basin-wide scales of variation of large scale circulation, the spatial inhomogeneity of the background flow is neglected. This results in a solution that is remarkably simple analytically though showing a number of previously unemphasized features. We thus consider perturbations of a spatially homogeneous nonzonal upper ocean flow that are generated by a specified initial perturbation or by harmonic forcing that starts at some initial time. Solutions in a periodic domain are found by numerically summing Fourier components, and are compared with solutions obtained by the method of steepest descents. Even very gentle nonzonal background flow dramatically changes the perturbations away from the simple β solutions. In the first case a Gaussian shaped initial disturbance evolves in a highly anisotropic manner. The growth of the amplitude of this transient disturbance depends strongly on the central wavenumber and spectral width of the initial disturbance. In the latter case the solution consists of a forced part varying at the annual period plus a transient part; the solution is again unstable.

**2. LONG PLANETARY WAVES IN A TWO-AND-A-HALF LAYER
MODEL OCEAN WITH A WIND-DRIVEN STEADY CIRCULATION**

Ivana Cerovečki and Roland A. de Szoeke

Submitted to *Journal of Physical Oceanography*

Abstract

To understand how the structure of the annual period planetary wave signal is affected by the underlying large scale circulation, we consider a model of long planetary waves linearized about a simple, steady, wind-driven subtropical (ST) gyre circulation in two layers overlying a very thick and quiescent third layer. The circulation divides the gyre into an eastern zone (EZ) where only the surface layer is in steady motion and a western zone (WZ) where both the surface and the middle layer are in steady motion because the potential vorticity of the middle layer is assumed to be homogenized. In both zones the upper two layers participate in the wave motion which is driven by wind stress fluctuating with the annual period, representing the seasonal cycle.

The equations governing the waves are spatially hyperbolic, and are numerically integrated along their characteristics. In the EZ the solution consists of two sets of baroclinic waves which are a combination of the classical first and the second baroclinic Rossby wave modes coupled by the background flow. In contrast to the case of quiescent background flow, each is strongly dispersive. (The rigid lid approximation filters out the barotropic mode in the long wave limit.) The local analytical WKB dispersion relation shows that one set of waves is much shorter than the other or the forcing, so it can have wavelengths small enough that relative vorticity may become important. For sufficiently strong background flow the waves become baroclinically unstable in the southwestern part of the ST gyre and grow westward. Baroclinic instability may be generated by local wind forcing or energy incident from the east. In the WZ the propagation of information from the western boundary of the ocean is neglected because the model cannot capture the physics of

the western boundary current; therefore the WZ solution consists of only one (long) set of baroclinic waves.

The numerical procedure for solving across the coupled characteristics is examined using an extension of the classical von Neumann analysis. The procedure can give rise to spurious instabilities, which occur when the aliased short wave resonates with the long wave. These are suppressed by a moderate amount of averaging over adjacent characteristics. In the southwest corner of the ST gyre the short wave part of the solution has wavelengths of the order of 10-100 km; these are prone to generate numerical instabilities via aliasing so that careful analysis is needed to distinguish this possibility from true baroclinic instability.

2.1. Introduction

Sea level fluctuations in the annual-period band have been examined by Chelton and Schlax (1996, hereafter CS) using TOPEX-Poseidon altimeter data. Figure 2.1 shows time-longitude plots of sea level bandpassed around the annual period for several latitudes in the Pacific subtropical gyre. Along the eastern boundary, positive and negative amplitudes alternate in time with nearly annual period. Each of these extrema propagates westward, its amplitude first decreasing and then increasing as the signal crosses into the western half of the basin. This is seen at the latitudes of the time longitude plots of Figure 2.1 in both the Pacific and the Atlantic. Comparison to the standard theory of free, linear planetary waves revealed significant discrepancies, the most important being observed phase speeds up to twice those of the standard theory (CS).

Killworth et al.(1997) showed that mean circulation could modify the planetary waves and cause enhanced zonal phase speeds. De Szoeke and Chelton (1999) examined the effect of constant potential vorticity regions in finite depth models of the large scale circulation (Keffer, 1985, Talley, 1988), showing that they were an important factor in the speedup. Dewar (1998) analyzed propagation of the annual frequency baroclinic planetary waves in the presence of the mean shear using a quasigeostrophic analysis with possible extensions to a layered planetary geostrophic system. Related studies of low frequency flow perturbations were carried out by Liu (1999), Huang and Pedlosky (1999), and Dewar and Huang (2001). Qiu et al. (1997) analyzed the propagation and decay of free and forced baroclinic Rossby waves in the off-equatorial oceans with emphasis on the influence of eddy dissipation on the sea surface height anomalies. White et al. (1998) considered biennial waves. Dewar

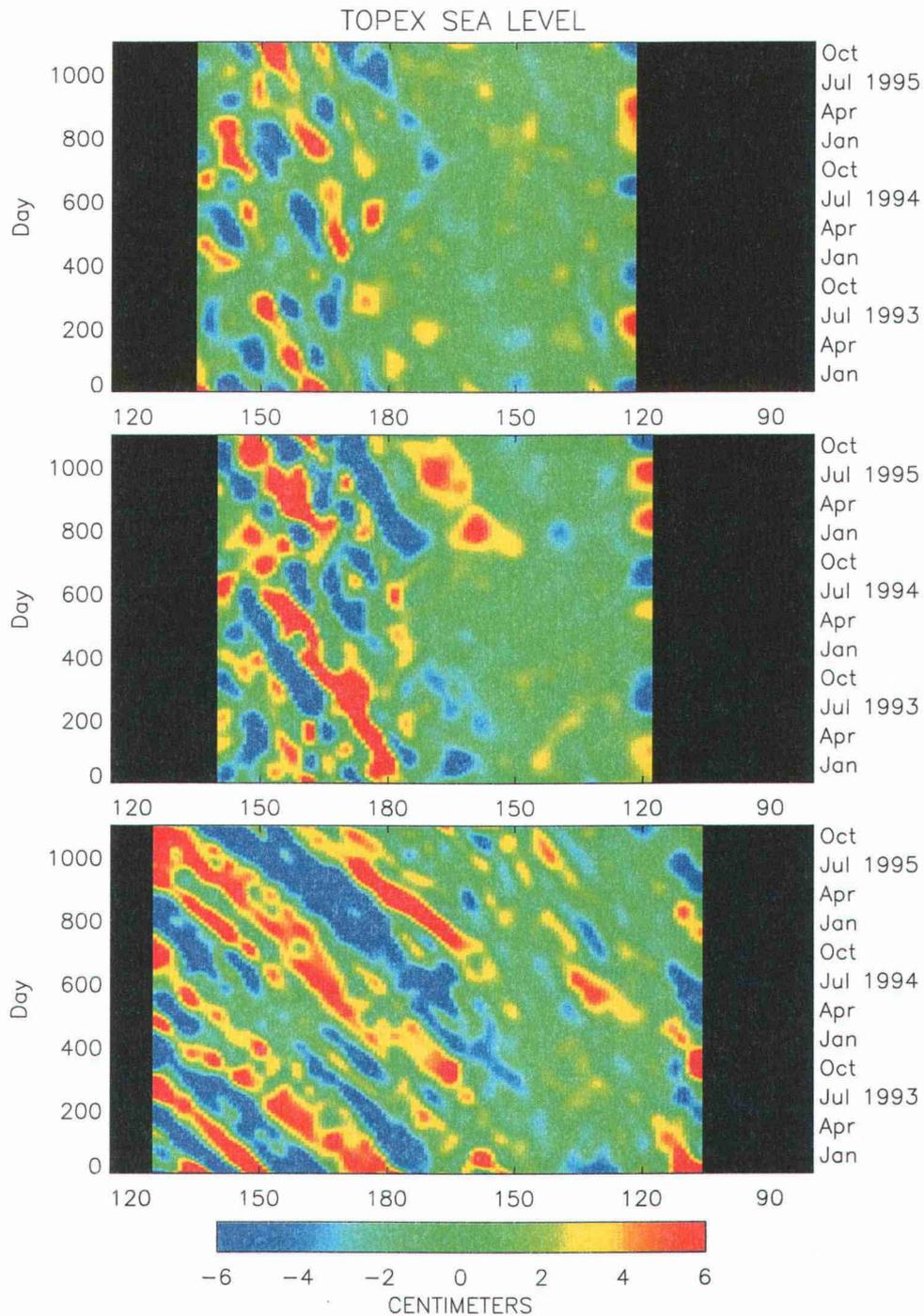


FIGURE 2.1. Time-longitude sections of annually filtered sea level for the midlatitude North Pacific. A prominent feature is the amplification of midlatitude Rossby waves in the western half of the basin (Chelton and Schlax, 1996, Figure 2).

and Morris (2000) analyzed the propagation of long, first mode baroclinic planetary waves in an eddy-resolving QG model.

The westward amplitude increase and the phase velocity discrepancy of the near-annual period Rossby waves provided the initial motivation for this work. The overall goal is to understand how long, baroclinic planetary waves are influenced by features of the underlying steady circulation. In this paper a steady, idealized anticyclonic Ekman pumping drives a steady circulation in a two-and-a-half layer planetary geostrophic model of the oceans, to be described in section 2.2. In this circulation an eastern and a western zone (EZ,WZ) can be distinguished. The EZ is a shadow zone of the eastern boundary, so that only the uppermost wind-driven layer is directly in motion. The WZ is a region of recirculation into and out of the western boundary in which potential vorticity (PV hereafter) is taken to be homogenized, motivated by a similar feature of the observed large-scale circulation (Keffer, 1985, Talley, 1988). This steady solution is linearly perturbed by an annually varying wind stress (section 2.3). The perturbation equations are set up and solved by the method of characteristics giving the time-dependent interface displacements of the upper and the middle layer (section 2.4). The integration was performed westward numerically using a modified Runge-Kutta procedure, as outlined in section 2.5. An example of the numerical solutions is given and discussed in section 2.6. Approximate WKB solutions are worked out in section 2.7. The numerical scheme for solving across discrete coupled characteristics is examined using a method similar to the classical von Neumann analysis (section 2.8). The features of the annually varying solution are discussed in section 2.9. A summary (section 2.10) completes the paper.

2.2. The Model

A layered planetary geostrophic model (Luyten et al., 1983, hereafter LPS) was used to model the steady circulation of the subtropical gyre (though with no outcrops in the region of consideration). Three layers; upper, middle and lower, labeled 1, 2 and 3 (Figure 2.2), are considered. The depth of the lower layer will ultimately be allowed to become very large. The density ρ_j is constant in layer j . The thickness of layer j is denoted by h_j ($j=1,2,3$), and the depth of the base of layer j is $z = -H_j = -\sum_{n=1}^j h_n$. The model describes the region below the surface mixed layer, and layer 1 is directly forced by the Ekman pumping w_E . Motion in each layer is geostrophic,

$$-fv_j = -\partial_x P_j, \quad fu_j = -\partial_y P_j, \quad (2.1)$$

where u_j and v_j are the components of the horizontal velocity in layer j , $P_j = (p_j + \rho_j g z) / \rho_0$ is the Montgomery function, whose horizontal gradient depends only on the lateral position (x, y) within each layer; p_j being the true dynamic pressure. The Coriolis parameter is linear in y , $f(y) = f_0 + \beta y$. Using the hydrostatic relation, the Montgomery functions for the upper and the middle layer can be expressed as

$$P_1 = \gamma_1 H_1 + \gamma_2 H_2 + P_3, \quad P_2 = \gamma_2 H_2 + P_3, \quad (2.2)$$

and for simplicity we shall hereafter refer to them as "pressure". Here

$$\gamma_j = g \frac{(\rho_{j+1} - \rho_j)}{\rho_0} \quad j = 1, 2 \quad (2.3)$$

are the reduced gravity parameters. The mass conservation equation gives

$$\partial_t h_j + J(P_j, \frac{h_j}{f}) = \begin{cases} -w_E & \text{for } j=1 \\ 0 & \text{for } j=2,3. \end{cases} \quad (2.4)$$

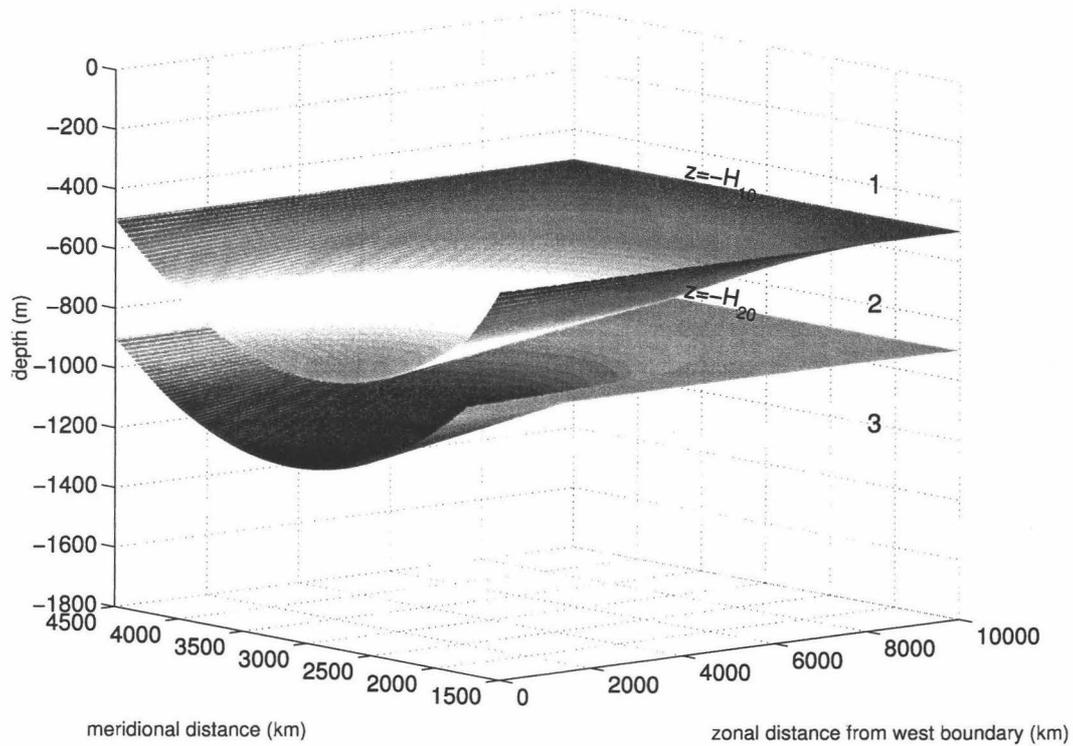


FIGURE 2.2. Three-dimensional perspective view of the layer base depth displacements due to the steady wind forcing looking into the basin from the southwest corner of the subtropical gyre. Only the upper layer is exposed to the Ekman pumping and the two lower layers do not outcrop. H_{10} and H_{20} are the upper and the middle layer depths at the eastern boundary. The base of the middle layer is flat in the region close to the eastern boundary, the middle layer has constant potential vorticity in the western part of the gyre. Parameters used in this example are: $H_{10}=500$ m, $H_{20}=900$ m, $\gamma_1 = 0.03$ m s⁻² and $\gamma_2 = 0.01$ m s⁻². The amplitude of the steady Ekman pumping defined in (2.7) $W_0=9 \cdot 10^{-7}$ m s⁻¹. The zonal extent of the basin is 10000 km and the meridional extent is 3000 km.

The sum of equations (2.4) gives the Sverdrup transport relation for the meridional geostrophic transport. This equation can be integrated from the eastern boundary (at $x = a$) westward to give

$$2H_3P_3 + \gamma_1H_1^2 + \gamma_2H_2^2 = \Theta(x, y, t) + \gamma_1H_{10}^2 + \gamma_2H_{20}^2 + 2H_3P_{30}, \quad (2.5)$$

where

$$\Theta(x, y, t) = -\frac{2f^2}{\beta} \int_x^a w_E(x', y, t) dx' \quad (2.6)$$

is the zonal integral of Ekman pumping from the eastern boundary $x = a$, H_{10} and H_{20} are the depths of the bases of the upper and the middle layers and P_{30} is the lowest layer Montgomery function, all at $x = a$. H_3 is a constant total depth. To satisfy the no normal flow condition, H_{10} and H_{20} are chosen to be constant along the eastern boundary.

2.2.1. Steady Model

Suppose steady anticyclonic Ekman pumping, a function of latitude only,

$$w_E^{(0)}(y) = -W_0 \left(\frac{f_0}{f} \right)^2 \sin \left[\frac{(y-b)\pi}{\Delta} \right] \quad (2.7)$$

acts on the upper layer. Superscript 0 denotes the variables related to the steady circulation. The basin occupies $b/3 < y < b$ so that $\Delta = 2b/3$ is the meridional extent of the basin. Ekman pumping is thus zero at the northern and the southern edges of the subtropical gyre.

The second of the planetary geostrophic continuity equations (2.4) shows that the PV of layer 2 is conserved, hence

$$\frac{h_2^{(0)}}{f} = \Gamma_2(P_2) \quad (2.8)$$

where $h_j^{(0)}/f$ is the potential thickness of layer j , the reciprocal of the planetary geostrophic potential vorticity. Γ_2 is determined (in part) as follows. Because $\Theta^{(0)}(x, b) = 0$, so that $H_1^{(0)} = H_{10}$ and $H_2^{(0)} = H_{20}$ along $y = b$, the northernmost potential thickness contour along which $\Gamma_2 = \frac{H_{20} - H_{10}}{f(b)}$, runs along the northern edge of the subtropical gyre $y = b$. It bifurcates at the Rossby Attractor (RA) point, where the eastward mean Sverdrup flow $u_S = u_1^{(0)} = -\frac{\gamma_1}{f} \frac{\partial H_1^{(0)}}{\partial y}$ carried by the surface layer exactly balances the westward Rossby-wave speed $c_R = -\frac{\beta \gamma_1 H_1^{(0)} (H_{20} - H_1^{(0)})}{f^2 H_{20}}$ (Luyten and Stommel, 1986, de Szoeke, 1992). To the west of RA the northern branch of this potential thickness contour continues along $y = b$, while the southern branch follows a locus $x_B(y)$ whose form will be established subsequently. The splitting of this PV isopleth at the RA point is a consequence of the indeterminacy of the direction of information propagation at this point where both $u_S + c_R$ and the vertical average of the meridional component of the Sverdrup transport vanish (Pedlosky, 1996). In this way the steady circulation divides the basin into an eastern zone (EZ) and a western zone (WZ), respectively east and west of $x_B(y)$. The function Γ_2 need not be the same in both zones. In both zones the solution satisfies equation (2.8) and the steady counterpart of the Sverdrup transport equation (2.5).

2.2.2. East zone steady solution

Since there is no normal flow at the eastern boundary and the middle layer is not directly exposed to the Ekman pumping, there is no flow at any point in that layer that can be reached from the eastern boundary along isopleths of its potential thickness. Hence $P_2^{(0)}/\gamma_2 = H_2^{(0)} = H_{20}$, constant. Only $H_1^{(0)}$ varies, as given by the Sverdrup equation (2.5): $H_1^{(0)} = \sqrt{\frac{\Theta^{(0)}(x, y)}{\gamma_1} + H_{10}^2}$. This fixes the circulation in the EZ, which is the analogue of the eastern shadow zone region of the LPS model.

The contours of potential thickness of the middle layer, $\frac{H_2^{(0)}-H_1^{(0)}}{f}$, that emanate from the eastern boundary (so defining the EZ) lie in the range $\frac{H_{20}-H_{10}}{f(b)}$, $\frac{H_{20}-H_{10}}{f(b/3)}$. Suppose a potential thickness contour emanates from the eastern boundary at latitude y_0 . This contour is described in the ocean interior by

$$\frac{H_{20} - \sqrt{\frac{\Theta^{(0)}(x,y)}{\gamma_1} + H_{10}^2}}{f} = \frac{H_{20} - H_{10}}{f(y_0)}. \quad (2.9)$$

The left side of (2.9) is the distribution of potential thickness in the middle layer as a function of x and y . The right side is the particular value at y_0 on the eastern boundary. The inverse potential thickness function Γ_2^{-1} (equation 2.8) maps all these onto the single value $P_2 = \gamma_2 H_{20}$.

The northernmost potential thickness contour with value $\Gamma_0 = \frac{H_{20}-H_{10}}{f(b)}$, occurs along $y=b$. However a branch of the contour with this value is given by

$$\frac{H_{20} - \sqrt{\frac{\Theta^{(0)}(x_B,y)}{\gamma_1} + H_{10}^2}}{f(y)} = \Gamma_0, \quad (2.10)$$

for y different from b . With the particular form for $\Theta^{(0)}$ obtained by inserting (2.7) into (2.6), this gives an expression for the contour $x_B(y)$ which marks the western boundary of the EZ:

$$x_B(y) - a = \frac{\gamma_1 a}{A \sin \left[\frac{(y-b)\pi}{\Delta} \right]} \left\{ \left[H_{20} \left(1 - \frac{f}{f(b)} \right) + \frac{f}{f(b)} H_{10} \right]^2 - H_{10}^2 \right\}. \quad (2.11)$$

As $y \rightarrow b$, both numerator and denominator of (2.11) go to zero. Hence, by L'Hospital's rule, the distance from the eastern boundary at which $x_B(y)$ intersects $y = b$ is

$$a - x_B(b) = \frac{2a\Delta\beta\gamma_1 H_{10}(H_{20} - H_{10})}{A\pi f(b)}. \quad (2.12)$$

The slope at which $x_B(y)$ intersects $y=b$ is given by

$$\frac{\partial x_B}{\partial y}(b) = \frac{4a\Delta}{\pi} \frac{\gamma_1}{\gamma_2} \left(\frac{\beta}{f(b)} \right)^2. \quad (2.13)$$

2.2.3. West zone steady solution

West of the RA, the two branches ($x = x_B(y)$ and $y = b$) of the bounding potential thickness contour with potential thickness Γ_0 enclose a region which we call the western zone (WZ). In this zone, following Young and Rhines (1982) and LPS (1983), and being motivated by observations (Keffer, 1985, Talley, 1988), we take potential thickness to be homogenized with the specified value Γ_0 (equation (2.10)). Hence

$$H_2^{(0)} = H_1^{(0)} + f\Gamma_0. \quad (2.14)$$

(More generally in the WZ, $\frac{H_2^{(0)} - H_1^{(0)}}{f} = \Gamma_2(\gamma_2 H_2^{(0)})$ where the function Γ_2 should be determined by western boundary current processes since the potential thickness contours of the middle layer in the WZ originate at the western boundary of the basin in the northern half of the WZ and rejoin the western boundary in the southern half of the basin (Figure 2.3).)

The potential thickness function Γ_2 in the WZ maps all the P_2 values onto the constant Γ_0 . An equation for $H_1^{(0)}$ is obtained by substituting (2.14) into the Sverdrup constraint (2.5); (2.14) then gives $H_2^{(0)}$.

Equation (2.5) shows that the requirement for continuous interface displacements and continuous wind forcing across the boundary between the zones gives continuous vertically integrated transport across $x_B(y)$. Since vertically integrated transport is confined to the upper layer in the EZ, but is distributed over the upper two layers in the WZ, the tangential velocity along x_B may be discontinuous across the boundary between the zones.

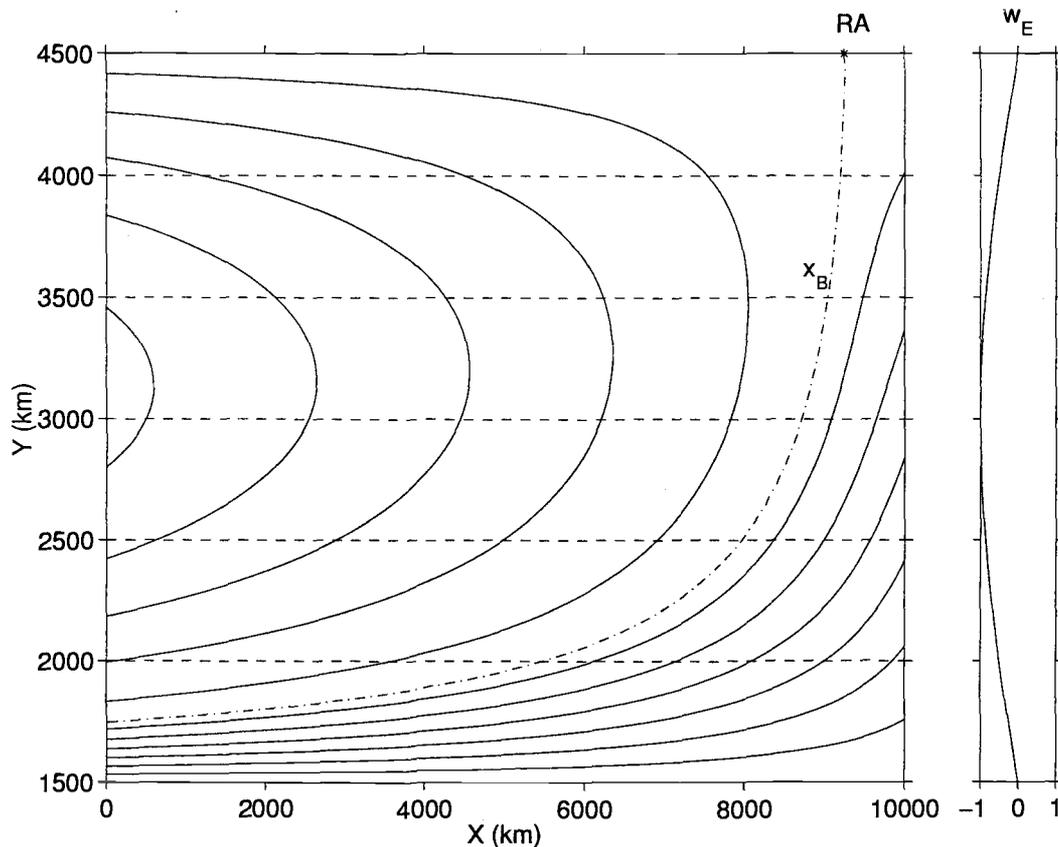


FIGURE 2.3. Characteristics of the equations governing annual perturbations about the mean flow (left panel) and the spatial structure of both the steady and the time-dependent Ekman pumping (right panel). Dash-dot line denoted by x_B is the boundary between the EZ and the WZ. The star denotes the Rossby Attractor point (RA). In both zones perturbation equations have two sets of characteristics. The first set (solid lines) are constant potential vorticity contours in the EZ and isopleths of middle layer base depth in the WZ. In both zones the second set of the characteristics are constant latitude lines, plotted as a dashed line. In the EZ both sets of characteristics emanate from the eastern boundary, in the WZ constant latitude lines emanate from the east, while the other set of characteristics emanates from the western boundary, so that the perturbation solution may be discontinuous along x_B . The solution at any point in the EZ depends on the Ekman pumping within the triangular region formed by the characteristics passing through that point and the eastern boundary values along the eastern wall of the triangular region. (The characteristics shown in this figure are for the steady flow example shown in Figure 2.2.)

2.2.4. An illustration of the steady solution

Figure 2.2 illustrates the steady solution that defines the background circulation. It shows a three-dimensional perspective view of the depths of the bases of the upper two layers, looking into the basin from the southwest corner of the subtropical gyre. The depth of the uppermost layer is constant along the eastern boundary and deepens westward. The depth of the base of the middle layer is constant everywhere in the EZ and deepens westward in the WZ.

The parameters used in this example are chosen to give a qualitative representation of the density profile in the north Pacific subtropical gyre: $H_{10}=500$ m, $H_{20}=900$ m, $\gamma_1 = 0.03$ m s⁻², $\gamma_2 = 0.01$ m s⁻². Since in the EZ only the upper layer moves and carries the whole Sverdrup transport, a somewhat large value for H_{10} has been chosen; with considerably smaller upper layer depth the flow is subsequently found to be unstable in the greater part of the entire EZ.

The amplitude of the Ekman pumping velocity defined in (2.7) is $W_0=9.1 \cdot 10^{-7}$ m s⁻¹. The latitudes of zero wind forcing representing the northern and the southern edge of the subtropical gyre are at $y = 4500$ km ($=b$) and $y = 1500$ km, so that the meridional extent of the basin is 3000 km. The zonal width of the basin is $a = 10\,000$ km.

2.3. The perturbation equations

The steady circulation described in the previous section is perturbed by a time-dependent flow driven by a time-varying wind stress curl with the same spatial structure as the steady wind stress curl, but fluctuating with the annual period. We thus consider perturbations (denoted by superscript 1) about the steady solution (denoted by superscript 0):

$$w_E = w_E^{(0)}(x, y) + w_E^{(1)}(x, y, t), \quad w_E^{(1)} = w_E^{(0)}(x, y) \sin(\sigma t) \quad (2.15)$$

$$H_j = H_j^{(0)}(x, y) + H_j^{(1)}(x, y, t) \quad j = 1, 2, 3, \quad (2.16)$$

$$P_j = P_j^{(0)}(x, y) + P_j^{(1)}(x, y, t) \quad j = 1, 2, 3. \quad (2.17)$$

Equations (2.2) relate $H_j^{(1)}$ and $P_j^{(1)}$, just like their steady counterparts. Put (2.15), (2.16) and (2.17) into the Sverdrup relation (2.5), linearize about the steady state and subtract the steady Sverdrup relation to obtain

$$2H_3P_3^{(1)} + 2\gamma_1H_1^{(0)}H_1^{(1)} + 2\gamma_2H_2^{(0)}H_2^{(1)} = \Theta^{(1)} + 2\gamma_1H_{10}^{(0)}H_{10}^{(1)} + 2\gamma_2H_{20}^{(0)}H_{20}^{(1)} + 2H_3P_{30}^{(1)} \quad (2.18)$$

where $H_{10}^{(1)}$ and $H_{20}^{(1)}$ are the time-dependent interface displacements of the upper and the middle layer along the eastern boundary and $P_{30}^{(1)}$ is the time dependent pressure in the lowest layer along the eastern boundary. The requirement of no flow through eastern boundary furnishes the boundary conditions $H_{20}^{(1)} = H_{10}^{(1)} = P_{30}^{(1)} = 0$. The zonal integral of time-dependent Ekman pumping $\Theta^{(1)}$ is given by the equation (2.6) with the time-varying Ekman pumping velocity $w_E^{(1)}$ on the rhs.

After linearization about the steady circulation, equations (2.4) for the middle and the lowest layers become

$$\partial_t(H_2^{(1)} - H_1^{(1)}) + J \left(P_2^{(0)}, \frac{H_2^{(1)} - H_1^{(1)}}{f} \right) + J \left(P_2^{(1)}, \frac{H_2^{(0)} - H_1^{(0)}}{f} \right) = 0 \quad (2.19)$$

and

$$\partial_t(-H_2^{(1)}) + J \left(P_3^{(1)}, \frac{H_3 - H_2^{(0)}}{f} \right) = 0. \quad (2.20)$$

The resulting equations determine $H_1^{(1)}$ and $H_2^{(1)}$ and completely describe the perturbation flow. Equations (2.2) and (2.18) can be used to eliminate $P_j^{(1)}$ in favor of $H_j^{(1)}$ in (2.19) and (2.20).

We now reduce our model to two layers lying over an infinitely deep quiescent third layer in both zones, which gives the so-called two-and-a-half layer model.

2.3.1. Perturbation equations in the EZ

Since there is no steady motion in the middle layer in the EZ, $H_2^{(0)} = H_{20} = \text{constant}$ and $P_2^{(0)} = 0$. The Sverdrup relation (2.18) thus reduces to

$$2H_3P_3^{(1)} + 2\gamma_1H_1^{(0)}H_1^{(1)} + 2\gamma_2H_{20}H_2^{(1)} = \Theta^{(1)}. \quad (2.21)$$

The middle layer equation (2.19) in the EZ becomes

$$\partial_t(H_2^{(1)} - H_1^{(1)}) + J\left(P_2^{(1)}, \frac{H_{20} - H_1^{(0)}}{f}\right) = 0. \quad (2.22)$$

To get the lowest layer equation (2.20) in the EZ, solve the Sverdrup equation (2.21) for $P_3^{(1)}$, insert into (2.20) and let $H_3 \rightarrow \infty$, which gives

$$\partial_t H_2^{(1)} - \frac{\beta\gamma_2 H_{20}}{f^2} \partial_x H_2^{(1)} - \frac{\beta\gamma_1 \partial_x (H_1^{(0)} H_1^{(1)})}{f^2} = -\frac{\beta\Theta_x^{(1)}}{2f^2}. \quad (2.23)$$

Equations (2.22) and (2.23) determine $H_1^{(1)}$ and $H_2^{(1)}$ in the EZ.

2.3.2. Perturbation equations in the WZ

In the WZ the Sverdrup relation (2.18) becomes

$$2H_3P_3^{(1)} + 2\gamma_1H_1^{(0)}H_1^{(1)} + 2\gamma_2H_2^{(0)}H_2^{(1)} = \Theta^{(1)}. \quad (2.24)$$

Since the assumption that the middle layer has constant PV in the WZ links the steady layer depths as shown in (2.14), equations (2.19) and (2.20) take on the following form:

$$\partial_t(H_2^{(1)} - H_1^{(1)}) + J\left(H_2^{(0)}, \frac{\gamma_2(H_2^{(1)} - H_1^{(1)})}{f}\right) = 0 \quad (2.25)$$

and

$$\partial_t H_2^{(1)} - \frac{\beta}{f^2} [\gamma_1 \partial_x (H_1^{(0)} H_1^{(1)}) + \gamma_2 \partial_x (H_2^{(0)} H_2^{(1)})] = -\frac{\beta \Theta_x^{(1)}}{2f^2}, \quad (2.26)$$

where $P_2^{(0)} = \gamma_2 H_2^{(0)}$ has been used. In (2.26) $P_3^{(1)}$ was eliminated from (2.24) with $H_3 \rightarrow \infty$. Equations (2.25) and (2.26) determine $H_1^{(1)}$ and $H_2^{(1)}$ in the WZ.

2.3.3. Sea surface displacement

An expression can be obtained for the sea surface displacement η in the two-and-a-half layer limit. This is useful for comparison to the annually varying field of sea surface elevation obtained from satellite altimetry.

The time-dependent part of Montgomery function for the upper layer $P_1^{(1)}$ is related to sea surface displacement by the hydrostatic relation $P_1^{(1)} = g\eta$. In the limit $H_3 \rightarrow \infty$, the transient Sverdrup relation (2.18) gives $P_3^{(1)} \rightarrow 0$, so that eq. (2.2) for $P_1^{(1)}$ gives

$$\eta = \frac{1}{g} (\gamma_1 H_1^{(1)} + \gamma_2 H_2^{(1)}). \quad (2.27)$$

2.4. Characteristics of the governing equations for $H_1^{(1)}$ and $H_2^{(1)}$:

Suppose the ocean is forced at frequency σ and the time-dependent Ekman pumping has the same spatial structure as the steady wind

$$\Theta^{(1)} = \Theta^{(0)} e^{-i\sigma t}. \quad (2.28)$$

Because the problem is linear, any amplitude for $\Theta^{(1)}$ can be assigned, but we suppress it for simplicity. Then the ocean layer depths oscillate similarly in time,

$$(H_1^{(1)}, H_2^{(1)}) = (w, z)e^{-i\sigma t}, \quad (2.29)$$

where the variables w, z are functions of position (x, y) only.

We solve the systems of equations (2.22) and (2.23) in the EZ, (2.25) and (2.26) in the WZ by the method of characteristics.

For notational simplicity abbreviate

$$C(x, y) = \frac{\beta\gamma_1 H_1^{(0)}}{f^2}, \quad c(y) = \frac{\beta\gamma_2 H_2^{(0)}}{f^2}, \quad (2.30)$$

$$\Psi(x, y) = \frac{\gamma_2(H_{20} - H_1^{(0)})}{f}, \quad (2.31)$$

$$\tilde{\Theta}(x, y) = -\frac{\beta\partial_x\Theta^{(0)}}{2f^2}. \quad (2.32)$$

2.4.1. In the EZ

Ψ is proportional to middle layer potential thickness. Isopleths of potential thickness are called geostrophic contours (Rhines and Young, 1982). Its x and y gradients give, respectively, the mean geostrophic meridional velocity V , and the sum of the mean geostrophic zonal velocity U and $\frac{\gamma_2}{\gamma_1}C - c$, a Rossby-wave speed. (In our case these gradients of Ψ give $\frac{\gamma_2}{\gamma_1}U$ and $\frac{\gamma_2}{\gamma_1}V$ since γ_2 was used in the definition of Ψ .) From

$$U = \Psi_y, \quad V = -\Psi_x \quad (2.33)$$

the ratio

$$\frac{V}{U} = \left(\frac{\partial y}{\partial x} \right) \Big|_{\Psi} \quad (2.34)$$

gives the slope of PV_2 isopleths.

Rewrite the EZ perturbation equations (2.22) and (2.23) using (2.29) - (2.32)

as

$$i\sigma z - J(z, \Psi) = i\sigma w \quad (2.35)$$

$$i\sigma z + cz_x + (Cw)_x = -\tilde{\Theta}. \quad (2.36)$$

where $P_2^{(1)} = \gamma_2 z e^{-i\sigma t}$ has been used. Eliminate w between these to get a second order differential equation for z :

$$CUz_{xx} + CVz_{xy} + [(CU)_x - i\sigma(c + C)]z_x + (CV)_xz_y + [\sigma^2 - i\sigma C_x]z = i\sigma\tilde{\Theta}. \quad (2.37)$$

Equation (2.37), hyperbolic in the space coordinates x, y everywhere in the basin, possesses two sets of real characteristics (Guenther and Lee, 1996), given by

$$\partial_x y = \begin{cases} 0 \\ V/U. \end{cases} \quad (2.38)$$

These are lines of constant latitude and lines of constant middle layer potential thickness, Ψ . We denote the former set by $y(x; y_0) \equiv y_0$ and the latter by $Y(x; y_0)$, where $Y(a; y_0) = y_0$ is the originating latitude of a Ψ isopleth at the eastern boundary. There is a unique relation between y_0 and $\Psi = \frac{\gamma_2(H_{20} - H_{10})}{f(y_0)}$ at the eastern boundary.

The governing equations (2.35) and (2.36) for waves in the EZ can be integrated westward along the two sets of the characteristics defined by (2.38) and shown in Figure 2.3. Their solution describes the propagation of two baroclinic modes, coupled by the baroclinic mean flow. In the case of no mean flow, they reduce to the modes where the interfaces are in phase (sinuous) and antiphase (varicose) in the two layers, although no such simple characterization is always possible in the coupled situation.

Time dependent interface displacement amplitude of the middle layer $H_2^{(1)}$ is obtained from the equation (2.35), which is written along its characteristics as a first order ordinary differential equation in x :

$$\partial_x \mathcal{Z}|_{y_0} = i\sigma(\mathcal{Z} - \mathcal{W}) \frac{\frac{\partial Y}{\partial y_0}|_x}{\frac{\partial \Psi}{\partial y_0}|_x}. \quad (2.39)$$

Here $\mathcal{Z}(x, y_0) = z(x, Y(x; y_0))$ and $\mathcal{W}(x, y_0) = w(x, Y(x; y_0))$ are $H_2^{(1)}$ and $H_1^{(1)}$ amplitudes considered as functions of x along the Y characteristics. (Observe that we have notated in upper case all the dependent variables defined on the set of the characteristics denoted by $Y(x; y_0)$. In contrast, we will notate in lower case all the dependent variables defined on the set of constant latitude lines denoted by $y(x; y_0)$; see Figure 2.4.) The $\frac{\partial Y}{\partial y_0}$ term (equal to one along the eastern boundary) indicates whether the characteristics are converging or diverging meridionally, as a function of x .

Equation (2.36) can be written along its characteristics, lines of constant latitude, as

$$\frac{\partial \phi}{\partial x}|_y = i\sigma z + \tilde{\Theta} \quad (2.40)$$

where

$$\phi = -cz - Cw. \quad (2.41)$$

In terms of $\Phi(x, y_0) = \phi(x, Y(x; y_0)) = -c\mathcal{Z} - C\mathcal{W}$, (2.39) becomes

$$\frac{\partial \mathcal{Z}}{\partial x}|_{y_0} = i\sigma \left[\left(1 + \frac{c}{C}\right) \mathcal{Z} + \frac{\Phi}{C} \right] \frac{\frac{\partial Y}{\partial y_0}|_x}{\frac{\partial \Psi}{\partial y_0}|_x}. \quad (2.42)$$

To obtain the solution, we integrate the equations (2.40) and (2.42). The actual numerical integration was performed in a stretched zonal coordinate, as described in the Appendix 2.11.

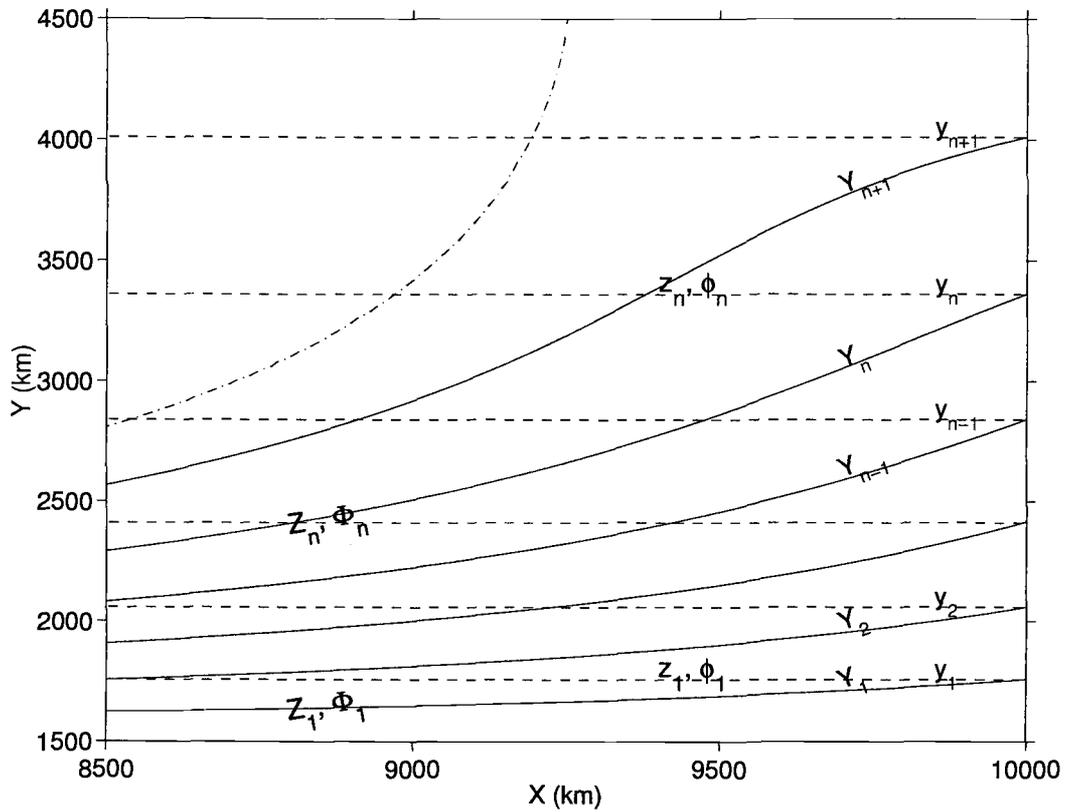


FIGURE 2.4. Detail of the two sets of the characteristics of the EZ solution: all the dependent variables defined on the set of the characteristics which are isopleths of the PV of the middle layer, denoted by $Y(x; y_0)$ (plotted as a solid line) are notated in upper case; all the dependent variables defined on the set of constant latitude lines characteristics denoted by $y(x; y_0)$ (plotted as a dashed line) are notated in lower case. Dot-dash line shows the boundary between the EZ and the WZ.

In the EZ, the westward Rossby wave speed exceeds the zonal geostrophic flow in the upper layer so that both sets of characteristics emanate from the eastern boundary and proceed westward without reversal. The middle layer PV (and its reciprocal Ψ) is conserved along the curved characteristics. The characteristics of the other equation, which comes from the Sverdrup equation, are east-west lines. The two sets of EZ characteristics cross one another. At any point in the EZ the solution depends on the initial data along the stretch of the eastern boundary bounded by the two characteristics that go through the solution point and on the forcing in the whole triangularly shaped region bounded by them.

2.4.2. In the WZ

Characteristics (geostrophic contours) of the perturbation equation for the time-dependent PV of the middle layer in WZ (2.25) are isopleths of depth $H_2^{(0)}$ of the base of the middle layer. Since in the northern half of the WZ the averaged zonal geostrophic flow exceeds the Rossby wave speed, these emanate from the western boundary in the north and loop around to rejoin it in the south, and the solutions of (2.25) are not determined by values of the independent variables along $x_B(y)$. Since our model does not include the western boundary current, it thus cannot give the boundary conditions needed for (2.25) and a further ad hoc assumption is required to specify its solutions. We make the simplest one, that the time dependent part of PV of the middle layer, given by $\frac{(H_2^{(1)} - H_1^{(1)})}{f}$, is constant, just like its steady counterpart. We set the constant to be equal to zero, assuming that the annual variability is too fast to disturb the homogeneity of the PV. With this assumption the solution of (2.25) is specified. This assumption also makes both interface displacements

continuous across the boundary between the zones as no information propagates into the WZ from the western boundary.

The other WZ equation is the same as equation (2.40) in the EZ. It is integrated westward along its characteristics, which are constant latitude lines to obtain ϕ (defined by 2.41). (The details of the calculation are given in the Appendix 2.11.) The initial conditions for the WZ calculation are obtained from the ϕ and \mathcal{Z} values along x_B which are the result of the EZ calculation.

Having chosen $\Gamma^{(1)} = 0$, time dependent interface displacements $H_1^{(1)} = z e^{-i\sigma t}$ and $H_2^{(1)} = w e^{-i\sigma t}$ are obtained from the resulting ϕ as $z = w = -\phi/(c + C)$. Sea surface elevation is obtained from (2.27).

2.5. Numerical integration

2.5.1. In the EZ

Two discrete sets of the characteristics are defined by $Y_j(x) = Y(x; y_{0j})$ and $y_j(x) = y(x; y_{0j}) \equiv y_{0j}$, $j = 1, 2, \dots, N$.

Equations (2.40), (2.41) and (2.42) are discretized and integrated westward using the fourth order Runge-Kutta procedure (Press et al., 1994) as though they were ordinary differential equations of the form

$$\frac{\partial Y_j}{\partial x} = \dot{Y}_j(x; y_{0j}; \mathcal{Z}_j, \Phi_j) \quad (2.43)$$

$$\frac{\partial \mathcal{Z}_j}{\partial x} = \dot{\mathcal{Z}}_j(x; y_{0j}; \mathcal{Z}_j, \Phi_j) \quad (2.44)$$

$$\frac{\partial \phi_j}{\partial x} = \dot{\phi}_j(x; y_{0j}; \mathcal{Z}_j, \Phi_j) \quad (2.45)$$

all for $j = 1, 2, \dots, N$. The discretized arrays of dependent variable obtained in the calculation are $\mathcal{Z}_j(x) = \mathcal{Z}(x; y_{0j})$, $\Phi_j(x) = \Phi(x; y_{0j})$, $\phi_j(x) = \phi(x; y_{0j})$ and

$z_j(x) = z(x; y_{0j})$ as shown schematically in Figure 2.4. (In the calculation it has been convenient to use a stretched coordinate, defined in the Appendix 2.11, as an independent variable instead of x .)

Each Runge-Kutta step advancing x furnishes Z_j on the Y_j characteristics and ϕ_j on the y_j characteristics. But \dot{Z} is a function of Φ as indicated in equation (2.44) and $\dot{\phi}$ is a function of z (equation 2.45). Therefore, after each Runge-Kutta step we need to obtain z_j on the y_j characteristics by interpolating the Z_j from the Y_j characteristics, and vice versa to obtain Φ_j from ϕ_j . In this way the integration is performed from the eastern boundary of the ocean to the western edge $x_B(y)$ of the EZ. Values of ϕ_j estimated along x_B provide the boundary conditions for the WZ integration. Finally, time-dependent upper layer interface displacement amplitude w is obtained from ϕ and z using the expression (2.41). Sea surface elevation is obtained from (2.27).

2.5.2. In the WZ

In the WZ, an equation of the same form as (2.40) is discretized and integrated westward from x_B as

$$\frac{\partial \phi_j}{\partial x} = \dot{\phi}_j(x; y_j; z_j, \phi_j) \quad (2.46)$$

using the Runge-Kutta procedure. The characteristics of this equation are (trivially) constant latitude lines, $y_j = y_{0j}$, $j = 1, 2, \dots, N$. (Westward integration was performed using a stretched variable ξ , defined in the Appendix 2.11.)

2.6. A preliminary numerical solution

Figure 2.5 shows a snapshot of the time-dependent middle layer interface displacement for a very weak steady background circulation, a case chosen to emphasize the two sets of waves of which the EZ time-dependent solution consists. In this calculation, as in all the subsequent numerical results shown in this paper, the time-dependent wind stress $\Theta^{(1)}$ (2.28) has the annual period, and the same spatial structure and amplitude as the steady wind stress; the choice for amplitude was motivated by fact that the seasonal variation in wind forcing is comparable to the amplitude of the mean wind forcing. The parameters used in this example are $W_0 = 7.7 \cdot 10^{-8} \text{ m s}^{-1}$, $H_{10}=300 \text{ m}$, $H_{20}=750 \text{ m}$, $\gamma_1 = 0.0245 \text{ m s}^{-2}$, $\gamma_2 = 0.01 \text{ m s}^{-2}$.

In the WZ the solution consists of one set of long baroclinic waves whose wavelengths are approximately 200 km - 500 km. Over most of the EZ, the time-dependent solution consists of two sets of baroclinic waves. Somewhat surprisingly one set of waves has scales much smaller than the scales of waves with no background flow. This result is similar to the one of Dewar (1998) who analysed propagation of planetary waves in the presence of vertically sheared flow by considering a three-layer flat-bottom basin QG system. He found that in the two-and-a-half layer limit the solution has small wavelength relative to the basin, with first-mode waves longer than the second one.

Since the goal of our work was to model westward propagating Rossby wave perturbations of the large scale circulation, we have made the so-called long wave approximation in neglecting the relative vorticity, so both sets of waves are in this sense "long" planetary waves. Because the wavelengths of the two sets of waves are so different, however, we will for convenience refer to them as the long and the short

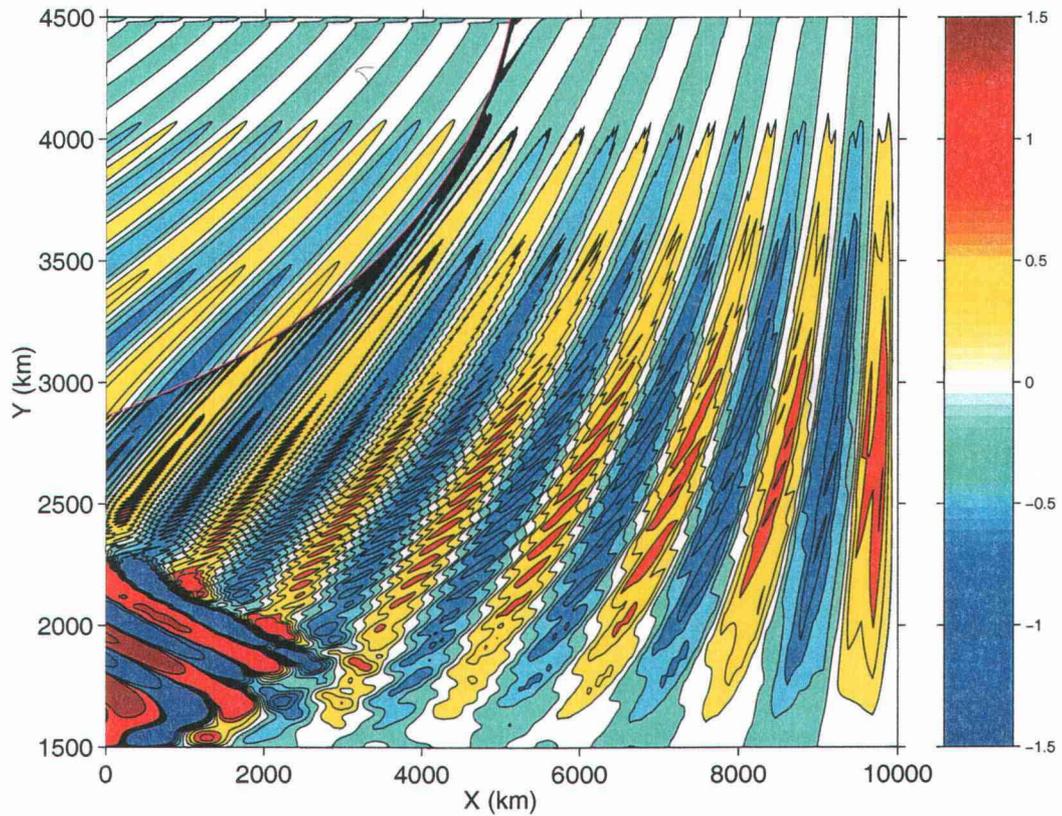


FIGURE 2.5. A snapshot of the time-dependent interface displacement of the base of the middle layer (m) obtained with very weak steady wind forcing of the background flow, $W_0=7.7 \cdot 10^{-8} \text{ m s}^{-1}$, a case which emphasizes the waves. The other parameters used in this example are $H_{10}=300 \text{ m}$, $H_{20}=750 \text{ m}$, $\gamma_1 = 0.0245 \text{ m s}^{-2}$ and $\gamma_2 = 0.01 \text{ m s}^{-2}$. The amplitude of the time-dependent Ekman pumping equals W_0 . The pink line is the boundary between EZ and WZ.

waves. But we must remember that the short wave field is not realistic when its wavelengths decrease below the baroclinic Rossby radius.

For the choice of parameters used in Figure 2.5 the wavelengths of the long waves range from approximately 2000 km in the southeastern part of the basin to 300 km in the northwestern part of the EZ. They are not greatly different in the WZ. The wavelengths of the short waves range from approximately 450 km in the southeast part of the basin to a fraction of a kilometer very close to the boundary x_B . The short waves cannot penetrate into the WZ.

The wavelengths of the short waves for parameters more representative of the steady circulation of the North Pacific subtropical gyre likewise range between approximately 500 km and a fraction of a kilometer near x_B . A scale estimate of the shorter baroclinic Rossby radius of deformation for this example is $\sqrt{\gamma_1 H_{10}}/f=38$ km. Short waves shorter than this scale are found in a band parallel to x_B and extending away from x_B about 1000 km equatorward and 2000 km eastward. Since only the long wave part of the solution enters the WZ, the WZ solution is not affected by the short waves not being modeled correctly all the way to the boundary x_B .

In the numerical calculation, the adaptive step size Runge-Kutta procedure provides almost unlimited zonal resolution, but the meridional resolution is limited by the meridional spacing of the characteristics. The short waves are therefore prone to being aliased, especially near x_B , where they are the shortest. Such aliasing is not obvious in case of weak background flow, but it will be seen to be present; when the background flow is more realistically strong (e.g. Figure 10 below) aliasing may become obvious and very troublesome.

In addition to the two sets of waves, a region of exponential growth in the southwestern corner of the subtropical gyre is a prominent feature of all our EZ

solutions. Gill et al. (1974) suggested on general grounds that this region may be prone to baroclinic instability. The nature of that instability for our particular model is explored more fully in a subsequent section of this paper. The possible aliasing of short waves however raises two questions regarding the numerical result shown in Figure 2.5. (i) How well are the stable waves modeled by the numerical scheme used? (ii) Is the growth observed in the southwestern part of the EZ a numerical artifact possibly related to such aliasing or is the flow there truly unstable?

We have worked out the approximate WKB free solutions (next section) to complete the identification of the small scale waves noted above with propagating Rossby waves and to see if the WKB solutions may be unstable. In order to see if the numerical scheme used to obtain the EZ solution gives rise to numerical instabilities, we have analysed that scheme by the classical methods of von Neumann stability analysis (Section 2.8.1, below). These WKB and von Neumann results are then applied to a particular numerical solution (Section 2.8.3).

2.7. WKB solution

2.7.1. In the EZ

The underlying WKB assumption is that the scale over which the mean flow changes is large compared to the wavelength of the waves (e.g. Bender and Orszag, 1978). We seek an approximate solution of the form $z \sim e^{iq(x,y)}$. The phase q is expanded as a series in powers of a small parameter ϵ where only the two highest terms are kept, so that $q = \frac{q_0 + \epsilon q_1}{\epsilon}$. Motivated by the plane wave solutions of (2.37) when all the coefficients are constant, we take ϵ to be of order σ^{-1} . In equation (2.37) the derivatives of z thus become

$$z_x = i \left(\frac{q_{0x}}{\epsilon} + q_{1x} \right) z \quad (2.47)$$

$$z_{xx} = \left[- \left(\frac{q_{0x}}{\epsilon} \right)^2 + i \frac{q_{0xx}}{\epsilon} - 2 \frac{q_{0x} q_{1x}}{\epsilon} + i q_{1xx} - (q_{1x})^2 \right] z \quad (2.48)$$

etc. The dispersion relation, obtained by collecting the largest terms of order ϵ^{-2} in the resulting equation, is

$$\sigma^2 - CUk^2 - CVkl + \sigma(c + C)k = 0. \quad (2.49)$$

Here $k = \frac{q_{0x}}{\epsilon}$, $l = \frac{q_{0y}}{\epsilon}$ are the zonal and the meridional component of the wavenumber vector.

In the southwestern part of the subtropical gyre, if U is sufficiently negative that

$$|U| > \left[\frac{\sqrt{C}Vl}{2\sigma} - \frac{(c + C)}{2\sqrt{C}} \right]^2, \quad (2.50)$$

the WKB solution shows westward spatial growth. The physical meaning of this apparent instability will be discussed later.

The rays, lines along which groups move according to WKB theory, denoted by $y(x) = \mathcal{Y}(x; \iota_0)$, where $\mathcal{Y}(a; \iota_0) = \iota_0$ is the initial position at the eastern boundary, have the slope

$$\frac{\partial \mathcal{Y}}{\partial x} = \frac{(\partial \Omega / \partial l)_{k,x,y}}{(\partial \Omega / \partial k)_{l,x,y}} = \frac{c_{gy}}{c_{gx}} = - \left(\frac{\partial k}{\partial l} \right)_{\sigma,x,y} = \frac{Vk}{\frac{\sigma^2}{Ck} + Uk}. \quad (2.51)$$

Here c_{gx} , c_{gy} are the x , y components of the group velocity. Let $\sigma = \Omega(k, l, x, y)$ denote the two solutions of the dispersion relation (2.49). There are two roots of (2.49) corresponding to a short and a long wave for given σ . Thus there are two sets of rays describing the group propagation of two types of baroclinic mode. Note that these rays are different from the characteristics given by (2.38), although the

rays corresponding to the short waves, like the characteristics for \mathcal{Z} defined above, do not cross x_B .

Collecting the terms of the next lower order in equation (2.37), substituted with (2.47) and (2.48), gives an equation for the evolution of amplitude along the rays,

$$\left. \frac{\partial q_1}{\partial x} \right|_{\iota_0} = \frac{i[CUk_x + CVk_y + (CU)_x k + (CV)_x l - \sigma C_x]}{\frac{\sigma^2}{k} + CUk}. \quad (2.52)$$

(Equation (2.49) has been used in the denominator.) Then the RHSs of (2.51) and (2.52) can be determined as follows, and these equations for the ray path and amplitude can be integrated westward.

Rays are started from the eastern boundary where the meridional component of the wavenumber vector l is specified. For given l the quadratic dispersion relation (2.49) is solved for one particular branch of k (long or short waves). By choosing a discrete set of initial points ι_{0j} , $j = 1, 2, \dots, N$, along the eastern boundary a discrete network of rays can be traced westward. Here the meridionally discretized equation (2.51) was integrated westward by a fourth-order numerical Runge-Kutta procedure.

Moving westward along the ray meridional wavenumber, l changes as follows. Differentiate the solution of the dispersion relation (2.49) denoted by $\Omega(k, l, x, y)$ with respect to y to get

$$0 = c_{g_x} k_y + c_{g_y} l_y + \left. \frac{\partial \Omega}{\partial y} \right|_{x, k, l}. \quad (2.53)$$

Since $q_{0xy} = k_y = l_x$, and equation (2.51) relates $\frac{c_{g_y}}{c_{g_x}}$ to the slope of the ray, it follows that

$$\left. \frac{\partial l}{\partial x} \right|_{\iota_0} = - \frac{1}{c_{g_x}} \left. \frac{\partial \Omega}{\partial y} \right|_x. \quad (2.54)$$

This equation can also be integrated westward with the Runge-Kutta procedure, furnishing the changing l . As l changes, k along the ray is obtained by solving the dispersion relation (2.49).

2.7.2. In the WZ

In the WZ the solution is obtained by solving equation (2.26). The approximate WKB solution to this equation, obtained in the same manner as in the EZ, yields a dispersion relation of the form

$$k = -\frac{\sigma f^2}{\beta(\gamma_1 H_1^{(0)} + \gamma_2 H_2^{(0)})}. \quad (2.55)$$

Analogous to the EZ calculation l is obtained from the ray equation (2.54). The terms of the next lower order give the amplitude equation along the rays (which are lines of constant latitude) as

$$\left. \frac{\partial q_1}{\partial x} \right|_y = \frac{i(\gamma_1 H_{1x}^{(0)} + \gamma_2 H_{2x}^{(0)})}{\gamma_1 H_1^{(0)} + \gamma_2 H_2^{(0)}}. \quad (2.56)$$

Values of z and w obtained along x_B from the EZ calculation provide the boundary conditions for the WZ calculation. The zonal wavenumber vector component tangential to x_B is continuous across the boundary and the zonal wavenumber vector component normal to x_B is discontinuous across the boundary between the zones.

2.7.3. An illustration of a WKB solution

Figure 2.6 shows a snapshot of the time-dependent displacement of the interface between the surface layer and the middle layer for the long wave WKB solution for the same set of parameters as in Figure 2.5. In the free WKB solutions the time-dependent wind forcing is neglected, so that the waves in Figure 2.6 are free annual

period long waves whose amplitudes are specified at the eastern boundary. For purposes of qualitative comparison with numerical solutions, the WKB calculation is started at the eastern boundary with $l = \frac{\pi}{\Delta}$, where Δ is the meridional extent of the basin. This corresponds approximately to the eastern boundary meridional scale of the idealized Ekman pumping (2.7) used in the numerical solutions. A similar short wave WKB solution (not shown) may be constructed. Comparison of the pattern of crests and troughs and their horizontal scales with the those of the numerical solution of Figure 2.5 suggests that the most of the variance of the signal in the numerical solution indeed consists of "free" waves, i.e. waves that are not generated locally by the time-dependent wind forcing.

The blank area in Figure 2.6 is the region where the WKB dispersion relation (2.49) indicates that the WKB solution becomes unstable. Since our WKB expressions for the ray paths hold only for the stable waves, only the part of the solution which consists of stable waves is shown in this figure, but it is important to note the general correspondence between the region where the numerical solution appears to be unstable and the regions where the WKB solutions are unstable.

2.8. Stability analysis of the numerical scheme used in the EZ

In order to examine the properties of the discrete numerical scheme used to obtain the EZ solution, an extension of the classical von Neumann analysis of the numerical scheme was performed. The comparison of the numerical dispersion relation to the analytical WKB dispersion relation provides a check for how well the numerical model represents the longer set of free long planetary waves in the EZ.

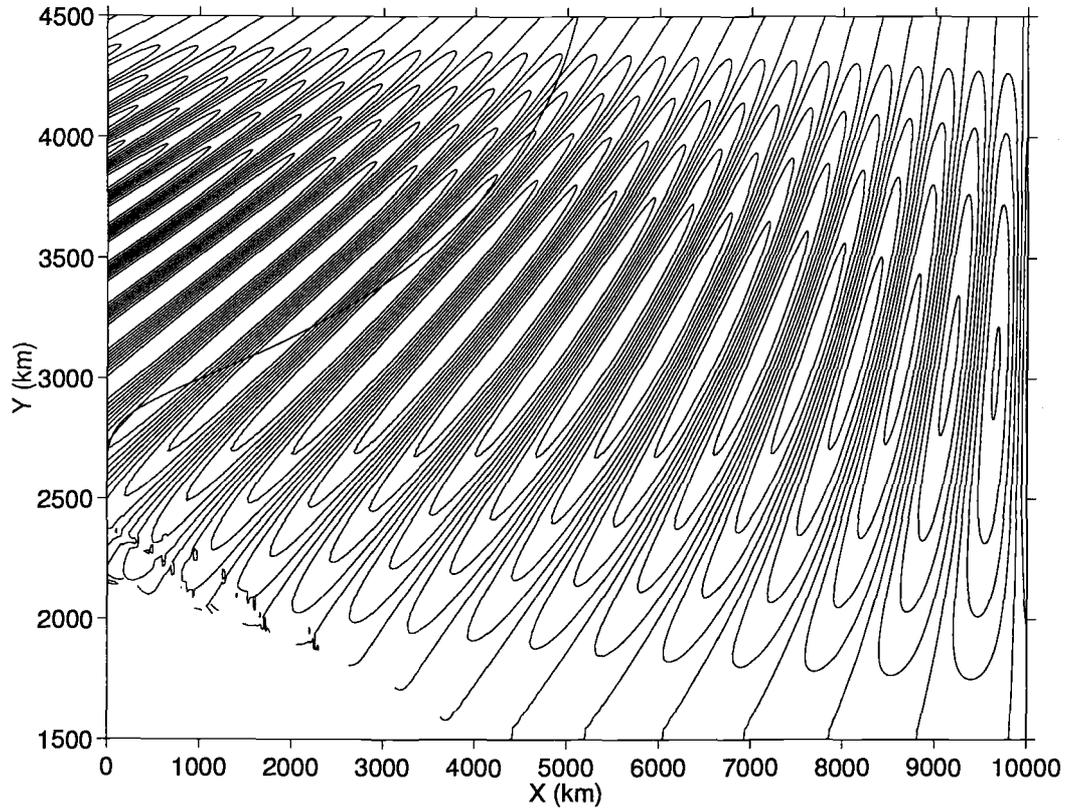


FIGURE 2.6. A snapshot of the annual period time-dependent middle layer interface displacement (m) for the long wave part of the WKB solution for the same set of parameters as in Figure 2.5. Contour levels are 0.25 m ranging from -1 m to 1 m. The solution shown in this figure does not include the slowly varying amplitude term e^{iq_1} . Since WKB solutions for the ray paths obtained in section 2.7 hold only for the stable waves, the region where the WKB solution becomes baroclinically unstable is left as a blank area.

The analysis was done using the χ, y coordinates (Appendix A), since those are the coordinates in which the numerical solution has been obtained. In this section, for simplicity, the analysis is outlined in x, y coordinates.

Numerical solutions in the EZ were obtained by integrating equations (2.40) and (2.42) to advance time dependent interface displacement of the middle layer, \mathcal{Z} , and a linear combination of the time dependent interface displacements of the upper and the middle layers, ϕ , defined by (2.41). The equations were integrated westward along the two sets of characteristics by means of the Runge-Kutta method. After each Runge-Kutta step, variables were interpolated from one set of the characteristics to the other. The two sets of the equations (2.40) and (2.42) can be schematically written as

$$\frac{d}{dx}\phi_n(x) = i\sigma z_n(x) \quad n = 1, 2, \dots, N \quad (2.57)$$

$$\frac{d}{dx}\mathcal{Z}_n(x) = i\sigma \left[\hat{d}\mathcal{Z}_n(x) + \hat{c}\Phi_n(x) \right] \quad n = 1, 2, \dots, N \quad (2.58)$$

where the time-dependent wind forcing has been neglected. The coefficients are assumed locally constant, so that each set of characteristics is a set of parallel straight lines, one zonal and one with slope V/U . For simplicity we assume that both sets of characteristics are meridionally equally spaced with the increment Δy . Comparison with the equations (2.40) and (2.42) shows that $\hat{d} = (c + C)/(CU)$ and $\hat{c} = 1/(CU)$, where the coefficients c, C, U are defined in section 2.4.

The two sets of characteristics divide the x, y plane into a periodically repeating pattern of identical cells. Figure 2.7 shows the cells. The dependent variables have correspondingly been discretized in the y direction so that the subscripts n index the characteristics.

Consider one of the cells (index j) closest to the eastern boundary of the basin. The x coordinates of the southeast and southwest corners of the j^{th} cell are

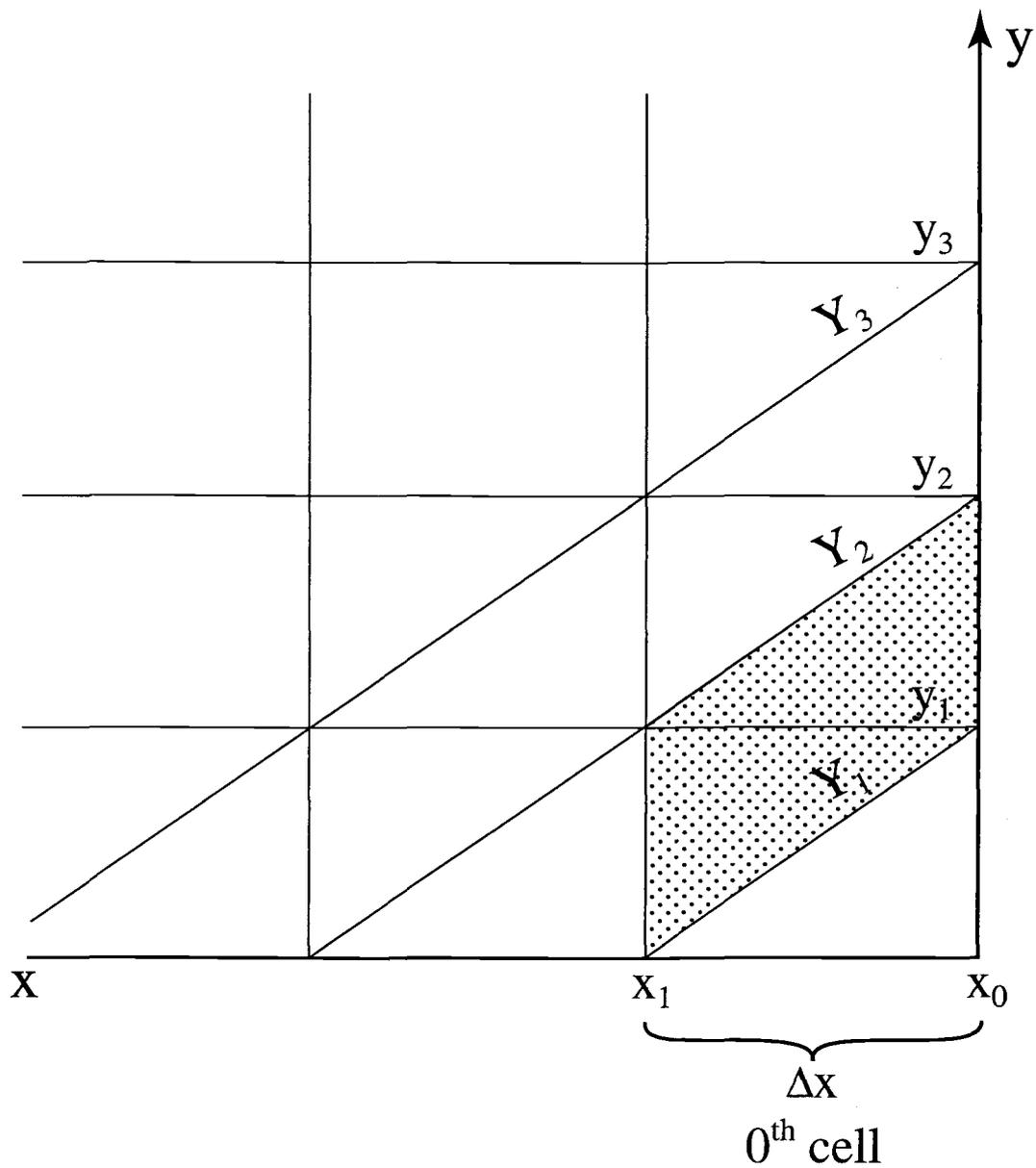


FIGURE 2.7. Two sets of characteristics which are parallel straight lines divide x,y plane into periodically repeating pattern of identical cells on which von Neumann type of analysis was performed.

denoted by x_j and x_{j+1} . (Though x increases towards the east, the indices of the cells increase westward, since that is the sense of integration.) The width of the cell is given by $\Delta x = x_j - x_{j+1}$.

After each Runge-Kutta step, \mathcal{Z} is obtained on the set of Y characteristics (tilted lines) and ϕ on the set of y characteristics (constant latitude lines). For equation (2.57) lower case z 's may be obtained by interpolating the upper-case \mathcal{Z} 's, while for equation (2.58) upper-case Φ 's may be interpolated among the lower-case ϕ 's. For simplicity consider interpolation which is linear among nearest neighbors:

$$z_n = (1 - \tau_j)\mathcal{Z}_{n+j} + \tau_j\mathcal{Z}_{n+j+1} \quad (2.59)$$

$$\Phi_n = (1 - \tau_j)\phi_{n-j} + \tau_j\phi_{n-j-1} \quad (2.60)$$

in the j^{th} cell ($x_{j+1} \leq x \leq x_j$) and $\tau_j = \frac{(x_j-x)}{\Delta x}$.

We look for solutions trigonometric in y and Y . Consider one Fourier constituent at a time:

$$\mathcal{Z}_n = \mathcal{Z}(x, l)e^{iln\Delta y}, \quad \phi_n = \phi(x, l)e^{iln\Delta y}. \quad (2.61)$$

Use the interpolation formulae (2.59) and (2.60) to write the system defined by (2.57) and (2.58) in terms of \mathcal{Z} and ϕ only, to obtain

$$\frac{d}{dx}\phi(x, l) = i\sigma f(\tau_j)\tilde{\mathcal{Z}}^{(j)}(x, l) \quad (2.62)$$

$$\frac{d}{dx}\tilde{\mathcal{Z}}^{(j)}(x, l) = i\sigma \left[\hat{d}\tilde{\mathcal{Z}}^{(j)}(x, l) + \hat{c}f^*(\tau_j)\phi(x, l) \right]. \quad (2.63)$$

Here $f(\tau) = (1 - \tau) + \tau e^{il\Delta y}$ denotes the interpolation function, $f^*(\tau)$ its complex conjugate and

$$\tilde{\mathcal{Z}}^{(j)}(x, l) = \mathcal{Z}(x, l)e^{ilj\Delta y}. \quad (2.64)$$

The phase shift introduced by (2.64) represents the j -fold shift of n th Y characteristic from the n th y -characteristic in the j th cell. The width of the cell is normalised by using the variable τ_j , so that $d\tau_j = -dx/\Delta x$. Dropping the j labels, (2.62), (2.63) become

$$\frac{1}{\Delta x} \frac{d}{d\tau} \begin{pmatrix} \phi \\ \tilde{Z} \end{pmatrix} = -i\sigma \begin{pmatrix} 0 & f \\ \hat{c}f^* & \hat{d} \end{pmatrix} \begin{pmatrix} \phi \\ \tilde{Z} \end{pmatrix}, \quad 0 \leq \tau \leq 1. \quad (2.65)$$

2.8.1. Numerical dispersion relation

Concentrating on one cell at a time, we seek to express the solution at the western edge of the cell, $\tau = 1$, as a function of the solution at the eastern edge of the cell, $\tau = 0$. Because the coordinate $\tau_j = \frac{(x_j - x)}{\Delta x}$ appears in the elements of the right hand side of (2.65), the integration across a cell must be performed numerically. The amplitudes $\phi(\tau)$ and $\tilde{Z}(\tau)$ are expressed as a linear combination of two functions

$$\phi(\tau) = A\varphi_1(\tau) + B\varphi_2(\tau) \quad (2.66)$$

$$\tilde{Z}(\tau) = A\zeta_1(\tau) + B\zeta_2(\tau). \quad (2.67)$$

The boundary conditions at the eastern edge of the cell are $\zeta_1(0) = 1$, $\zeta_2(0) = 0$, $\varphi_1(0) = 0$, $\varphi_2(0) = 1$. This defines two systems to be solved

$$\frac{1}{\Delta x} \frac{d}{d\tau} \begin{pmatrix} \varphi_m \\ \zeta_m \end{pmatrix} = -i\sigma \begin{pmatrix} 0 & f \\ \hat{c}f^* & \hat{d} \end{pmatrix} \begin{pmatrix} \varphi_m \\ \zeta_m \end{pmatrix} \quad m = 1, 2. \quad (2.68)$$

The systems are solved numerically by integrating westward across a cell using the Runge-Kutta procedure, which provides the values of $\zeta_1(1)$, $\varphi_1(1)$, $\zeta_2(1)$ and $\varphi_2(1)$ at the western edge of the cell. These are the elements of the propagation matrix that relates the solution at the western edge of the cell to the solution at the eastern edge of the cell

$$\begin{pmatrix} \phi \\ \tilde{\mathcal{Z}} \end{pmatrix} \Big|_{\tau=1} = \begin{pmatrix} \varphi_1(1) & \varphi_2(1) \\ \zeta_1(1) & \zeta_2(1) \end{pmatrix} \begin{pmatrix} \phi \\ \tilde{\mathcal{Z}} \end{pmatrix} \Big|_{\tau=0}. \quad (2.69)$$

When the boundary between the two cells is crossed, both ϕ and $\tilde{\mathcal{Z}}$ must be continuous. The latter requirement applied to (2.64) shows that

$$\tilde{\mathcal{Z}}^{(j+1)} = \tilde{\mathcal{Z}}^{(j)} e^{il\Delta y} \quad (2.70)$$

at $\tau_j = 1 + \epsilon$, $\tau_{j+1} = 0 - \epsilon$ which combined with (2.69) gives the matching condition

$$\begin{pmatrix} \phi \\ \tilde{\mathcal{Z}} \end{pmatrix} \Big|_{\tau_{j+1}=0-\epsilon} = \begin{pmatrix} 1 & 0 \\ 0 & e^{il\Delta y} \end{pmatrix} \begin{pmatrix} \phi \\ \tilde{\mathcal{Z}} \end{pmatrix} \Big|_{\tau_j=1+\epsilon}. \quad (2.71)$$

The propagation matrix, relating the solution an infinitesimal distance to the west of one edge of the cell to the solution an infinitesimal distance to the west of the edge of the next cell, is then obtained by combining (2.69) and (2.71) to give

$$\begin{pmatrix} \phi \\ \mathcal{Z} \end{pmatrix} \Big|_{\tau=1-\epsilon} = \begin{pmatrix} \varphi_1(1) & \varphi_2(1) \\ e^{il\Delta y} \zeta_1(1) & e^{il\Delta y} \zeta_2(1) \end{pmatrix} \begin{pmatrix} \phi \\ \mathcal{Z} \end{pmatrix} \Big|_{\tau=0-\epsilon}. \quad (2.72)$$

Diagonalize the propagation matrix in (2.72) to obtain

$$N = H \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix} H^{-1}, \quad (2.73)$$

where K_1, K_2 are its eigenvalues. Integration over each cell corresponds to multiplying by the matrix N one time. Eigenvalues of the propagation matrix can be written as $K_m = e^{i\overline{k}_m \Delta x}$, $m=1,2$, where \overline{k}_m is the zonal component of the wavenumber vector averaged over the width Δx of a cell. From the eigenvalues K_m , $m=1,2$, of the propagation matrix zonal wavenumbers \overline{k}_m are estimated as

$$\overline{k}_m = \frac{\ln(K_m)}{i\Delta x} + \frac{2n\pi}{\Delta x} \quad n = 0, 1, 2, \dots \quad (2.74)$$

so that the numerical dispersion relation consists of many branches; as $\Delta x \rightarrow 0$ the $n = 0$ branch reduces to the WKB dispersion relation (2.49). This procedure is

repeated for a range of specified real values of the meridional wavenumber vector component l .

Figure 2.8 shows examples of the numerical dispersion relation so obtained (with cubic interpolation) and the corresponding analytical WKB dispersion relation obtained by solving equation (2.49). The dispersion relation plots are shown in the wavenumber coordinates corresponding to the χ, y coordinates in which the numerical calculations have been performed. Both analytical and numerical dispersion relation plots are obtained by fixing frequency σ , varying the meridional wavenumber component l (real) and solving the resulting quadratic equation for the zonal wavenumber k (complex in general).

The figure encompasses the range of the wavenumbers occupied by the lowest branch of the numerical dispersion relation (2.74): $|k| < \frac{\pi}{\Delta x}$, $|l| < \frac{\pi}{\Delta y}$. These limits are set by the meridional spacing and the slopes of the characteristics. Beyond these bounds, the numerical dispersion relation would be periodic in l and k . The meridional resolution is indeed limited by the meridional spacing of the characteristics. The zonal resolution is however not restricted to the lowest branch of the numerical dispersion relation, but may be much finer because the integration along the characteristics was performed by using an adaptive-step Runge-Kutta method. This enables us to successfully resolve waves whose zonal wavenumbers lie outside the zonal limits of the numerical box characterizing the lowest branch of the dispersion relation. In the representation of Figure 2.8 the higher zonal aliases, though apparently resolved, are folded back into the range $|k| < \frac{\pi}{\Delta x}$. This is important because when a branch from a higher alias intersects a lower alias a kind of instability may occur, the origin of which is purely due to periodicity of the cells formed by the intersecting characteristics.

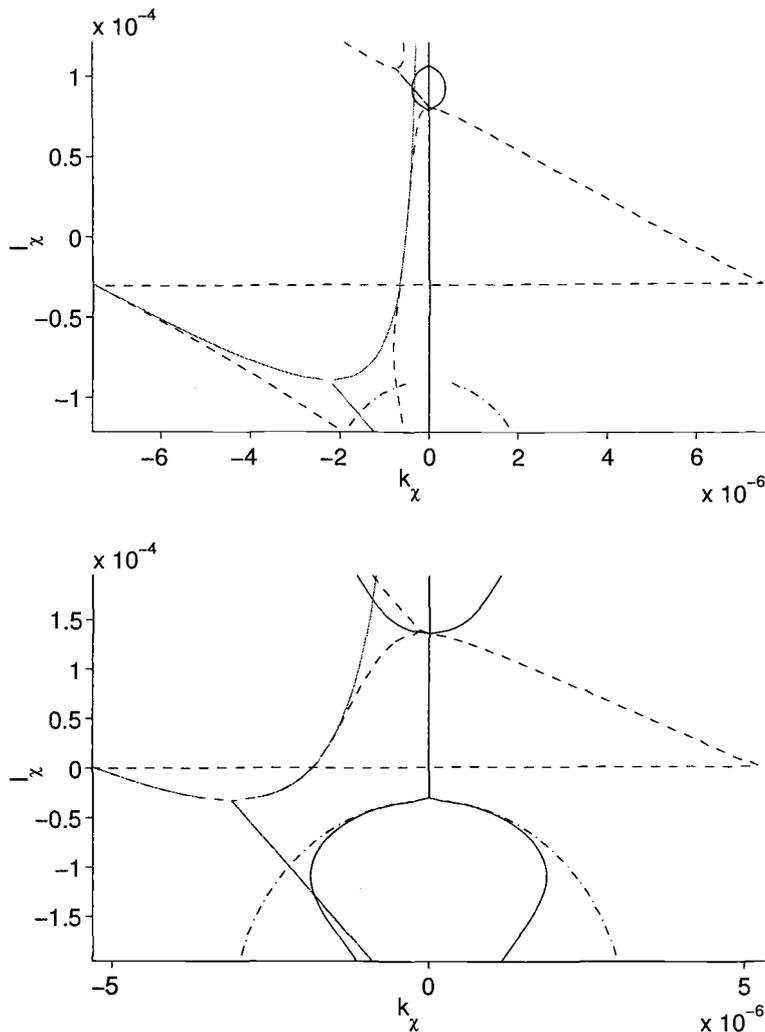


FIGURE 2.8. Numerical dispersion relation (2.74) and the corresponding local analytical WKB dispersion relation (2.49) in χ, y coordinates (units: radians/km) for the parameters of Figure 2.5. We fix the frequency σ , vary the meridional wavenumber l (real) and solve the quadratic dispersion relation for the zonal wavenumbers k (complex in general). The real part of k is plotted as a light solid line for the analytical dispersion relation and as a dashed line for the numerical dispersion relation. The imaginary part of k , representing the spatial growth, is plotted as a dash-dot line for the analytical dispersion relation and as a dark solid line for the numerical dispersion relation. The figure encompasses the range of the wavenumbers occupied by the lowest branch of the numerical dispersion relation (2.74): $|k| < \frac{\pi}{\Delta x}$, $|l| < \frac{\pi}{\Delta y}$. In both panels numerical instability appears for $l > 0$ as a consequence of discretization and the resulting aliasing: on account of finite meridional resolution the baroclinically unstable region for $l < 0$ gets partly aliased in the numerical dispersion relation from the $-\frac{l}{\Delta y}$ limit to $\frac{l}{\Delta y}$ limit. The spurious numerical instability occurs in the vicinity of the point in the wavenumber space where the aliased short wave branch crosses the long wave branch. Both plots are for the numerical solution shown in Figure 2.11, the upper panel is for a point $x = 6700$ km, $y = 3050$ km and the lower panel is for $x = 2000$ km, $y = 2000$ km.

Figure 2.8 illustrates the important features of the numerical dispersion relation: (i) quite good agreement with the analytical dispersion relation for k , l not too large, (ii) preservation of the instability which is present in the analytical dispersion relation on longer wavelengths that are well resolved by numerics (typically for $l < 0$) and (iii) the possible occurrence of a new, numerical instability. To suppress this last, it will be necessary to introduce some degree of averaging across the characteristics.

Figure 2.8 shows that the numerical instability (iii) is a consequence of discretization and the resulting aliasing. In the upper panel the baroclinically unstable region present for $l < 0$ in the analytical dispersion relation is so shrunk in the numerical dispersion relation by discretization error that only its alias into $l > 0$ wavenumbers on which there is no corresponding physical instability in the analytical dispersion relation survives. In the example shown in the lower panel the analytically unstable region that occurs for $l < 0$ in the analytic dispersion relation wraps around in y in such a way that it partly remains in $l < 0$ close to the $-\frac{l}{\Delta y}$ limit and it partly gets aliased into $l > 0$ close to the $\frac{l}{\Delta y}$ limit.

In both cases the alias into $l > 0$ results in numerical instability of waves that are completely stable on the analytical dispersion relation. In both cases the spurious numerical instability occurs in the vicinity of the point in wavenumber space where the extension of the zonally aliased short wave branch would cross the long wave branch. This is very similar to the way in which in which the long and short wave branches of the analytical dispersion relation would cross for $l < 0$ in the absence of the analytically unstable region in $l < 0$. The numerical instability can be suppressed by averaging over adjacent characteristics.

2.8.2. Scale selective damping: 3 point averaging

In order to suppress the numerical instabilities illustrated in Figure 2.8, averaging over adjacent characteristics has been introduced, which acts as scale selective damping. Averaging over three adjacent characteristics replaces equation (2.61) by

$$\bar{\phi}_n = s_w \phi_{n-1} + c_w \phi_n + s_w \phi_{n+1} \quad (2.75)$$

$$\bar{\mathcal{Z}}_n = s_w \mathcal{Z}_{n-1} + c_w \mathcal{Z}_n + s_w \mathcal{Z}_{n+1} \quad (2.76)$$

where c_w is the weight assigned to the central value of the variable that has been averaged and s_w are the weights of the two sidepoints used in averaging. For the averaging not to affect the largest scales, the weights must satisfy $c_w + 2s_w = 1$. Averaging the variables on the rhs of the equations (2.57) and (2.58) over three adjacent characteristics introduces scale selective damping. In the j^{th} cell the Fourier transformed equations (2.62) and (2.63) correspondingly become

$$\frac{d}{dx} \phi(x, l) = i\sigma f(x) \tilde{\mathcal{Z}}^{(j)}(x, l) \mathcal{A} \quad (2.77)$$

$$\frac{d}{dx} \tilde{\mathcal{Z}}^{(j)}(x, l) = i\sigma \left[\hat{d} \tilde{\mathcal{Z}}^{(j)}(x, l) + \hat{c} f^*(x) \phi(x, l) \right] \mathcal{A} \quad (2.78)$$

where $\mathcal{A} = \left[2s_w \cos(l\Delta y) + c_w \right]$ introduces the scale selective damping and $\tilde{\mathcal{Z}}^{(j)}(x, l)$ is given by (2.64).

Although the averaging is introduced for numerical considerations, it can be given a physical interpretation and interpreted as the equivalent meridional diffusivity \mathcal{K} . For suppose the RHS of equations (2.35) and (2.36) contain meridional diffusion terms of the form $\mathcal{K} \frac{d^2(z-w)}{dy^2}$ and $\mathcal{K} \frac{d^2 z}{dy^2}$. For illustration consider the latter (the simpler) one; it would appear in the equation (2.77) as

$$\mathcal{K} \tilde{\mathcal{Z}}^{(j)}(x, l) \frac{e^{il(n+1)\Delta y} - 2e^{iln\Delta y} + e^{il(n-1)\Delta y}}{\Delta y^2}. \quad (2.79)$$

For small values of Δy , \mathcal{A} can be approximated by the expression $1 - s_w(l\Delta y)^2$. If we assume periodicity in y , the operator $-\mathcal{K} \frac{d^2}{dy^2}$ can be expressed as $\frac{4\mathcal{K}}{\Delta y^2} \sin^2(\frac{l\Delta y}{2})$. Comparing the two expressions shows that performing the three point average in the limit of small Δy corresponds to adding the term $\sigma s_w(l\Delta y)^2 \mathcal{Z}^{(j)}$ to the original unaveraged equation (2.62). Equating the two expressions for small values of (Δy) gives; $\frac{4\mathcal{K}}{\Delta y^2} \sin^2(\frac{l\Delta y}{2}) = \sigma s_w(l\Delta y)^2$, which implies $\mathcal{K} = \sigma s_w(\Delta y)^2$.

Figure 2.9 shows the same examples shown in Figure 2.8, but here the numerical dispersion relation has been obtained by averaging over three adjacent characteristics with the central weight $c_w=0.4$. With an average $\Delta y=15$ km, this corresponds to introducing numerical diffusion with $\mathcal{K} =14$ m² s⁻¹. Compared to commonly accepted values of diffusivity due to mesoscale geostrophic turbulence this is a miniscule value. Averaging has entirely suppressed the spurious numerical instability in both the examples shown in Figure 2.9. Although the averaged numerical dispersion relation does not agree so well with the analytical dispersion relation as its unaveraged counterpart does, the two still remain quite close for the wavelengths of the longer set of planetary waves upon which this work focuses.

2.8.3. Application of the stability analysis of the numerical scheme used in the EZ

The numerical solutions display three kinds of numerical instability resulting in westward growing waves, as well as westward growth that reflects true instability of the analytical solution. The important conclusion of this section is that the place in the basin at which the solution becomes unstable depends on the numerical resolution in two ways: 1) finer meridional resolution enables us to resolve short waves further towards the west, but does not necessarily guarantee more accurate solution because 2) changing the resolution irregularly changes the way in which the numeri-

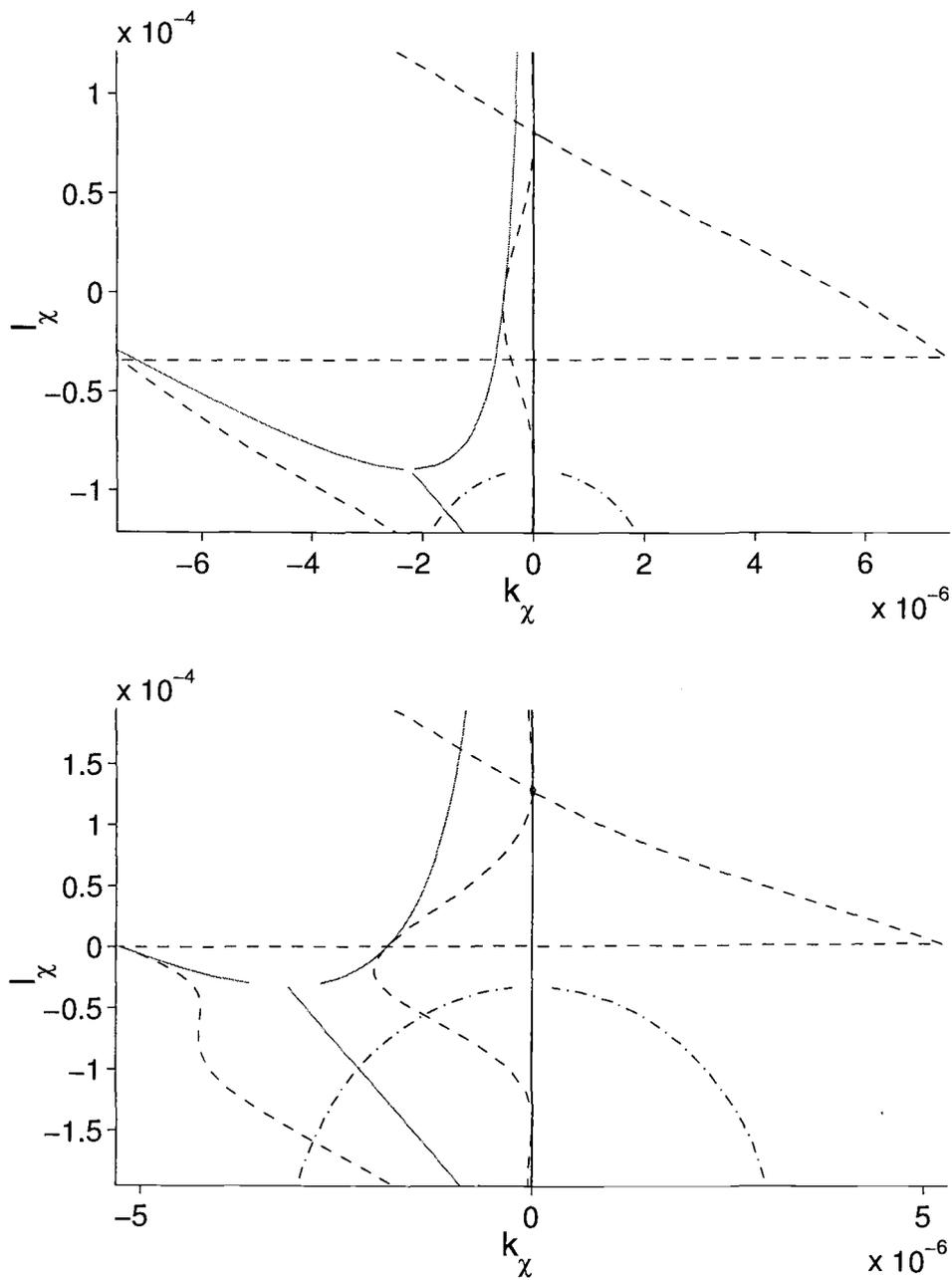


FIGURE 2.9. .

Analytical and averaged numerical dispersion relation for the case shown in Figure 2.8. The latter has been obtained by using the three point average with the central weight 0.4, which has suppressed the spurious numerical instability in both examples.

cal solution aliases waves with small meridional wavelengths. That aliasing depends critically on the particular meridional wavenumber for which finite meridional numerical resolution coupled with the slope of one set of the characteristics results in wrap around, from $k=-\frac{\pi}{\Delta x}$ to $k=\frac{\pi}{\Delta x}$, of the short wave branch of the dispersion relation ($\Delta x = \Delta y/\text{slope}$).

As is further discussed in Section 2.9, the true instability may be excited locally by wind forcing and/or by refraction of waves incident from the east into the unstable region of the dispersion relation. In this section we first check how accurately the numerical model represents the longer set of stable waves and then illustrate the three different ways in which numerical instability may occur:

(1) The lowest branch of the dispersion relation folds over in k when k reaches $|\frac{\pi}{\Delta x}|$, there is accompanying crossing of long and aliased short branches of the dispersion relation, and resulting instability over a limited range of $l > 0$. This possibility is further discussed in Section 2.8.3.2, below.

(2) The lowest branch of the dispersion relation again folds over in k when k reaches $|\frac{\pi}{\Delta x}|$, there is again crossing of long and aliased short branches of the dispersion relation, and resulting instability over a range of l which now however extends up to $\frac{\pi}{\Delta y}$ because the analytically unstable region for $l < 0$ extends to the lower limit $l = -\frac{\pi}{\Delta y}$ of resolution and beyond so that it is partly aliased into $l > 0$. This possibility is further discussed in Section 2.8.3.3, below.

(3) The lowest branch of the dispersion relation does not fold over in k but nonetheless the meridional wavelength gets so short that waves with $l > 0$ get aliased into $l < 0$, and particularly into the analytically unstable region that occupies $l < 0$ (or into what remains of that region in the numerical dispersion relation). In this case, waves with poleward propagating phases approach a latitude at which they

begin to be aliased into equatorward propagating unstable waves with $l < 0$. This possibility is further discussed in Section 2.8.3.5, below.

2.8.3.1. Resolution of long waves in the numerical solution

To check how accurately the numerical model represents the longer set of stable waves, we have graphically estimated the wavelengths from the numerical solution shown in Figure 2.12 at the location A and plotted them on the corresponding unaveraged and averaged local dispersion relation diagrams in Figure 2.10. (This is a very crude method of estimating the wavelengths especially since the WKB maps of the wavelengths show that the wavelengths of the long set of waves change appreciably on a scale of the order of the long wave wavelength). The upper panel shows the unaveraged numerical dispersion relation and the lower panel shows the averaged numerical dispersion relation ($c_w=0.4$) for the point A. The zonal (λ_z) and meridional (λ_m) wavelengths of the long wave part of the solution are $\lambda_z=920$ km and $\lambda_m=2180$ km respectively. The star indicates the wavelength estimate which falls very closely on the corresponding stable wave branch of the unforced solution. The long wave branch of averaged numerical dispersion relation agrees well with the analytical dispersion relation at the wavelengths associated with the long waves. Similar wavelength estimates for the long wave part of the numerical solution continue to fall very closely on the long wave branch of the dispersion relation as far westward in the basin as one can follow the long waves. We thus conclude that the the longer set of the stable planetary waves has been modeled accurately.

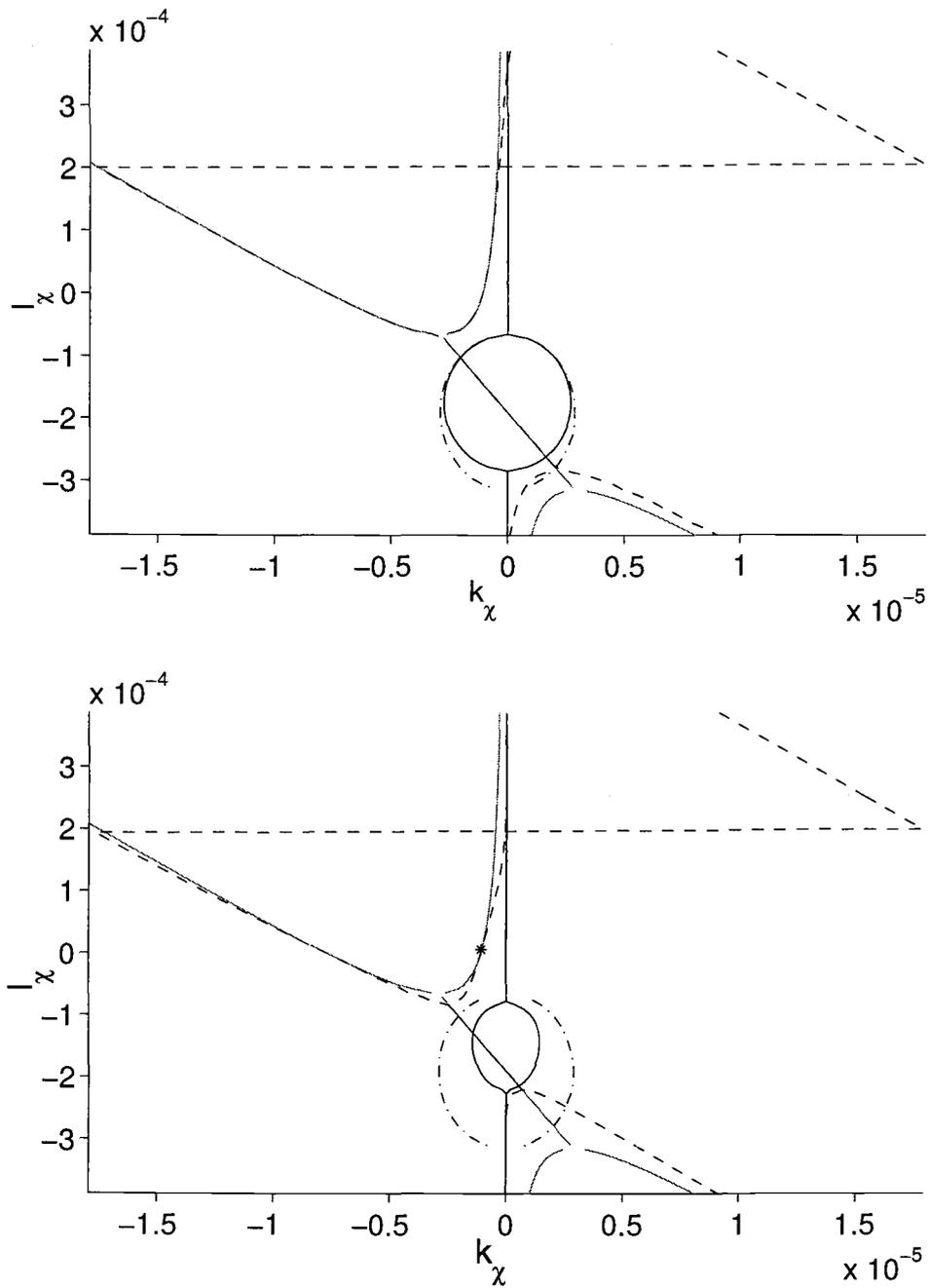


FIGURE 2.10. Local analytical and numerical dispersion relation for a point A shown in Figure 2.12. The upper panel shows the analytical and numerical dispersion relation for the unaveraged solution shown in Figure 2.11 and the lower panel shows the analytical and numerical dispersion relations for the averaged numerical solution shown in Figure 2.12. The star indicates the wavelength estimate of the long wave part of the solution at $x = 9000$ km, $y = 3500$ km: $\lambda_z = 920$ km, $\lambda_m = 2180$ km.

2.8.3.2. Numerical instability with no analytical counterpart

Figure 2.11 shows a snapshot of the time-dependent interface displacement of the middle layer for the same set of parameters as in Figure 2.5, but with the amplitude of Ekman pumping increased to $W_0=1.23 \cdot 10^{-7} \text{ m s}^{-1}$. This example has been chosen since it clearly illustrates how numerical instability may be generated. Westward exponentially growing small scale features occupy a 1000 km wide region parallel to x_B between $2000 \text{ km} < x < 8000 \text{ km}$. On the dispersion relation of Figure 2.8 (upper panel) this feature corresponds to excitation of meridional wavenumbers for which alias induced folding of the dispersion relation in k causes the two branches of the dispersion relation to cross with instability resulting as discussed in Section 2.8.1.

2.8.3.3. Numerical instability due to resolution induced meridional wrap-around in the wavenumber plane

In the western part of the EZ (in the region of Figure 2.11 occupying approximately $x < 3000 \text{ km}$, $y > 2000 \text{ km}$) the short waves are clearly visible as small scale oscillations in the contour lines of the exponentially growing solution. At this location their meridional wavelength estimated as from the Figure 2.11 equals exactly twice the meridional spacing of the characteristics (approximately 20 km) indicating that the short waves are barely resolved in the meridional direction. In the numerical dispersion relation for this location (Figure 2.8, lower panel), differently from that of Figure 2.8, upper panel, the analytically unstable region in $l < 0$ gets wrapped around in y in such a way that it partly remains in $l < 0$ and it partly gets aliased into $l > 0$. For $l > 0$ waves with the meridional wavelengths shorter than 45 km thus become numerically unstable resulting in westward exponential growth.

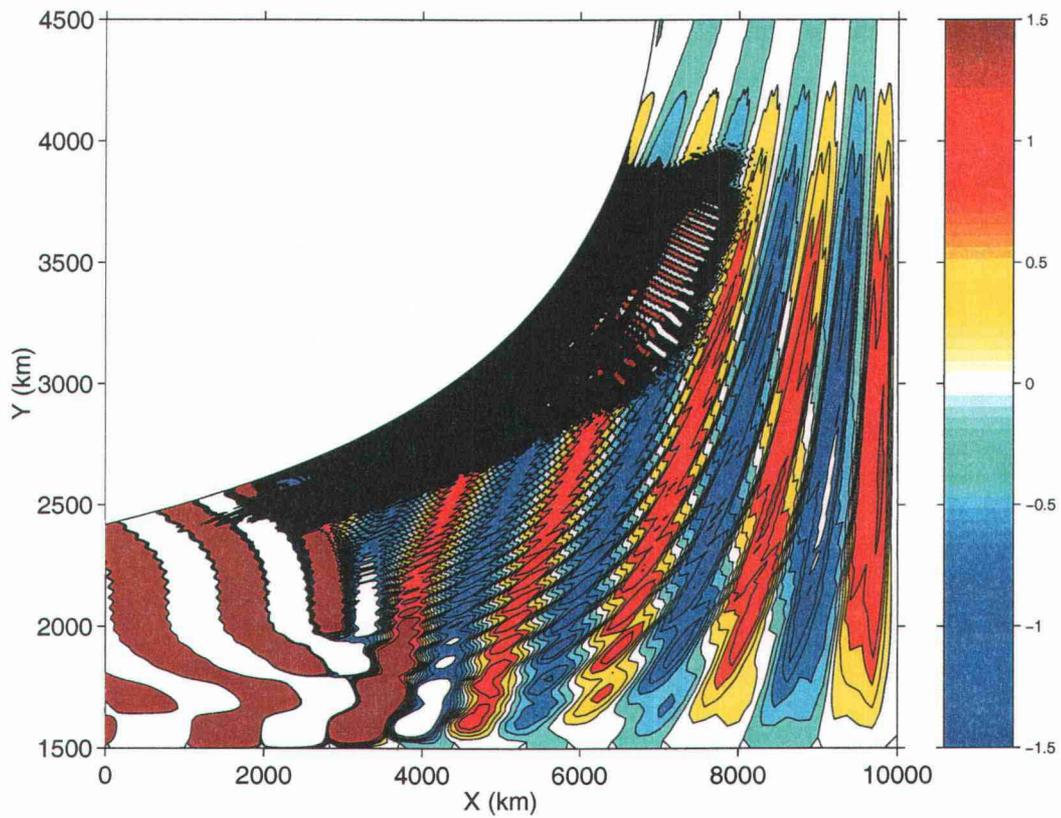


FIGURE 2.11. A snapshot of the time-dependent interface displacement of the base of the middle layer (m) for the same set of parameters as in Figure 2.5 with the amplitude of Ekman pumping increased to $W_0=1.23 \cdot 10^{-7} \text{ m s}^{-1}$.

2.8.3.4. Effects of averaging

Numerical instability can be suppressed by averaging over adjacent characteristics. Figure 2.12 shows the averaged numerical solution (with central weight $c_w=0.4$) corresponding to the unaveraged solution shown in Figure 2.11. Averaging over three adjacent characteristics has completely removed exponentially growing small scale features formerly present in the middle of the EZ (Section 2.8.3.2, above) as well as the exponential growth formerly present in the region occupying approximately $x < 3000$ km, $y > 2000$ km (Section 2.8.3.3, above).

The averaging has little effect in those areas of the gyre that were not occupied by numerical instabilities before the averaging was carried out. The exponential growth of the solution in the far southernmost part of the basin (the region occupying approximately $x < 4000$ km, $y < 1600$ km in Figure 2.11) has however not been reduced much by averaging. The waves here have long meridional wavelengths (the wavefronts are almost meridional) so averaging across characteristics, which acts as scale selective damping, does not affect them significantly. In this part of the basin the unstable region of the numerical dispersion relation (not shown) agrees well with the analytically unstable dispersion relation for smaller absolute values of l although averaging suppresses the growth closer to the resolution limit for $l < 0$. We later further discuss (Section 2.9.1.2.) why we believe that the true solution is unstable in this region.

In the southwestern corner of the EZ (in the region extending approximately $x < 4000$ km, 2000 km $< y < 2150$ km in Figure 2.12) wavefronts bend slightly in the form of a letter "S" as if the waves around $y = 2100$ km traveled faster than waves around $y = 2000$ km. This is an artifact caused by the underlying properties of the

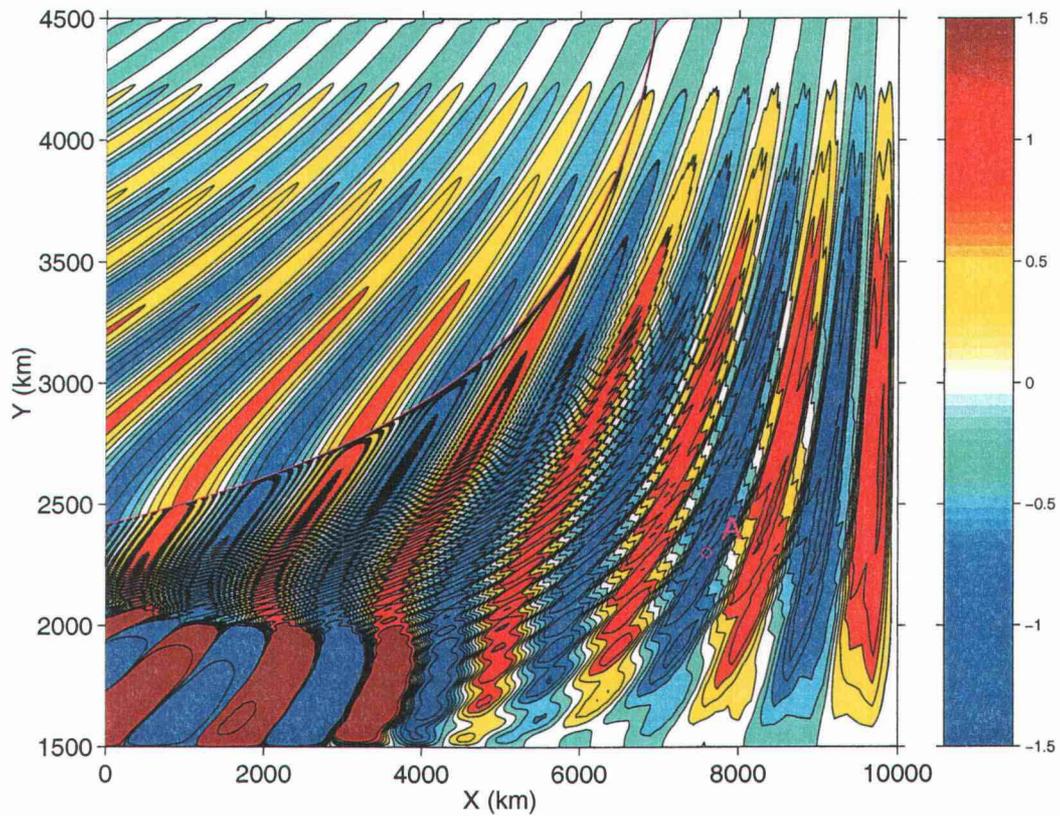


FIGURE 2.12. A snapshot of the time-dependent interface displacement of the base of the middle layer (m) for the same set of parameters as in Figure 2.11. The numerical solution shown in this figure is obtained by averaging over three adjacent characteristics with the central weight 0.4 which introduces scale selective damping suppressing the numerical instabilities.

aliased short waves present at this location in the unaveraged solution (Figure 2.11) which averaging with the central weight $c_w=0.4$ has not removed completely.

2.8.3.5. Numerical instability induced by northward propagating short waves being aliased into southward propagating unstable waves

Both numerical and WKB solutions show that both sets of waves refract so that the direction of phase propagation is towards the northwest, and the wavelengths become quite short; this means that in terms of a local plane wave $\exp(-i\sigma t + ikx + ily)$ the meridional wavenumber vector component for the short wave may become large and positive. If $\frac{dZ}{dy}$ in equation (2.37) were replaced by its finite difference equivalent $\frac{Z(y+\Delta y)-Z(y-\Delta y)}{2\Delta y}$, then the dispersion relation (2.49) would become

$$CUk^2 + CVk \frac{\sin(l\Delta y)}{\Delta y} - \sigma(c + C)k - \sigma^2 = 0. \quad (2.80)$$

The analytic dispersion relation, given by (2.49) shows that sufficiently negative values of l are always unstable. The important point is that aliasing of positive values of l greater than $\frac{\pi}{\Delta y}$ changes the sign of the term that involves l , potentially causing the aliased solutions to be unstable.

We believe that this is the process that dominates the behavior of the short wave in the southwest corner of the subtropical gyre shown in Figure 2.5, and results in aliased short wave solutions that grow westwards and appear to propagate towards the southwest (occupying approximately the region $x < 2000$ km, $y > 1600$ km).

To further investigate this possibility, we additionally obtained numerical solutions that primarily consist of either the long or the short wave part of the solution. These numerical solutions are generated at the eastern boundary by a

wavemaker artificially designed to give primarily either long or short waves. Time-dependent wind forcing is omitted.

Boundary conditions generating primarily one or the other set of free waves were obtained by reducing the original problem defined by (2.35) and (2.36) to a one-dimensional problem in which both sets of the characteristics become constant latitude lines. The dispersion relation corresponding to (2.49) then becomes

$$\left(\frac{\sigma}{k}\right)^2 + \frac{\sigma}{k}(c + C) - CU = 0, \quad (2.81)$$

where y dependence becomes parametric. Assume that at the eastern boundary z has the form $z(a, y) = 1e^{ik(y)x}$ and select one of the two roots of (2.81) for k . The eastern boundary amplitude of ϕ (equation 2.41) is then obtained from the unforced equation (2.36) as $\phi(a, y) = -\frac{\sigma}{k(y)}z(a, y)$. When these are used to specify z and ϕ at the eastern boundary, the result is a solution dominated by the wave corresponding to the chosen root k of (2.81).

Figure 2.13 shows a snapshot of the time-dependent interface displacement of the base of the middle layer for the free wave solution generated by a wavemaker at the eastern boundary that primarily generates the short wave part of the solution. (The parameters used to describe the background steady circulation in this example are the same as in Figure 2.5.) The structure of this short wave solution in the southwestern corner of the subtropical gyre indicates that the meridional resolution is not sufficient to resolve the short waves in this part of the basin. The unstable region occupies the same part of the basin as in Figure 2.5 and has the same structure, with phases propagating towards the SW. We believe that the SW propagation is associated with meridional aliasing of the short wave, as discussed above.

Figure 2.14 shows a snapshot of the time-dependent interface displacement of the base of the middle layer in which the waves are generated by a wavemaker at the

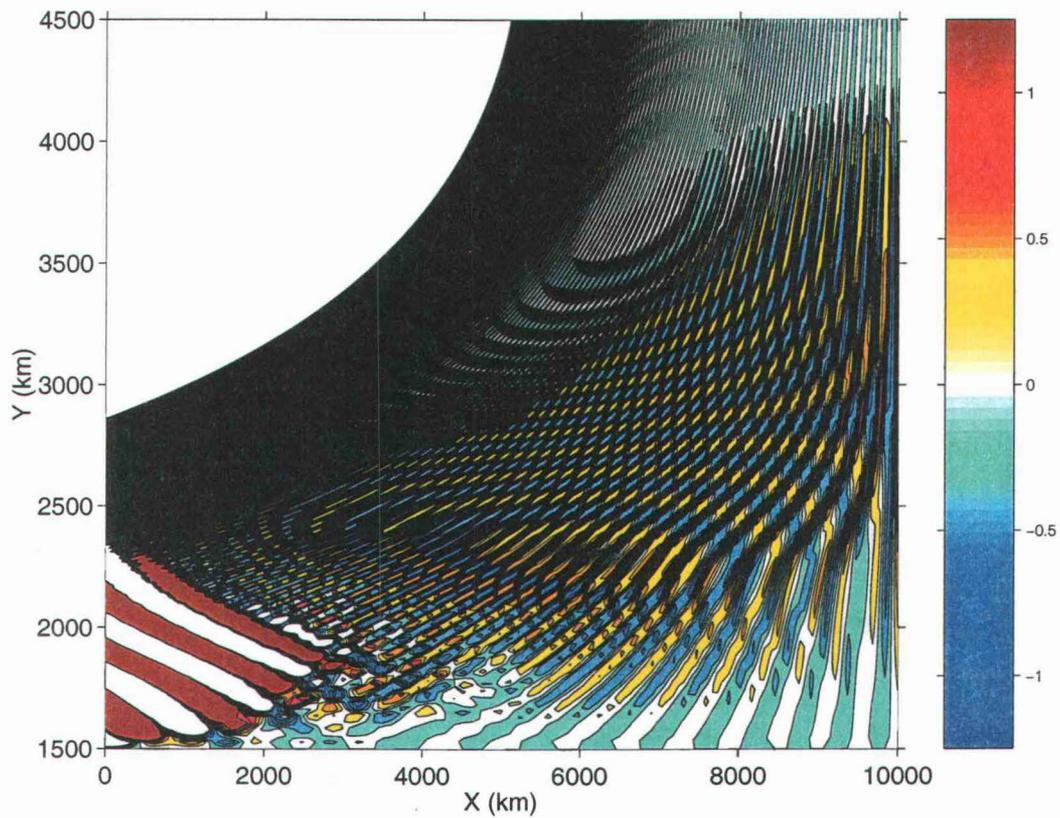


FIGURE 2.13. A snapshot of the time-dependent interface displacement of the base of the middle layer (m). The free annual period waves are forced by a wavemaker at the eastern boundary that primarily generates the short wave part of the solution. The parameters describing the background steady circulation are the same as in Figure 2.5.

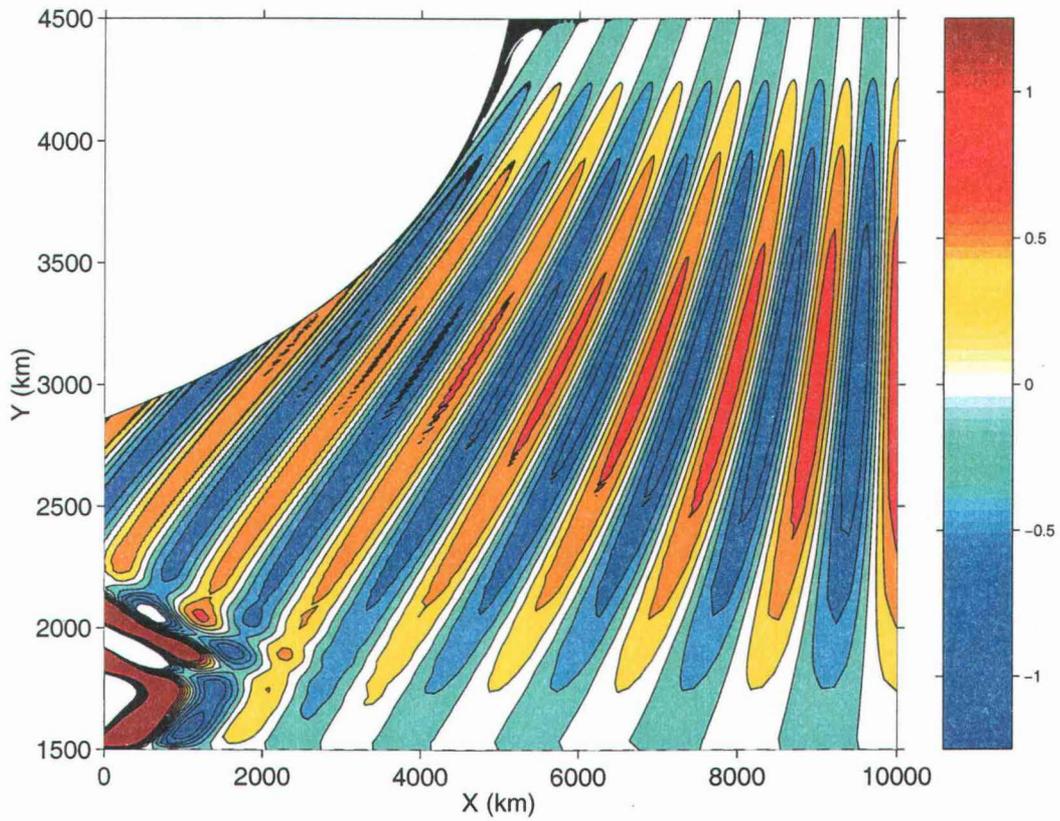


FIGURE 2.14. A snapshot of the time-dependent interface displacement of the base of the middle layer (m). The free annual period waves are forced by a wavemaker at the eastern boundary that primarily generates the long wave part of the solution. The parameters describing the background steady circulation are the same as in Figure 2.5.

eastern boundary that primarily generates the long wave part of the solution. The structure of the long wave solutions in the far southwestern part of the subtropical gyre is more complex than that for the short waves. In the very southwesternmost part of the domain, the long wave solution of Figure 2.14 appears to grow westward and to propagate towards the northwest, while some further distance poleward away from the southern boundary of the domain, the solution grows westward but propagates towards the SW. We believe that the southwest propagating feature is in fact the alias of the small amount of the short wave that remains in our numerical attempt to isolate the long wave because the eastern boundary conditions that are intended to choose one wave or the other do not fully take into account the meridional variation of the steady circulation and so slightly mix the two waves. We believe however that the growth of the northwest propagating part of the numerical solution in the southwestern most part of the basin is caused by true baroclinic instability.

2.9. Discussion

2.9.1. The possibility of true baroclinic instability in the southwestern part of the subtropical EZ

2.9.1.1. Satisfaction of the Charney-Stern criterion

For sufficiently strong background circulation our numerical solutions grow westward in southwestern part of the subtropical EZ. Gill et al. (1974) were apparently the first to point out that westward flow in this region is prone to baroclinic instability on account of the upper layer being thinner than the lower one. Dewar (1998) modeled the propagation of long planetary waves through a background

circulation using a three-layer flat-bottom quasigeostrophic model. In the southwestern corner of the subtropical gyre his solutions develop small scale structures; about which he comments ...”they seem to be related to a reversal with depth of potential vorticity gradient... which makes it potentially baroclinically unstable”. Dewar and Huang (2001) also find, in their QG model, small-scale structure in the southwestern part of the subtropical gyre, and they note that in that part of the basin the meridional gradient of the PV changes sign with depth, so the flow satisfies the necessary criterion for the baroclinic instability. Samelson (1999) found baroclinic instability extending to very short (km) wavelengths in perturbations linearised around the steady mid basin flow profile of a PG ocean circulation model (Samelson and Vallis, 1997). This instability arises from reversal in the water column of the potential vorticity gradient. Differently from the present calculation, a similar profile from the southern part of the basin was unstable only to much longer wavelength perturbations.

We therefore check if the background flow in the southwestern part of our subtropical gyre (where it may be idealized as nearly zonal) satisfies the Charney-Stern criterion, necessary for the occurrence of baroclinic instability in the presence of zonal background flow (Pedlosky, 1987). Figure 2.15 shows the contours of the meridional PV gradient in the upper layer for the numerical solution shown in Figure 2.11. West of the zero line the meridional derivatives of PV in the upper and the middle layer have opposite signs and the solution may correspondingly be unstable.

2.9.1.2. Meridional flow effects

How good is the idealization of zonal background flow in the southwestern part of the gyre? The local (WKB) analysis in which the background circulation

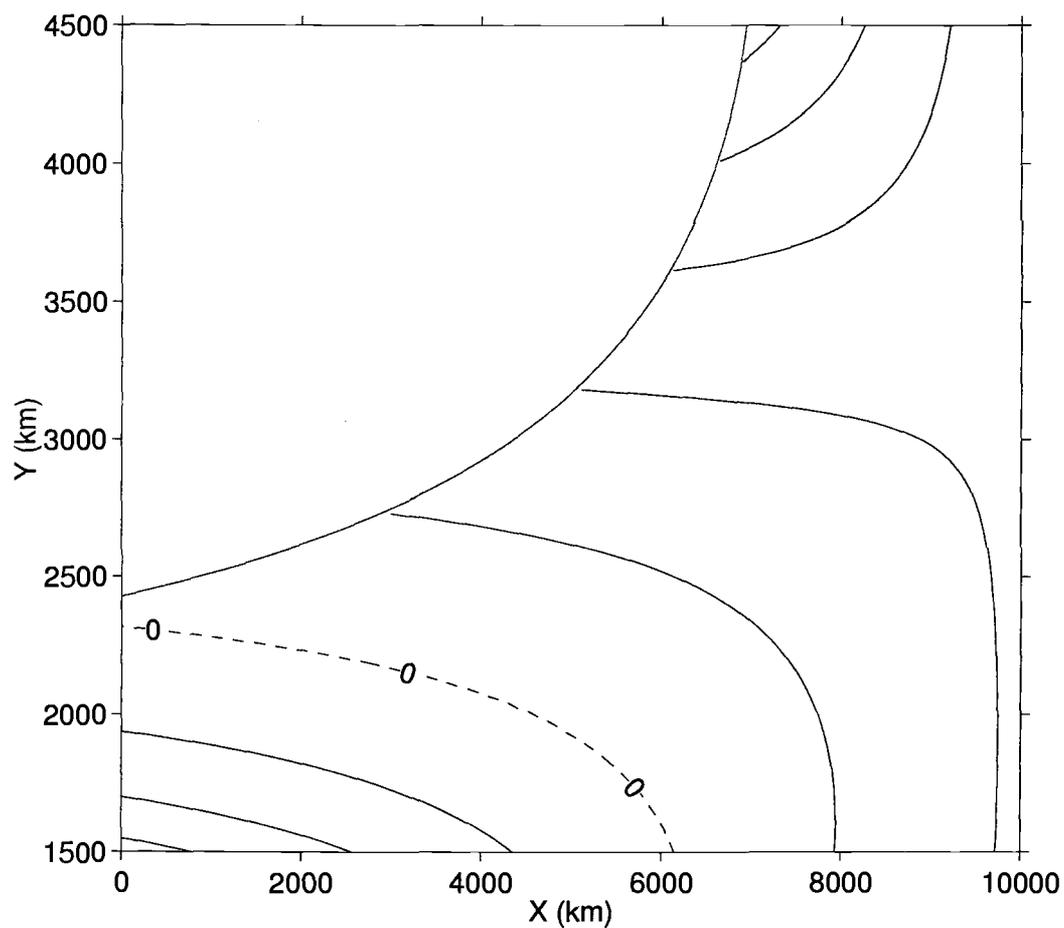


FIGURE 2.15. Meridional potential vorticity gradient of the steady flow in the upper layer ($10^{-13} \text{ m}^{-2} \text{ s}^{-1}$) for the solution shown in Figure 2.11. West of the zero line meridional derivatives of potential vorticity in the upper and the middle layer have the opposite sign and the Charney-Stern criterion for the instability of zonal background flow is satisfied.

is assumed spatially homogeneous has resulted in two sets of dispersion curves in the wavenumber plane of Figure 2.16. In the presence of meridional background flow they may be extrapolated towards a single point of intersection. In the vicinity of that point the horizontal wavenumber becomes complex, corresponding to zonal growth or decay, in a manner reminiscent of the occurrence of instability in e.g. Hayashi and Young (1987).

The locus of unstable wavenumbers depends strongly on the size of the local meridional flow. This is illustrated in Figure 2.16 showing the local analytical dispersion relation for points with fixed $x=2000$ km and varying $y = 1500$ km, 1502 km, 1510 km and 2500 km in the unstable region of Figure 2.11.

At the southern boundary of the subtropical gyre, at $y = 1500$ km, the meridional component of the background flow is zero, so the planetary geostrophic dispersion relation (2.49) becomes independent of l and the wave branches are parallel lines in the wavenumber plane. The zonal background flow is sufficiently strong and westward to make the analytical solution unstable (Figure 2.16 a). The solution consists of a complex conjugate pair of waves that grow and decay in the zonal direction. Along the southern boundary of the subtropical gyre the analytical dispersion relation similarly displays unstable solutions to the west of $x = 6145$ km for Figure 2.11.

Even artificially weak meridional background flow qualitatively dramatically changes the analytical dispersion relation. A little distance to the north of the gyre southern boundary, the meridional component of the background flow becomes different from zero. This non zero meridional flow component makes the zonal perturbations (which are not stabilized by β effect) more capable of releasing potential energy of the background flow (Pedlosky, 1987). The presence of the meridional background flow introduces l dependence into the dispersion relation, and the two

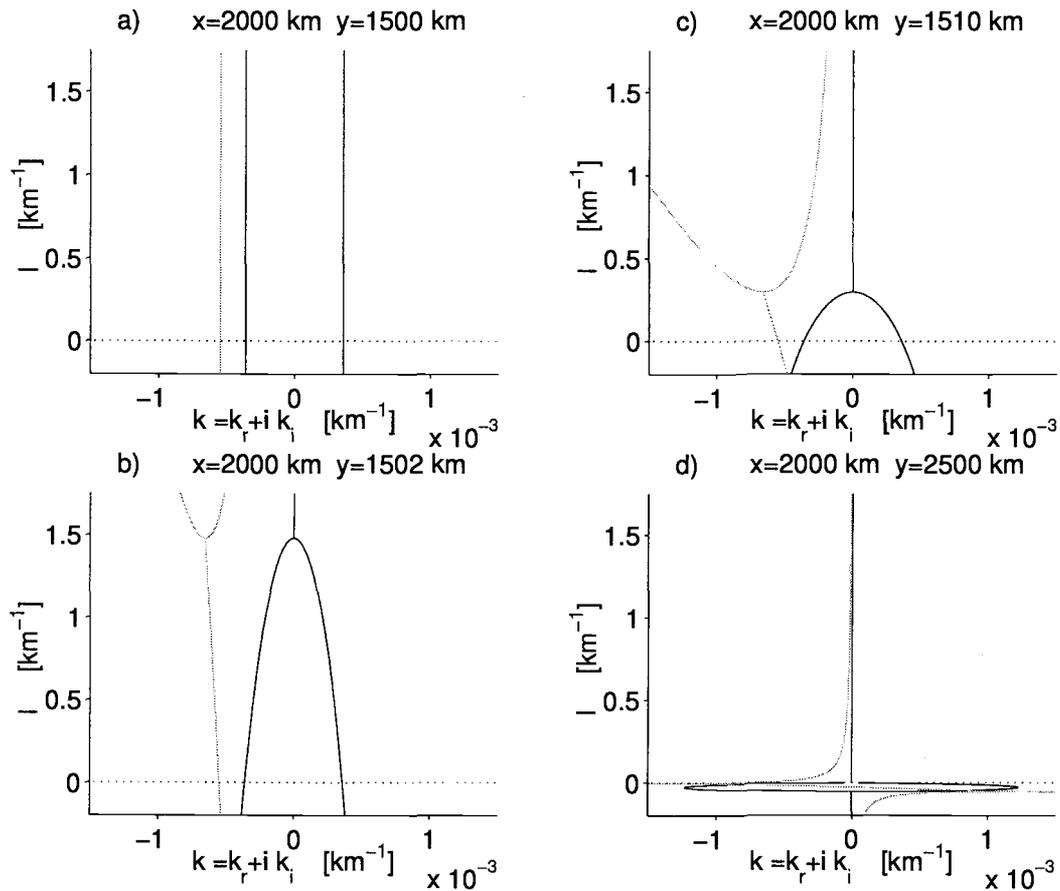


FIGURE 2.16. Local analytical dispersion relation for four points in the southwestern part of the subtropical gyre with the fixed $x=2000$ km and varying y : 1500 km, 1502 km, 1510 km and 2500 km, for the numerical solution shown in Figure 2.11. Real zonal wavenumber k is plotted as a light solid line and the imaginary zonal wavenumber k is plotted as a dark solid line. Comparison of panels a-d illustrates that a very small change in the meridional component of the background flow dramatically changes the range of the wavenumbers for which the solution is unstable; it also significantly changes the phase speed of the short wave part of the EZ solution.

wave branches are no longer parallel lines in the wavenumber plane. They may be extrapolated towards a single point of intersection (Figure 2.16, panels b-d); in the vicinity of that point the horizontal wavenumber becomes complex generating zonally growing and decaying waves. Sufficiently close to the southern edge of the gyre the unstable region is partly in $l > 0$ (panels b and c) and at locations further north it moves down to be entirely in $l < 0$ (panel d). Notice however that the analytical dispersion relation in this form, in which we have solved (2.49) for a range of real values of l , may not always be directly useful in discussing the instability of the southwestern part of the subtropical gyre because there it is possible that either or both of k and l may be complex.

2.9.1.3. Local forcing

The analysis required to determine whether local forcing can excite the instability is sufficiently intricate that it will be dealt with in a separate paper (Cerovečki and deSzoeke, in preparation). The most important result of that work is that when localized periodic perturbation of a spatially homogenous background flow is initiated, a growing transient disturbance propagates away from the region of forcing leaving behind a purely periodic (in time) response that grows spatially. This indicates that both local wind forcing and incidence of energy from the east could excite the instability.

2.9.1.4. Refraction

In the dispersion relation (2.49) both stable wave branches for positive values of l have a westward zonal component of the group velocity, while for negative values of l the zonal component of the group velocity of the stable short waves is westward,

and the zonal component of the group velocity for stable long waves is eastward. As now to be discussed, this is an important feature of the solution when considering if/how a group of incident stable waves may refract into the region where the local analytical dispersion relation indicates that there are unstable solutions.

For simplicity assume that instead of changing continuously in the horizontal direction, the background flow has a meridional discontinuity to the east of which only stable waves have been generated and to the west of which the dispersion relation indicates that there are some possibly unstable solutions. What happens to an incoming wave group of stable waves as it arrives to such a discontinuity? Figure 2.17 shows the local analytical dispersion relation for two points in the southernmost part of a subtropical gyre driven by steady wind forcing with an amplitude of realistic strength (shown in Figure 2.18). The right panel of Figure 2.17 shows the dispersion relation for a point in the southeast part of the basin where the numerical solution consists only of stable waves; the left panel shows the dispersion relation for a point in the southwestern part of the basin where the numerical solution becomes unstable. Assume that these two dispersion relations correspond to the regions to the east and to the west of the discontinuity, respectively. Consider an incident group of stable waves with specified meridional wavenumber vector component l arriving at the discontinuity from the west. What kind of reflected and/or transmitted waves will result?

Comparison of the dispersion relations in the two regions indicates that there are four possibilities: (i) If the meridional component of the wavenumber vector l of the incoming wave is positive and larger than the largest value of l for which the solution in the region to the west of the discontinuity becomes unstable, no reflected waves can be generated because the zonal component of the group velocity is westward for both incoming waves. In order to satisfy both matching conditions

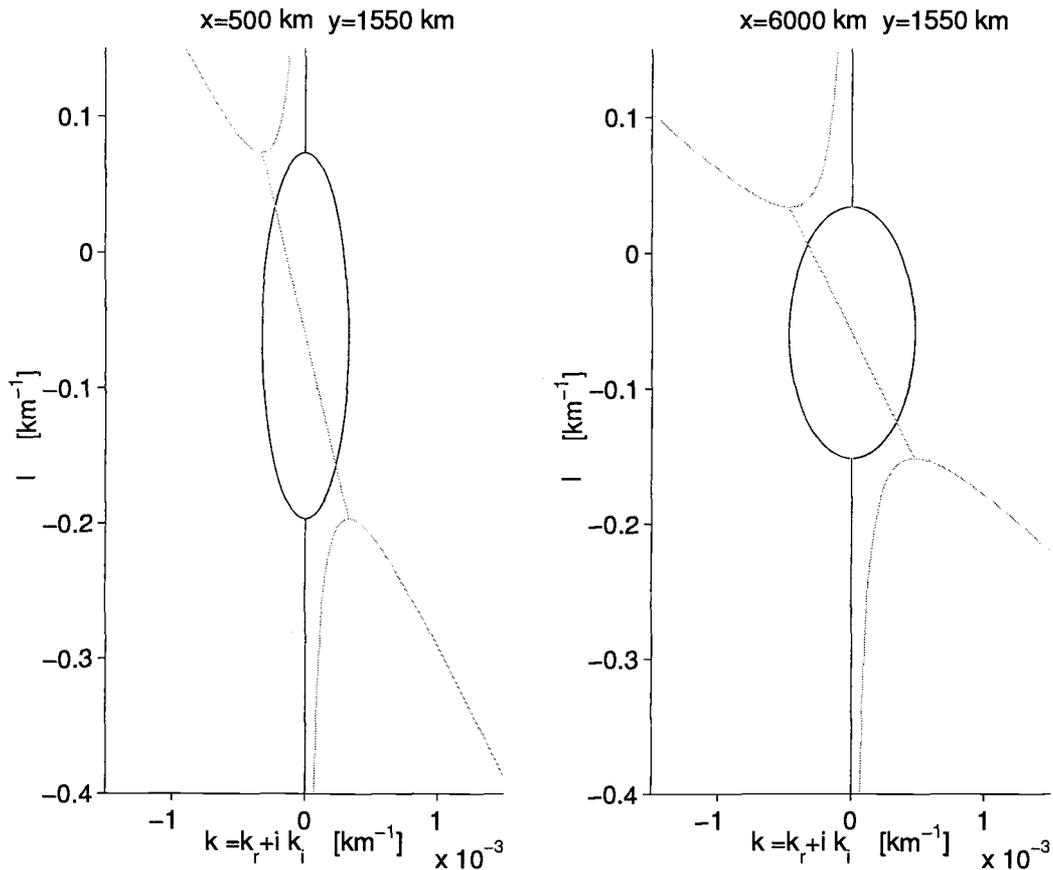


FIGURE 2.17. Local analytical dispersion relation for two points in the southernmost part of the EZ for the numerical solution shown and described in Figure 2.18 with coordinates $x=500$ km, $y=1550$ km (left panel) and $x=6000$ km, $y=1550$ km (right panel). Assume that there is a meridional discontinuity in the background flow such that the dispersion relation in the right panel is representative for the region to the east of the discontinuity in which region only the stable waves have been excited; the dispersion relation in the left panel is representative for the region to the west of the discontinuity. The comparison of two dispersion relations shows that the incident group of stable waves of different meridional wavelengths when arriving at the discontinuity will generate either stable or unstable waves, depending on their meridional wavelengths.

across the discontinuity two transmitted waves will be required. (ii) If the incoming stable wave has a positive value of l within the range of l for which both solutions to the west of the discontinuity are unstable, again no reflected waves can be generated and two transmitted waves will be generated, one growing and one decaying to the west of the discontinuity. (iii) If the incoming stable wave has a negative value of l within the range of l for which both solutions to the west of the discontinuity are unstable, then a reflected wave can be generated since for $l < 0$ the long wave has eastward zonal group velocity. The transmitted wave in the region to the west of the discontinuity may then be a decaying wave. (iv) If the incoming stable wave has a negative value of l below the range of l for which both solutions to the west of the discontinuity are unstable, then again a reflected wave can be generated since for $l < 0$ the long wave has eastward zonal group velocity. The result will either be a reflected wave and a transmitted wave, or two transmitted waves. The important point of this discussion is that, even when l is assumed to be real, it is possible for wave energy incident from the east to refract into the unstable part of the analytical dispersion relation.

2.9.1.5. Relative vorticity

The wavelengths of the waves which become baroclinically unstable in the southernmost part of the gyre are sufficiently long (of the order of 1000 km, see Figures 2.5, 2.11 and 2.18) for relative vorticity to be neglected. On the other hand the dispersion relation plots indicate that there are some possible baroclinically unstable solutions at shorter wavelengths (see e.g. Figure 2.16 b) for which the relative vorticity may become important. Therefore it would be interesting to compare the analytical dispersion relations with and without relative vorticity to check its influ-

ence on the stability properties of the waves. We have considered two-and-a-half layer QG equations with relative vorticity and no steady flow in the middle and the lowest layer (e.g. Pedlosky, 1987). With relative vorticity included, one can fix the frequency and carry out a similar WKB analysis. The zonal wavenumber component now obeys a fifth order dispersion relation which can be solved numerically; again there are complex roots. With relative vorticity terms included, the unstable region that corresponds to the unstable region that may arise in the PG limit tends to be smaller, but the presence of the additional wave branches introduces new unstable regions where some modes may be evanescent.

2.9.2. Numerical solution for the parameters representative of North Pacific subtropical gyre

Figures 2.18 and 2.19 show a snapshot of the time-dependent interface displacement of the middle layer and sea surface elevation for a set of parameters representative of a density profile in the North Pacific subtropical gyre; the amplitude of the Ekman pumping is of the realistic strength ($H_{10}=500$ m, $H_{20}=900$ m, $\gamma_1 = 0.03$ m s⁻² and $\gamma_2 = 0.01$ m s⁻², $W_0=9.1 \cdot 10^{-7}$ m s⁻¹). Stronger steady wind forcing pushes the boundary x_B towards the eastern boundary of the ocean, making the EZ narrow. The waves in the EZ are unstable in a large part of the gyre. Exponential growth in the southernmost part of the gyre is caused by baroclinic instability. At the southern edge of the subtropical gyre $y = 1500$ km the dispersion relation (2.49) indicates that the flow is unstable for $x < 8010$ km. Exponential growth in a band extending approximately 150 km equatorward from to the boundary x_B is caused by short waves becoming numerically unstable. This causes the WZ solution to grow exponentially in the band approximately 350 km wide extending

poleward from the boundary x_B to about $y=2100$ km. Bending of the contour lines in the WZ for $2000 \text{ km} < y < 2200 \text{ km}$ is introduced by averaging.

In the WZ the solution consists of one set of long planetary waves that propagate along the constant latitude lines. The ratio of the phase speeds obtained from the approximate WKB solution (2.55) to the phase speed in the case of no background mean flow in a two-and-a-half layer system (e.g. de Szoeke and Chelton, 1999) is only slightly above one. For the set of parameters listed above the ratio ranges from only 1.07 to 1.23 because our two-and-a-half layer model cannot give rise to the larger speedup of the Rossby waves that occurs in the presence of a constant PV pool in a finite depth model (see e.g. de Szoeke and Chelton, 1999).

2.9.3. Comparison with the observations and with other solutions

The altimeter mainly sees the long waves. Figure 2.20 shows time-longitude plots for the numerical solution of Figure 2.5 analogous to those of CS. Both appear to show westward propagating waves, but they differ in important ways that are discussed in the following subsections.

The phase velocity determined from the slope of the wave signal in the time-longitude plot for $y=2505$ km, shown in middle panel of Figure 2.20, is 6.1 cm s^{-1} . This speed is a longitudinal average. The corresponding phase velocity of the long wave in the absence of the background circulation obtained from (2.49) is 3.91 cm s^{-1} , so that a speed increase by the factor 1.56 is associated with the background flow. When the geostrophic circulation in (2.49) is set to zero, but the deepening of the thermocline associated with the background flow is retained, the average phase velocity at $y=2505$ km is 4.41 cm s^{-1} (e.g. at $x=2000$ km, the phase velocity is 4.69 cm s^{-1} ; at $x=5000$ km it is 4.42 cm s^{-1} ; at $x=8000$ km it equals

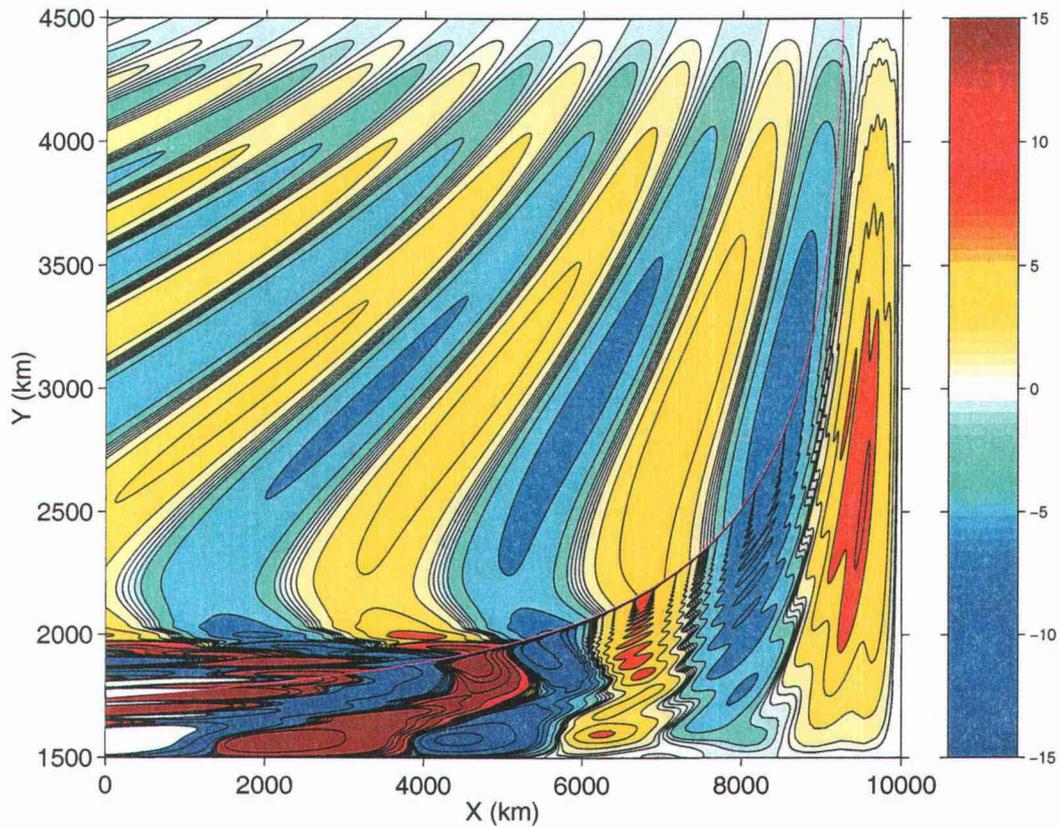


FIGURE 2.18. A snapshot of the time-dependent interface displacement of the base of the middle layer (m). The parameters used in this example are representative of a density profile in the North Pacific and the amplitude of the Ekman pumping is of realistic strength ($H_{10}=500$ m, $H_{20}=900$ m, $\gamma_1 = 0.03$ m s⁻² and $\gamma_2 = 0.01$ m s⁻², $W_0=9.1 \cdot 10^{-7}$ m s⁻¹). This solution was obtained by using the five point average with the central weight $5/8$, side weights $1/4$ and the weight of the outside points $-1/16$.

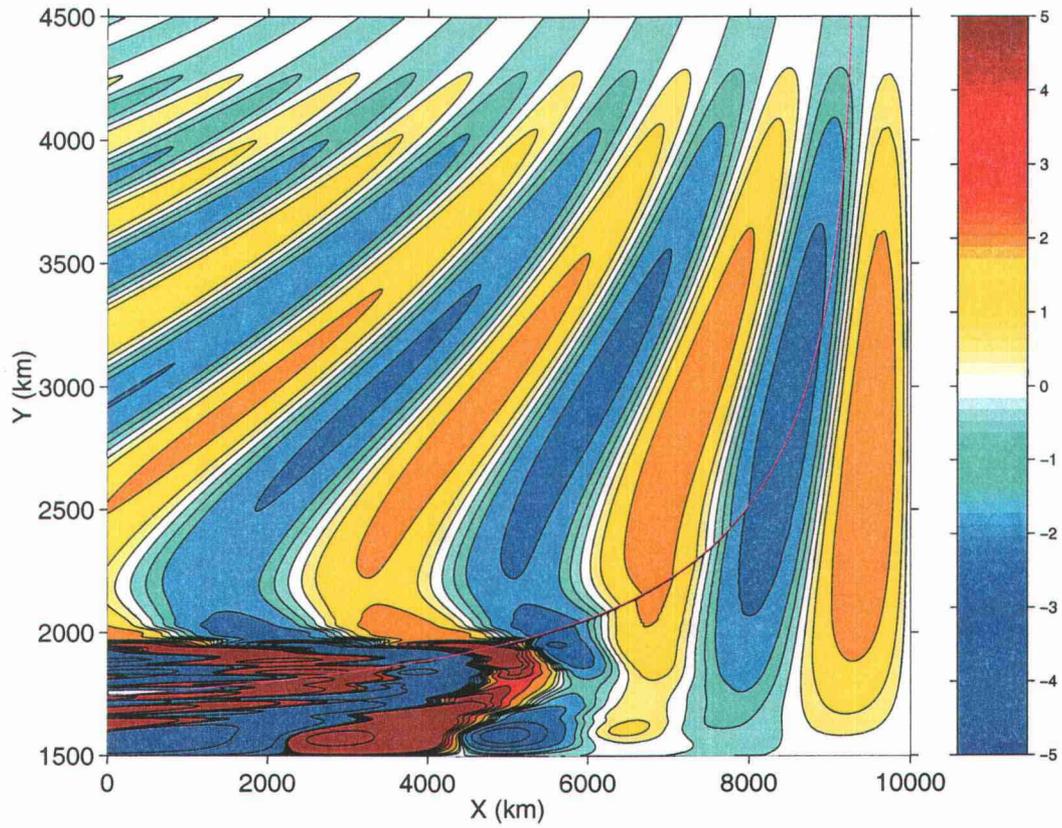


FIGURE 2.19. A snapshot of the sea surface elevation (cm) for the parameters as in Figure 2.18.

4.12 cm s⁻¹). This suggests that the motion of the water associated with the background flow has a somewhat larger effect on the Rossby wave propagation than the westward deepening of the thermocline that is associated with it.

9.3.1 The CS plots appear to show westward propagation of every maximum, while in the time-longitude plots for the numerical solution every other node remains fixed in time, indicating that the numerical solution consists of a propagating annually varying part plus a much larger scale annually varying part.

It is easy to understand from a one dimensional idealization of the numerical solution with the constant coefficients why the numerical solution contains both parts. In that idealization, $\frac{dz}{dy}$ is neglected in equation (2.37) so that z satisfies

$$CUz_{xx} - i\sigma(c + C)z_x + \sigma^2z = i\sigma\tilde{\Theta}. \quad (2.82)$$

with the boundary condition $z = 0$ specified at the eastern boundary $x = a$. In this equation $\tilde{\Theta}$ is independent of x for the particular wind forcing chosen here. The solution thus clearly consists of the particular x -independent solution $z = \frac{i\tilde{\Theta}}{\sigma}$ plus a sinusoidal solution that satisfies the homogeneous equation and is equal to the negative of the particular solution at $x = a$. Schematically we have $\cos(\sigma t) - \cos(\sigma t + k(x-a)) = -2\sin(\sigma t + k(x-a)/2)\sin(k(x-a)/2)$, for which every other node (at $k(x-a)/2 = -n\pi$) remains fixed in time while the ones in between propagate westward. White (1977) was the first one to remark on the factor 2 appearing in the effective zonal wavenumber. As this argument would suggest, in the numerical solutions Figure 2.13 and 2.14 for short and long waves excited by specified layer depths at the eastern boundary, all the nodes propagate westward in time. This is an important difference between our solutions and the results of CS, but in the Indian ocean the situation is different.

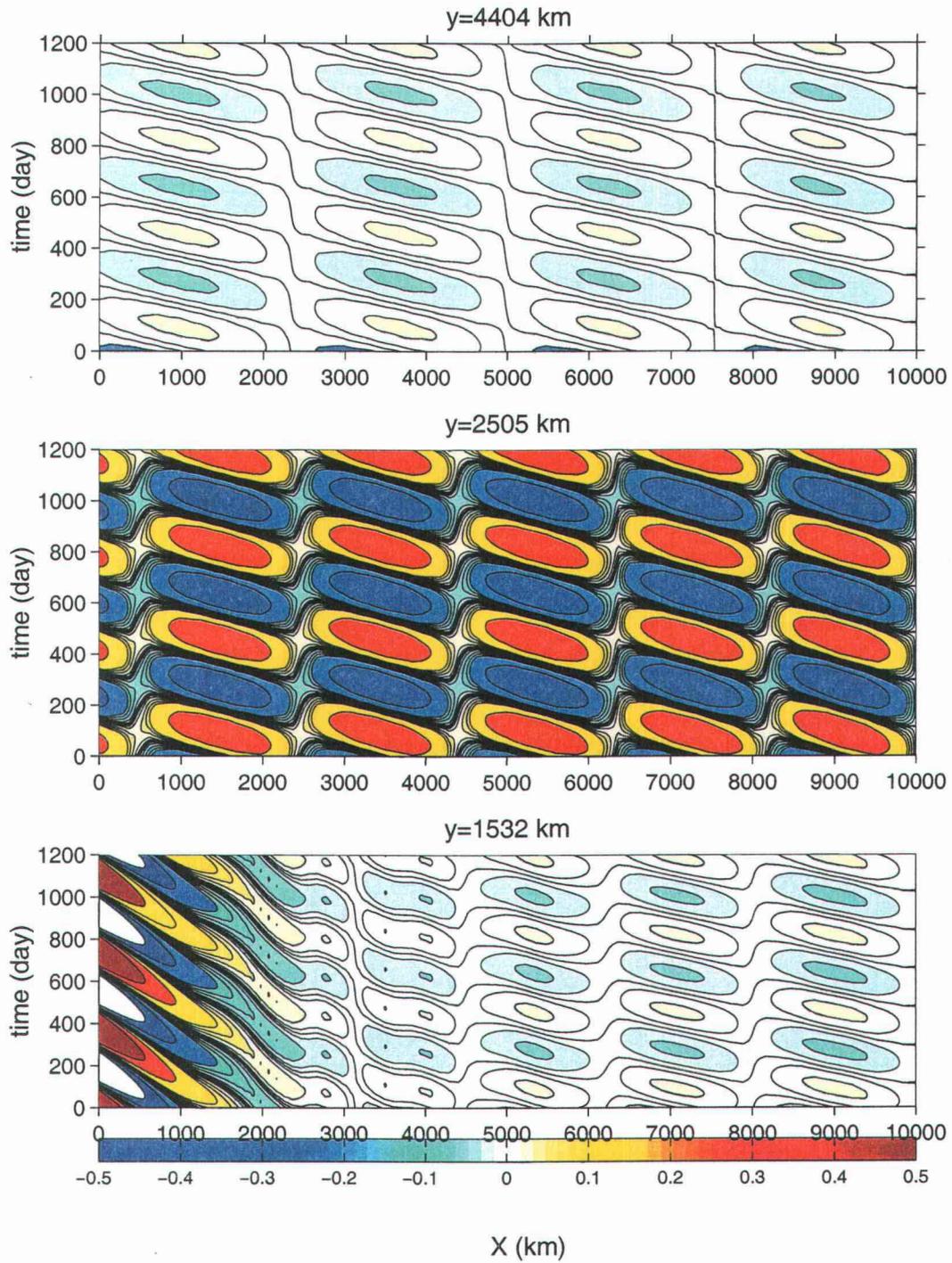


FIGURE 2.20. Time-longitude sections for the numerical solution of Figure 2.5.

9.3.2 The foregoing discussion is very closely related to that of Wang et al. (2001) who consider the annual Rossby wave in the Indian ocean. They do not consider the eastern boundary to be an impenetrable coast, but rather suppose that annual period energy from the Straits of Indonesia propagates westward into the Indian ocean there to be added to the annual period waves generated by local winds. In their analysis there is thus a clear reason for the wave field to consist of an entirely free part, and a part like our solutions that contains both a propagating and a non propagating part. In their solutions these two parts interfere constructively in the eastern and the western parts of the basin with an intervening minimum, and their analysis of the radar altimetry suggests a similar pattern in the observations. Our solutions do not show any such pattern because we do not specify a free wave component that is independent of the local wind forcing.

9.3.3 Differently from Wang et al. (2001), CS suggest that the amplitude of the annual signal increases abruptly as one goes from east to west across the Pacific or the Atlantic. Tailleux and McWilliams (2000) suggest that this is associated with mid ocean bottom relief (such as the Hawaiian Island Chain or the Mid-Atlantic ridge), their scattering calculations suggest that short waves will be much more pronounced to the west of the relief than to the east. This is not clearly apparent in the sections of CS.

Our model of course has no bottom relief, but alternatively suggests that at least for sections in the equatorward part of the ST gyre, baroclinic instability of the gyre may lead to an abrupt westward increase in the amplitude of the annual period signal.

9.3.4 In our model and in models like ours but with finite depth, the annual period signal is very different on different sides of the bounding characteristic x_B . Short waves are only present to the east of x_B , and when the total depth is finite (de

Szoeke and Chelton, 1999) the speed of westward propagation is much greater to the west of x_B than to the east. It would be correspondingly important to know where x_B is in the real ocean. It is not however easy to say because x_B is different for different flat bottom internal modes (the altimeter presumably responds primarily to the lowest mode), and depends on how the circulation changes the propagation properties of that mode, something that cannot be unambiguously deduced from a hydrographic reconstruction of the subtropical gyre flows. Bottom relief might further complicate matters by changing the vertical structure of the modes (as in Hallberg, 1997). The point is that an observational determination of x_B is important in interpreting satellite altimetry, but not easy to construct.

9.3.5 Roemmich and Gilson (2001) studied high resolution XBT transects in the North Pacific Ocean, at an average latitude of 22° N together with TOPEX/Poseidon altimetric data to determine the structure and transport characteristics of the mesoscale eddy field. The eddy field that they study is strongly expressed in both the XBT data and the TOPEX/Poseidon data.

They comment that ...”the ’eddies identified in Fig. 3 are the same features described by Chelton and Schlax (1996) as Rossby waves.” The data do not allow determination of whether the features are in fact eddies, having somewhat similar meridional and zonal scales, or whether they are Rossby waves, having very different scales along and across the propagation direction.

Roemmich and Gilson (2001) construct a ”composite eddy” of all their eddies (their Figure 5) and observe that its vertical structure is tilted westward towards the surface, in the way required for meridional eddy heat transport. They do not construct separate composites in the eastern and western halves of the basin, so it is not possible to say whether the features they observe are baroclinically unstable

in the western part of the basin, as is the case for the annual period waves of the present calculation.

In a more dynamically complete model, the annually forced long wave signal found in the present idealized model might become an eddy field whose intensity is annually modulated. It is possible that if the degree of nonlinearity is not too intense, the eddies although nonlinear may still be observed to propagate in a manner closely related to the way in which the linearized disturbances propagate. Something along these lines seems to occur in models of shear instability in coastal currents (Feddersen, 1998); there individual disturbances are highly nonlinear and eddy like but their speeds of propagation nonetheless fall fairly closely on the dispersion relation for linearized perturbations.

Roemmich and Gilson (2001) note that the eddy field has considerable variability on seasonal to interannual time-scales and suggest that "the eddy field may be a major contributor to time variability of northward heat transport." If their mesoscale eddy field is related to the annual period altimetry signal studied by Chelton and Schlax, then it might be that studies such as the present one of the propagation of annual period Rossby waves through the steady circulation can suggest how interannual changes in that circulation might modify the meridional heat transport.

2.10. Summary

Wavelike annual period signals are prominent in satellite altimetry of the ocean surface as well as in sufficiently detailed hydrographic data. In order to understand how the structure of the annual period signal is affected by the underlying large scale circulation, we consider a model in which long planetary waves are lin-

earized about a schematic, wind-driven, steady subtropical gyre in a three layer ocean. The steady circulation divides the gyre in an east zone (EZ) where only the uppermost layer is in steady motion, and a west zone (WZ) where the uppermost and middle layers are in steady motion because the potential vorticity of the middle layer is assumed homogeneous. In both zones the upper two layers participate in the wave motion, which is driven by a wind stress fluctuating at the annual period, but the lowest layer is quiescent because it is assumed to be infinitely deep.

The waves are governed by spatially hyperbolic equations having two families of characteristics. One family is zonal and carries information from the eastern boundary into the WZ; characteristics of the other family emanate from the eastern or western boundaries, but do not cross the boundary between EZ and WZ. The EZ wave equations are numerically integrated westward away from the eastern boundary along their characteristics, only the westward integration along the family of zonal characteristics can continue into the WZ. Because the model domain excludes the western boundary current region, the part of the solution that propagates along characteristics emanating from the western boundary is neglected.

In the EZ two sets of baroclinic waves are apparent in the solutions. They resemble sinuous and varicose vertical modes in certain limits. The wavelengths of one set of waves are much smaller than those characterizing the basin or the wind field. Each is a combination, coupled by the background flow, of the classical first and second mode baroclinic Rossby waves with quiescent background flow, but in marked contrast with the case of quiescent background flow, each is strongly dispersive. Both sets of waves are generated by wind. As the short waves propagate westward in the presence of the background circulation, they get refracted into shorter wavelengths. These wavelengths are shorter than the ones of the second internal baroclinic mode. WKB solutions agree well with the numerical waves both

with regards to local wavelength and direction of propagation as well as to the overall basin wide distribution of wave energy. The WKB solutions allow construction of each set of waves independently of constraints on resolution imposed by the numerics. It is important to note that, on account of the way in which the background flow modifies the wave dispersion relations, the short waves can become so short in the vicinity of x_B that the neglect of relative vorticity common to all planetary geostrophic calculations is no longer self consistent.

In the southwest portion of the subtropical gyre the annual period numerical solutions may become baroclinically unstable and grow exponentially westward. The WKB solutions additionally confirm the possibility of this instability in the southwestern subtropical gyre. In the WKB solutions, instability occurs near the point in the horizontal wavenumber plane where the two stable wave branches, extrapolated across the unstable wavelengths, intersect. The instability is excited by wave energy that refracts into the unstable region from more easterly regions of wind forcing or directly by the local wind forcing.

The numerical procedure is examined using an extension of the classical von Neumann stability analysis. It is found that the procedure gives rise to spurious instabilities when finite resolution aliases one of the stable wave branches so that it crosses the other at wavenumbers that are stable in the analytic WKB solution. The important result is that naive refinement of the resolution does not necessarily improve the accuracy of the solution. The numerical instabilities may be suppressed by a moderate amount of averaging over adjacent characteristics, corresponding to a lateral diffusivity of about $14 \text{ m}^2 \text{ s}^{-1}$ in our examples, but careful analysis is needed to distinguish between true and numerical instability in the southwestern subtropical gyre.

Numerical time-longitude plots of sealevel qualitatively resemble those of Chelton and Schlax (1996) but there are some important differences. To begin with, our time-longitude plots show a gradual westward increase of wave amplitude (except for sections that penetrate the baroclinically unstable region of the southwestern subtropical gyre) whereas both the sections of CS at 21° N, 32° N and 39° N as well as the section of Roemmich and Gilson (2001) show an abrupt increase in amplitude at roughly the longitude of the Hawaiian Island Chain. Tailleux and McWilliams (2000) suggest that mid basin ridge bottom topography may result in propagating Rossby waves west of the topography. The only similarly abrupt westward change in our computations is on time-longitude sections that penetrate into the baroclinically unstable region of the southwestern subtropical gyre. Our time-longitude plots also differ from those of CS and of Roemmich and Gilson (2001) in the presence of clearly defined stationary nodes through which the waves propagate, indicating that in addition to the westward traveling waves our solution also contains a very large scale annually varying component. Such a situation is however found in the analysis of Wang et al. (2001) of sealevel in the Indian Ocean. These authors attribute the stationary node to destructive interference between the response of the ocean to local annual period winds and eastern boundary radiation of Rossby waves. Finally, the similarity between the annual period time-longitude plots of CS and the time-longitude plots of "eddy locations" made by Roemmich and Gilson (2001) for both altimetry and XBT data suggest to us that much of the eddy signal of Roemmich and Gilson is the annual Rossby wave field studied here. Roemmich and Gilson construct a composite eddy from all their data. On a zonal transect through the composite eddy, each isotherm is offset westward relative to the one immediately below, a structure consonant both with baroclinic instability and with poleward heat transport by the eddy. If indeed much of the eddy signal of Roemmich and

Gilson is the annual Rossby wave field studied here, then studies such as those of the present paper may be useful in studying interannual variation of eddy heat transport in the major ocean gyres.

2.11. Appendix

2.11.1. Perturbation equations in the EZ

The constant-latitude characteristics cross the boundary between the EZ and WZ, though the geostrophic contour characteristics do not. (Indeed, the boundary is defined by the geostrophic contour $Y(x_B(y); b) = b$.) To ensure that both sets of characteristics "arrive" at the boundary between EZ and WZ together, a stretched coordinate is introduced by means of

$$\xi = -a \frac{(x - a)}{(x_B(y) - a)} \quad (2.83)$$

in terms of which $x_B(y)$ is mapped into $\xi = -a$, and the eastern boundary is at $\xi = 0$. A further logarithmic transformation is made in order to increase the resolution in the region of the EZ close to x_B by means of

$$\chi = a \ln \left(\frac{\xi + a}{a} \right). \quad (2.84)$$

The eastern boundary corresponds to $\chi = 0$ and x_B to $\chi = -\infty$. In terms of χ equation (2.42) for \mathcal{Z} becomes

$$\left. \frac{\partial \mathcal{Z}}{\partial \chi} \right|_{y_0} = \frac{-i\sigma \left[\left(1 + \frac{c}{C}\right) \mathcal{Z} + \frac{\Phi}{C} \right] \frac{\partial Y}{\partial \chi}}{\frac{\partial \Psi}{\partial x}}. \quad (2.85)$$

It is integrated westward on a set of the Y characteristics obtained from

$$\left. \frac{\partial Y}{\partial \chi} \right|_{y_0} = \frac{-\Psi_x}{\Psi_y \frac{\partial \chi}{\partial x} \Big|_y - \Psi_x \frac{\partial \chi}{\partial y} \Big|_x}. \quad (2.86)$$

We advance ϕ westward by integrating (2.40) on constant latitude lines, as

$$\left. \frac{\partial \phi}{\partial \chi} \right|_y = \frac{i\sigma z + \tilde{\Theta}}{\frac{\partial \chi}{\partial x} \Big|_y}. \quad (2.87)$$

2.11.2. Perturbation equations in the WZ

In the WZ the same equation as the equation (2.40) in the EZ has been integrated westward along its characteristics, which are constant latitude lines, as

$$\frac{\partial \phi}{\partial \xi} = \frac{\partial x}{\partial \xi} \Big|_y (i\sigma z + \tilde{\Theta}). \quad (2.88)$$

The stretched variable ξ defined by (2.83) has been used, which maps the curved boundary $x_B(y)$ into a straight line.

References

- Bender, C. M. and S. A. Orszag, 1978: Advanced mathematical methods for scientists and engineers. McGraw-Hill, Inc., 593 pp.
- Chelton, D. B., and M. G. Schlax, 1996: Global observations of oceanic Rossby waves. *Science*, **272**, 234–238.
- de Szoeke, R. A., 1992: Effects of Alongshore Wind Stress at the Eastern Boundary on Circulation in the Ocean Pycnocline. *J. Phys. Oceanogr.*, **29**, 500–511.
- de Szoeke, R. A., and D. Chelton, 1999: The modification of long planetary waves by homogeneous potential vorticity layers. *J. Phys. Oceanogr.*, **22**, 247–267.
- Dewar, W. K., 1998. On "too-fast" baroclinic planetary waves in the general circulation. *J. Phys. Oceanogr.*, **28**, 1739–1758.
- Dewar, W. K., M. Y. Morris, 2000: On the propagation of baroclinic waves in the general circulation. *J. Phys. Oceanogr.*, **30**, 2637–2649.
- Feddersen, F., 1998: Weakly nonlinear shear waves. *J. Fluid Mech.*, **372**, 71–91.

- Gill, A. E., Green, J. S. A., Simmons, A. J., 1974: Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies. *Deep-Sea Research*, **21**, 499–528.
- Hallberg, R. 1997: Localized coupling between surface and bottom-intensified flow over topography. *J. Phys. Oceanogr.*, **27**, 977–998.
- Hayashi, Y.-Y., Young, W. R., 1987: Stable and unstable shear modes of rotating parallel flows in shallow water. *J. Fluid Mech.*, **184**, 477–504.
- Hendershott, M. C., 1989: The ventilated thermocline in quasigeostrophic approximation. *J. Mar. Res.*, **47**, 33–53.
- Huang, R. X., and J. Pedlosky, 2000: Climate variability induced by anomalous buoyancy forcing in a multilayer model of the ventilated thermocline. *J. Phys. Oceanogr.*, **30**, 3009–3021.
- Huang, R. X., and J. Pedlosky, 1999: Climate variability inferred from a layered model of the ventilated thermocline. *J. Phys. Oceanogr.*, **29**, 779–790.
- Keffer, T., 1985: The ventilation of the world's oceans: Maps of the potential vorticity field. *J. Phys. Oceanogr.*, **15**, 509–523.
- Killworth, P. D., D. B. Chelton, and R. A. de Szoeke, 1997: The speed of observed and theoretical long extratropical planetary waves. *J. Phys. Oceanogr.*, **27**, 1946–1966.
- Liu, Z., 1999: Planetary wave modes in the thermocline: Non-Doppler-shift mode, advective mode and Green mode. *Quart. J. Roy. Meteor. Soc.*, **125**, 1315–1339.

- Luyten, J., and H. Stommel, 1986: Gyres driven by combined wind and buoyancy flux. *J. Phys. Oceanogr.*, **16**, 1551–1560.
- Luyten, J., J. Pedlosky, and H. Stommel, 1983: The ventilated thermocline. *J. Phys. Oceanogr.*, **13**, 293–309.
- Pedlosky, J., 1987: *Geophysical Fluid Dynamics*. 2 d ed. Springer-Verlag, 710 pp.
- Pedlosky, J., 1996: *Ocean Circulation Theory*. Springer-Verlag, 453 pp.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1994: *Numerical Recipes in FORTRAN*. Cambridge University Press, 963 pp.
- Qiu, B., W. Miao, and P. Mueller, 1997: Propagation and decay of forced and free baroclinic Rossby waves in off-equatorial oceans. *J. Phys. Oceanogr.*, **27**, 2405–2417.
- Rhines, P. B., and W. R. Young, 1982: A theory of the wind-driven ocean circulation. I. Mid-ocean gyres. *J. Mar. Res.*, **40**, 559–596.
- Roemmich, D. and Gilson, J., 2001: Eddy transport of heat and thermocline waters in the North Pacific: a key to interannual/decadal climate variability? *J. Phys. Oceanogr.*, **31**, 675–687.
- Samelson, R. M., 1999: Note on a baroclinic analogue of vorticity defects in shear. *J. Fluid Mech.*, **382**, 367–373.
- Samelson, R. M., and G. K. Vallis, 1997: Large-scale circulation with small diapycnal diffusion: the two-thermocline limit. *J. Mar. Res.*, **55**, 223–275.
- Talley, L. D., 1988: Potential vorticity distribution in the North Pacific. *J. Phys. Oceanogr.*, **18**, 89–106.

- Tailleux, R. and J. C. McWilliams, 2000: Acceleration, Creation and Depletion of Wind-Driven, Baroclinic Rossby Waves over an Ocean Ridge. *J. Phys. Oceanogr.*, **30**, 2186–2213.
- Young, W. R., and Rhines, P. B., 1982: A theory of the wind-driven ocean circulation. II. Gyres with western boundary layers. *J. Mar. Res.*, **40**, 849–872.
- Wang, L. P., C. J. Koblinsky, S. Howden., 2001: Annual Rossby wave in the southern Indian Ocean: Why does it "appear" to break down in the middle ocean? *J. Phys. Oceanogr.*, **31**, 54–74.
- White, W., 1977: Annual forcing of baroclinic long waves in the tropical North Pacific. *J. Phys. Oceanogr.*, **7**, 50–61.
- White, W., Y. Chao, and C.-K. Tai, 1998: Coupling of biennial oceanic Rossby waves with the overlying atmosphere in the Pacific basin. *J. Phys. Oceanogr.*, **28**, 1236–1251.

3. THE INITIAL VALUE PROBLEM FOR LONG PLANETARY
WAVES
IN A TWO-AND-A-HALF LAYER MODEL IN THE PRESENCE OF
NONZONAL MEAN FLOW

Ivana Cerovečki and Roland A. de Szoeke

To be submitted to *Journal of Physical Oceanography*

Abstract

We consider planetary geostrophic perturbations of a nonzonal upper ocean flow that are generated by localised forcing turned on at an initial time and thereafter oscillating periodically. The resulting solution consists of a transient and a sympathetically oscillating part. Both are obtained by numerically summing Fourier components. A related transient solution is obtained analytically as an asymptotic solution for large time by the method of steepest descents.

Even very gentle background flow dramatically changes long planetary waves away from the usual simple β solutions. With nonzonal mean flow, long planetary waves are radially nondispersive and azimuthally dispersive (i.e. phase speed $\frac{\sigma}{\kappa}$ is only a function of the direction of the wavenumber vector α). A group of long planetary waves in the presence of a nonzonal background flow evolves toward a very anisotropic shape. Only the transient solution may grow spatially (but not the sympathetic solution). The growth of the amplitude of the transient solution depends strongly on the central wavenumber and the spectral width of the initial disturbance. The direction of propagation of the sympathetic solution depends on the central wavenumber and spectral width of the forcing.

3.1. Introduction

Baroclinic instability is an important classical problem in geophysical fluid dynamics. Much of the published work deals with purely zonal flow and seeks growth rates of spatially periodic disturbances. Pedlosky (1987, 2002) has pointed out the importance of considering nonzonal flow. Our interest grew out of a study of annual period Rossby waves propagating through a specified background subtropical circulation (Cerovečki and de Szoeko, submitted). In that study emphasizing annual

period perturbations, numerical solutions indicated the existence of a region in the far southwest subtropical gyre where such perturbations grew zonally towards the west. Such spatially growing instabilities have been studied extensively in other contexts (Briggs 1964, Huerre and Monkewitz, 1990). The goal in this paper is to apply similar methods to our problem, that is, to study the stability of time-periodic perturbations of an ocean subtropical gyre.

In the study by Cerovečki and de Szoeke quoted above, westward growth was observed, but its origin was not entirely clear. The three possibilities are:

(i) the perturbation solution generated by local winds is indeed unstable and grows westward;

(ii) the perturbation solution has propagated westward through inhomogeneous background flow from more easterly regions of wind stress and has been refracted into unstable wavenumbers that grow westward;

(iii) the perturbation solution, though physically stable, appears to grow westward because of numerical instability.

This paper examines the first of these possibilities. The spatial inhomogeneity of the background flow is neglected, giving a solution that is remarkably simple analytically though showing a number of previously unemphasized features. One way to proceed is *ab initio* to suppose time-periodic perturbations and look for spatial growth, though Briggs (1964, p. 12) points out the hazards of this. We therefore consider perturbations generated by localised forcing that is initiated at time $t = 0$ and thereafter oscillates periodically. The resulting solution consists of a transient and a sympathetically oscillating part. Both are obtained by numerically summing the Fourier components. We further consider perturbations generated by a specified initial disturbance where the transient solution is obtained analytically as an asymptotic solution valid for large time by using the method of steepest descents.

The result obtained in this way is verified by comparison to the solution obtained by numerically summing Fourier components.

3.2. The model

Theories of large scale ocean flow commonly make use of the planetary geostrophic approximation, in which the open ocean flow itself is geostrophic except for a surface Ekman layer. A layered planetary geostrophic model (Luyten et al., 1983, hereafter LPS) was used to describe the steady circulation of the subtropical gyre (though with no outcrops in the region of consideration). Three layers, upper, middle and lower, labeled 1, 2 and 3, are considered. The depth of the lower layer will ultimately be allowed to become very large. The momentum balance is geostrophic,

$$-fv_j = -\partial_x P_j, \quad fu_j = -\partial_y P_j, \quad (3.1)$$

where u_j and v_j are horizontal fluid velocities in the zonal (x) and meridional (y) directions within layer j ; $P_j = (p_j + \rho_j g z) / \rho_0$ is the Montgomery function, which depends only on the lateral position (x, y) within each layer; p_j being the true dynamic pressure. The vertical coordinate z is positive upwards. The Coriolis parameter is linear in y , $f(y) = f_0 + \beta y$. The density ρ_j is constant in layer j and ρ_0 is a reference density. If H_j is the depth of the base of layer j , then the hydrostatic assumption causes the Montgomery function in layer j to be related to that in layer $j + 1$ by

$$P_j = \gamma_j H_j + P_{j+1} \quad j = 1, 2 \quad (3.2)$$

in which

$$\gamma_j = g \frac{(\rho_{j+1} - \rho_j)}{\rho_0} \quad j = 1, 2 \quad (3.3)$$

are the reduced gravity parameters. If the layer thickness $h_j = H_{j-1} - H_j$ is introduced, the mass conservation equation for layer j becomes

$$\partial_t h_j + J(P_j, \frac{h_j}{f}) = \begin{cases} -w_E & \text{for } j=1 \\ 0 & \text{for } j=2,3 \end{cases} \quad (3.4)$$

in which J denotes the Jacobian and w_E is the wind-induced vertical velocity at the base of the surface Ekman layer. The sum of equations (3.4) gives the Sverdrup transport relation:

$$J(P_3, \frac{H_3}{f}) - \frac{\beta}{f^2} \partial_x \left(H_3 P_3 + \frac{1}{2} \gamma_1 H_1^2 + \frac{1}{2} \gamma_2 H_2^2 \right) = -w_E. \quad (3.5)$$

This may be integrated westward away from the eastern boundary, at which the condition of no normal flow and of geostrophy require that the H_j be independent of the alongshore coordinate.

Many steady solutions of these equations appear in the literature (e.g. LPS). An important feature is that as long as the deeper layers do not ever surface to be exposed to the wind, the steady flow is confined to the upper layer. In this paper we consider small amplitude perturbations of such a flow defined in a basin whose eastern boundary is $x = a$. In fact, since we restrict our attention to perturbations whose horizontal length scales are much smaller than the basin-wide scales of variation of the large scale circulation, we actually consider – and simply prescribe ab initio – spatially homogeneous steady upper layer flow u_1, v_1 which we assume to have been generated by solving these equations with a specified wind forcing, and then study time-dependent perturbations of that flow which are excited by a specified initial perturbation or by harmonic forcing turned on at some initial time.

Indicating perturbation quantities by superscript ⁽¹⁾ and steady quantities by superscript ⁽⁰⁾, the perturbed, linearized version of the mass conservation equation for the middle and the lower layer are readily found to be

$$\partial_t(H_2^{(1)} - H_1^{(1)}) + J \left(P_2^{(1)}, \frac{H_{20} - H_1^{(0)}}{f} \right) = 0 \quad (3.6)$$

$$\partial_t(-H_2^{(1)}) + J \left(P_3^{(1)}, \frac{H_3 - H_{20}}{f} \right) = 0 \quad (3.7)$$

where $H_2^{(0)} = H_{20} = \text{constant}$ is the depth of the bottom of the stagnant middle layer. The perturbation Sverdrup equation is

$$2H_3P_3^{(1)} + 2\gamma_1H_1^{(0)}H_1^{(1)} + 2\gamma_2H_{20}H_2^{(1)} = \Theta^{(1)} \quad (3.8)$$

where

$$\Theta^{(1)}(x, y, t) = -\frac{2f^2}{\beta} \int_x^a w_E^{(1)}(x', y, t) dx' \quad (3.9)$$

is the zonal integral of the time-dependent Ekman pumping from the eastern boundary.

We now reduce the model to two layers lying over an infinitely deep quiescent third layer, which gives the so-called two-and-a-half layer model. We solve the Sverdrup equation (3.8) for $P_3^{(1)}$, insert into (3.7) and let the total depth $H_3 \rightarrow \infty$, which gives

$$\partial_t H_2^{(1)} - \frac{\beta\gamma_2 H_{20}}{f^2} \partial_x H_2^{(1)} - \frac{\beta\gamma_1 \partial_x (H_1^{(0)} H_1^{(1)})}{f^2} = -\frac{\beta\Theta_x^{(1)}}{2f^2}. \quad (3.10)$$

Equations (3.6) and (3.10) determine the time-dependent perturbations $H_1^{(1)}$ and $H_2^{(1)}$ of nonzonal background flow.

3.3. The perturbation equations

Introduce the less cumbersome notation

$$(H_1^{(1)}, H_2^{(1)}) = (w, z). \quad (3.11)$$

The perturbation equations (3.6) and (3.10) for the two sets of long baroclinic planetary waves then become

$$z_t + J(z, \Psi) = w_t \quad z_t - cz_x - (Cw)_x = \tilde{\Theta}, \quad (3.12)$$

where for notational simplicity we have introduced the abbreviations

$$C(x, y) = \frac{\beta\gamma_1 H_1^{(0)}}{f^2} \quad c(y) = \frac{\beta\gamma_2 H_{20}}{f^2}. \quad (3.13)$$

Ψ is proportional to the potential thickness of the middle layer

$$\Psi(x, y) = \frac{\gamma_2(H_{20} - H_1^{(0)})}{f}, \quad (3.14)$$

and $\tilde{\Theta}$ is the forcing derived from the time-dependent Ekman pumping

$$\tilde{\Theta}(x, y, t) = -\frac{\beta\partial_x\Theta^{(0)}}{2f^2}e^{-i\sigma t}, \quad t > 0. \quad (3.15)$$

$\Theta^{(0)}$ is the steady counterpart of (3.9). U and V are gradients of potential thickness of the middle layer, given by:

$$U = \Psi_y = \frac{\gamma_2}{\gamma_1}(u_g + c_R) \quad V = -\Psi_x = \frac{\gamma_2}{\gamma_1}v_g, \quad (3.16)$$

where u_g, v_g are the mean geostrophic velocity components and $c_R = \frac{\gamma_2}{\gamma_1}C - c$, a Rossby-wave speed.

3.4. Plane wave solutions

Consider the waves that the system (3.12) supports. Eliminate w between these two equations to get a second order differential equation for z in which forcing is momentarily omitted:

$$CUz_{xx} + CVz_{xy} + (c + C)z_{xt} - z_{tt} = 0. \quad (3.17)$$

This is a hyperbolic equation in spatial variables x and y . To perform local stability analysis of the perturbations, we consider a point in the basin, obtain the solution for the steady flow at that position, assume that the flow is spatially homogeneous everywhere in the basin, and find plane wave solutions of the equation above.

For plane waves of the form $z \sim e^{i(kx+ly-\sigma t)}$, the corresponding dispersion relation is

$$\sigma^2 + \sigma(c + C)k - CUk^2 - CVkl = 0. \quad (3.18)$$

This is quadratic in frequency σ and zonal wavenumber vector component k , and linear in the meridional wavenumber vector component l . Coefficients C , c , U and V are completely determined by the background circulation and considered constant. By setting

$$k = \kappa \cos \alpha \quad l = \kappa \sin \alpha, \quad (3.19)$$

it follows from (3.18) that the zonal phase velocity $\frac{\sigma}{k}$ depends only on the wavenumber vector orientation α and not its magnitude κ .

Figure 3.1 shows a dispersion relation diagram obtained by fixing the frequency σ , varying the meridional wavenumber component l (real) and solving the resulting quadratic equation (3.18) for the zonal wavenumber k (complex in general). The parameters of the stratification and the circulation are those appropriate to the

southeastern part of the subtropical gyre where the background flow is towards the southwest. The frequency is 1 cpy. The upper panel is the dispersion relation diagram for the case of no background circulation ($u_g = v_g = 0$), showing the two nondispersive baroclinic Rossby modes of this two-and-a-half layer system (dashed lines). Even relatively weak background flow completely changes the nature of the solution (lower panel): there is no longer a simple separation into the first and the second baroclinic mode as the two sets of waves become coupled by the nonzonal background flow. As the meridional component of the background circulation becomes different from zero, the long wave branch of the dispersion relation evolves from the fast Rossby wave and the much shorter wave branch of the dispersion relation evolves from the slow Rossby wave (dashed lines). Although relative vorticity has been neglected (the long wave approximation has been made) so that both sets of waves are in that sense long planetary waves, since their wavelengths are so very different, we refer to them for convenience as the long and the short waves. The long wave is the first internal baroclinic mode in the absence of the background flow. The short wave is strongly refracted in the background circulation and its wavelengths become much shorter than the wavelengths of the corresponding second internal baroclinic mode in the absence of the background flow.

Additionally, the presence of the background circulation gives rise to a region in wavenumber space where the fixed-frequency annual-period waves have unstable solutions which grow zonally (lower panel, solid line). The unstable solutions appear where the stable branches approach most closely. Even a weak meridional background flow can make the waves unstable as it makes the zonal perturbations (not stabilized by the effects of β) more capable of releasing potential energy of the background flow (Pedlosky, 1987).

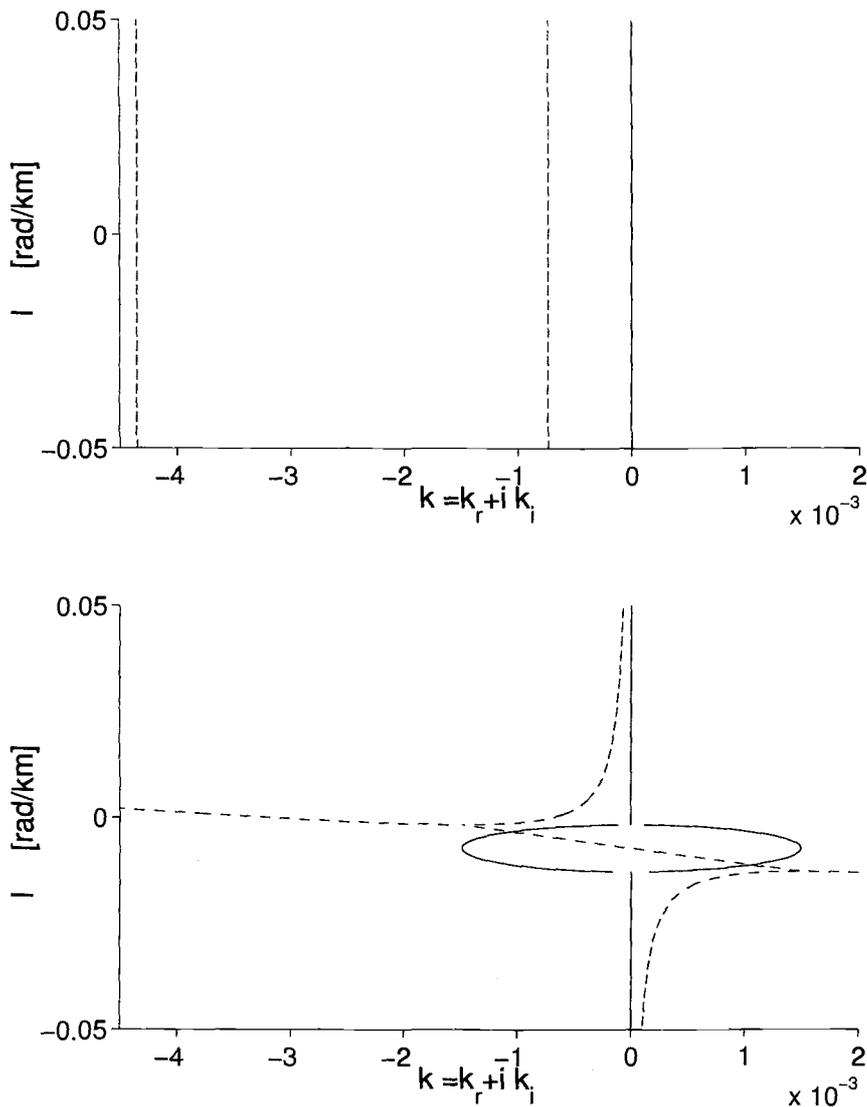


FIGURE 3.1. The dispersion relation diagram obtained by fixing the frequency σ , varying the meridional wavenumber component l (real) and solving the quadratic equation (3.18) for the zonal wavenumber k (complex in general). The upper panel shows the two nondispersive long baroclinic planetary waves of the two-and-a-half layer system (dashed line) for the case of no background geostrophic flow. The lower panel shows that the presence of even relatively weak background flow completely changes the dispersion relation: the modes are coupled and unstable (the imaginary part of k is plotted as a solid line). The parameters describing the background flow are $H_{10}=350$ m, $H_{20}=700$ m, $\gamma_1 = 0.0245$ m s $^{-2}$, $\gamma_2 = 0.01$ m s $^{-2}$. The point considered has the coordinates $x=9000$ km, $y=2500$ km in the basin with the zonal extent of 10000 km. The amplitude of the Ekman pumping is $4.7 \cdot 10^{-7}$ m s $^{-1}$. The frequency is 1 cpy.

3.5. The initial value problem: a transient solution

Consider a disturbance generated by localised forcing that starts at some initial time and thereafter oscillates periodically. The resulting solution consists of a transient part which satisfies a homogeneous initial value problem (IVP), and a sympathetic part that oscillates at the frequency of the forcing. Consider the transient solution first.

Outline

(i) For simplicity first analyse the IVP with the initial conditions of the undisturbed upper layer interface (w) and a Gaussian shaped middle layer interface displacement (z):

$$w(x, y, 0) = 0 \qquad z(x, y, 0) = e^{-\frac{(x^2+y^2)}{L^2}}, \qquad (3.20)$$

where L is the width of the Gaussian. The governing equations are (3.12), with the forcing omitted. The IVP is solved in a doubly periodic domain by numerically summing Fourier components (subsections 3.5.1.1 and 3.5.1.2 below).

(ii) We further consider the initial wave group which is a product of the Gaussian envelope and a plane wave of a specified wavelength:

$$w(x, y, 0) = 0 \qquad z(x, y, 0) = e^{-\frac{(x^2+y^2)}{L^2}} e^{i(k_0x+l_0y)}, \qquad (3.21)$$

where k_0 and l_0 are specified wavenumber components. This enables us to put energy into the wavenumbers of our choice (subsection 3.5.1.3 below).

The results obtained by numerical summation of Fourier components will reveal a number of limitations associated with this method, which provides motivation to work out the analytical solution of the problem (i).

(iii) We thus solve (i) analytically to obtain the asymptotic solution for large time using the method of steepest descents (subsection 3.6).

3.5.1. The transient solution obtained by numerically summing Fourier components

3.5.1.1. Gaussian disturbance centered at $k = 0, l = 0$; no background flow

Figure 3.2 shows the dispersion relation (3.18) for the same stratification parameters as those of Figure 3.1, with no background flow (upper panel). The dispersion relation is displayed in an alternative way: frequency is contoured vs. k and l . In the IVP we specify an initial condition in the wavenumber plane and at every point in the wavenumber plane solve the quadratic dispersion relation (3.18) for the two values of frequency which we denote by σ^- and σ^+ . In the case of no background flow these two solutions σ^- and σ^+ are the classical first and the second baroclinic mode, respectively. In order to see readily how each of them evolves when the background circulation is introduced, we display the two parts of the solution separately (in which case each component individually satisfies only the initial condition that the middle layer interface displacement is a Gaussian shaped disturbance, the second of (3.20)).

In all the cases where the solution is obtained numerically, the method of solution is to advance each Fourier component in time with an initial amplitude and phase derived from the initial condition, and with a frequency determined from the dispersion relation. The basin considered is 2000 km by 2000 km.

Figure 3.2 for σ^- shows that the westward phase speed is independent of k (left panel). The imaginary part of σ^- is zero since the waves are stable in the absence of the background circulation (right panel). Dashed lines are contours of the amplitude of the initial condition in the wavenumber plane. The initial disturbance is a Gaussian centered at $k = 0, l = 0$ with the width of 94 km and the amplitude

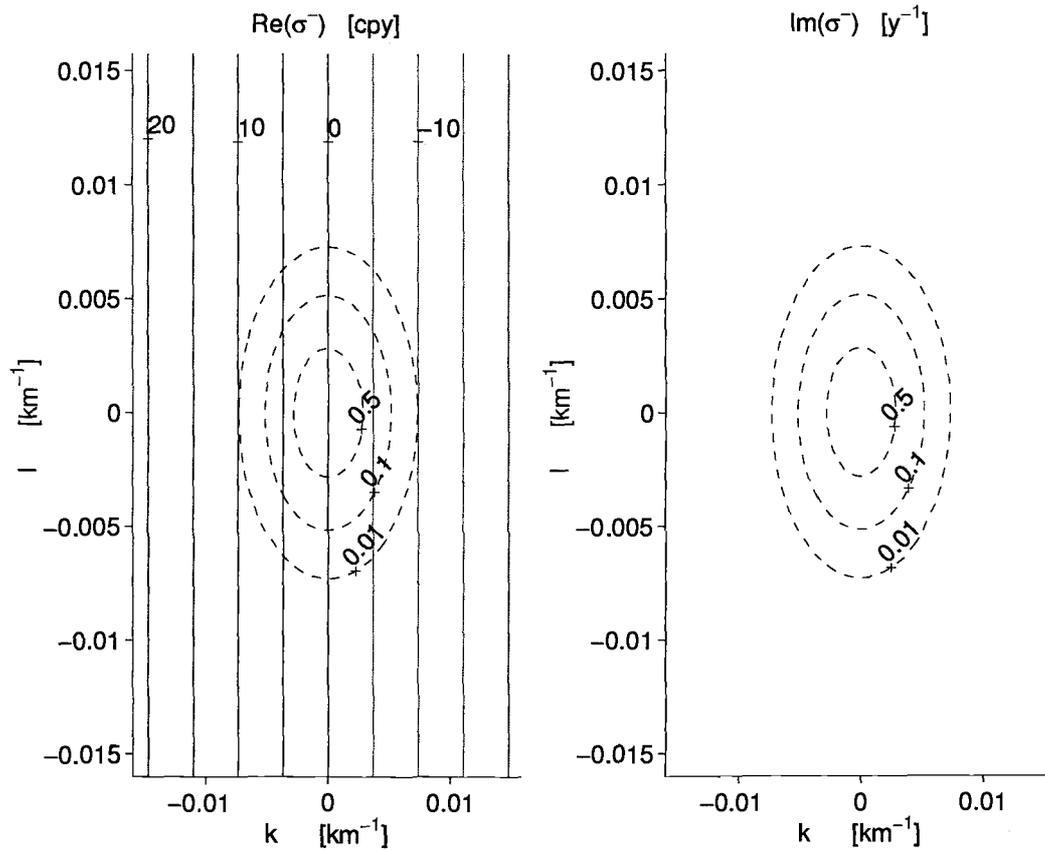


FIGURE 3.2. The dispersion relation (3.18): contours of frequency vs. k and l for the σ^- solution in the absence of the background flow. The westward phase speed is independent of k (left panel) and the waves are stable (imaginary part of σ is zero, right panel). Dashed lines are contours of the amplitude of the initial condition in the wavenumber plane. The initial Gaussian disturbance is centered at $k = 0$, $l = 0$ and has the width of 94 km.

1 m. The dispersion relation for σ^+ (not shown), which corresponds to the second baroclinic mode, looks very similar, except that the phase speeds are slower.

Figure 3.3 shows a snap-shot of the time-dependent interface displacement of the middle layer for the σ^- part of the solution after 150 days (upper panel). The initially symmetric disturbance has retained its shape since the long planetary waves in the case of no background circulation are nondispersive. The entire disturbance has propagated westward. This is the classical first baroclinic Rossby wave. The lower panel shows the analogous plot for the σ^+ part of the solution; this wave is slower and it corresponds to the classical second baroclinic Rossby wave.

*3.5.1.2. Gaussian disturbance centered at $k = 0, l = 0$;
spatially homogeneous background flow towards SW*

We next consider the case of spatially homogeneous background flow towards SW, and compare it to the case of no background circulation.

Figure 3.4 shows the dispersion relation plot (displayed as in Figure 3.2) for the σ^+ part of the solution: the most prominent difference is that the imaginary part of σ , which represents the temporal growth, is no longer zero (right panel). Dashed lines again show the contours of the amplitude of the initial condition in the wavenumber plane, which is a Gaussian disturbance centered at $k = 0, l = 0$. Energy has been put into Fourier components that are associated with the temporal growth and decay so the corresponding parts of the solution will grow and decay in time. The other root of the dispersion relation σ^- is the complex conjugate of this one.

The dispersion relation for the σ^+ part of the solution shows that the zonal component of the group velocity $\frac{\partial \sigma}{\partial k} |_l$ is positive where $\frac{\partial l}{\partial k} |_\sigma$ is positive, a region which forms a narrow radial wedge on Figure 3.4, (left panel).

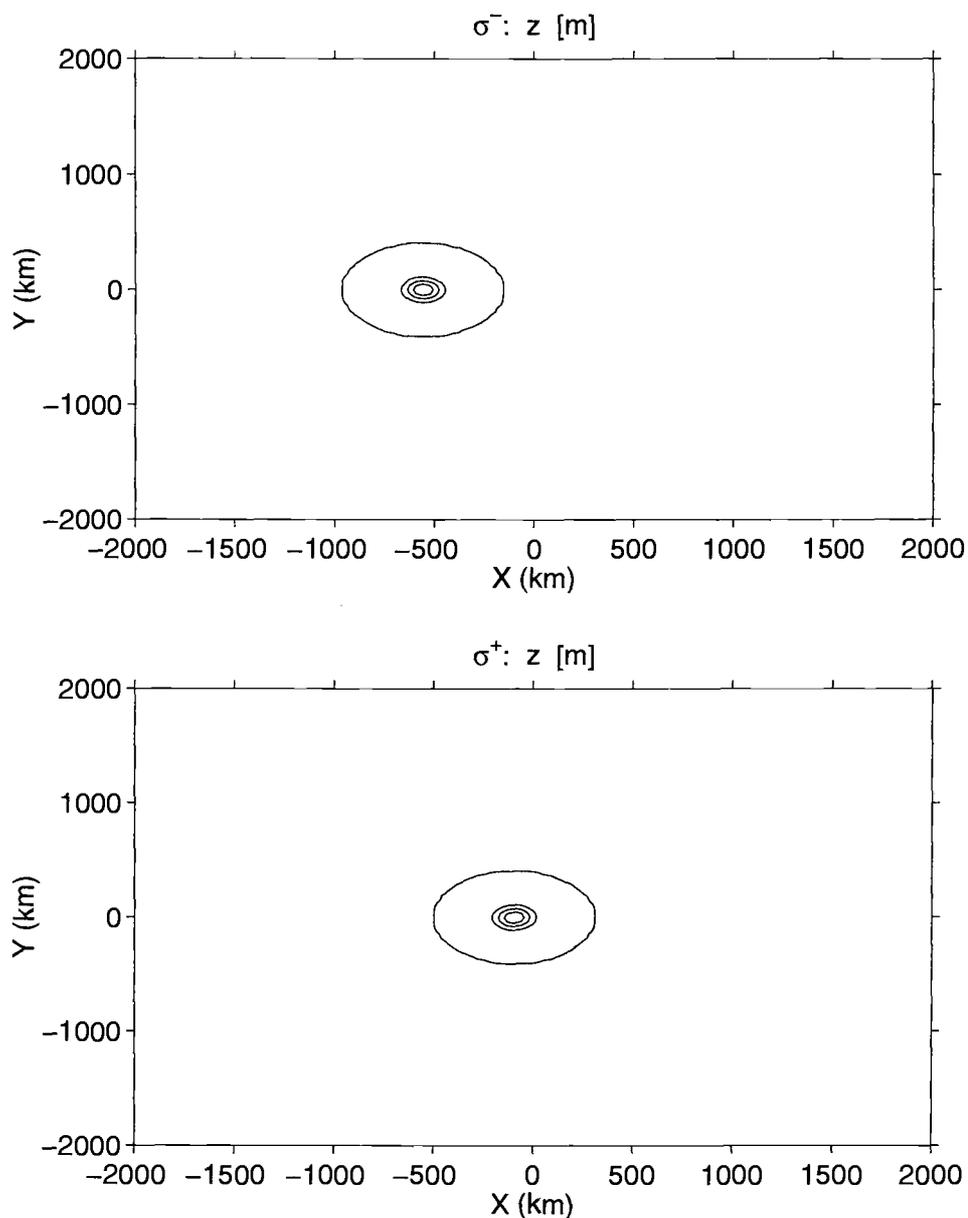


FIGURE 3.3. A snap-shot of the time-dependent interface displacement of the middle layer (z in m) for the σ^- and σ^+ part of the solution after 150 days (upper and lower panel, respectively). The disturbance has propagated westward. Initially symmetric disturbance has retained its shape since the long planetary waves in the case of no background circulation are nondispersive. The σ^- solution is the classical first baroclinic Rossby wave and the σ^+ solution is the classical second baroclinic Rossby wave. Contour levels are from 0 to 1 m, with the interval 0.25 m.

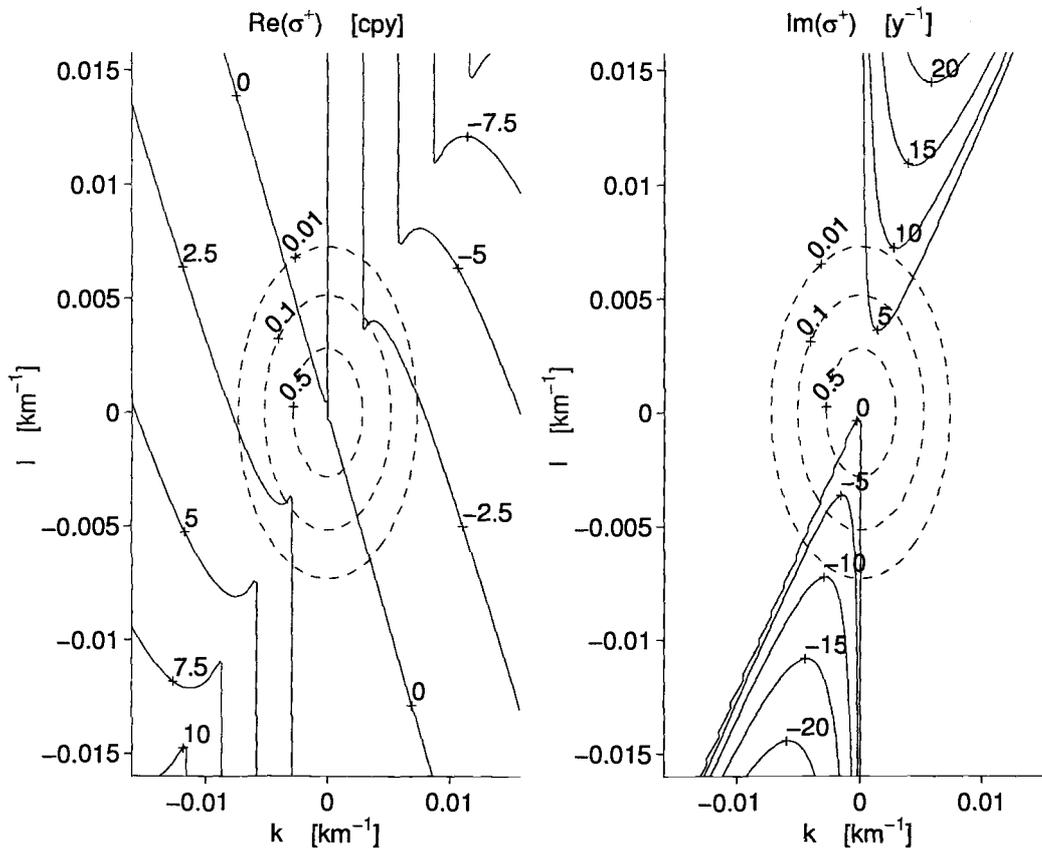


FIGURE 3.4. The dispersion relation diagram for the σ^+ part of the solution in the presence of the SW background flow: contours of σ vs. k and l . The zonal component of the group velocity $\frac{\partial \sigma}{\partial k} |_l$ is positive where $\frac{\partial l}{\partial k} |_\sigma$ is positive, a region which forms a narrow radial wedge (left panel). Energy has been put into Fourier components that are associated with the temporal growth and decay (right panel). Dashed lines show the contours of the amplitude of the initial Gaussian disturbance centered at $k = 0$, $l = 0$, with the width of 94 km.

Figure 3.5 is a snap shot of the time-dependent interface displacement of the middle layer as a function of x and y after 150 days. The upper panel shows the σ^- part of the solution, which would correspond to the first baroclinic Rossby mode in the case of no background flow. Due to the presence of the background circulation, the initially symmetric Gaussian disturbance has acquired very anisotropic shape. The largest amplitude is at that position in the basin where the first baroclinic mode was centered in the case of no background flow. Elsewhere, the wave amplitude is different from zero along certain directions which are completely determined by the background flow i.e. by the gradient of the background potential vorticity. From the analytic expression for the asymptotic solution we obtain the angle that these rays form with the zonal direction to be

$$\arctan\left(\frac{\frac{U}{C} + \frac{1}{4}\left(1 + \frac{c}{C}\right)^2}{\frac{V}{C}}\right) \quad (3.22)$$

where the coefficients are defined in (3.13) and (3.16). The amplitude of the solution has grown in time, as the dispersion relation suggested. The overall orientation of the alternating pattern of positive (solid line) and negative (dashed line) displacements corresponds to the orientation of crests and troughs of the unstable waves in Figure 3.4. At later times the orientation remains the same, but the separation between alternating positive and negative displacements becomes shorter as shorter waves have higher temporal growth. The lower panel shows the analogous figure for σ^+ part of the solution, which corresponds to the second baroclinic mode in the case of no background flow. Again the largest amplitude is at that location in the basin where the second baroclinic mode would arrive in the case of no background flow. As in the upper panel, preferred directions along which the amplitude is different from zero are determined by (3.22).

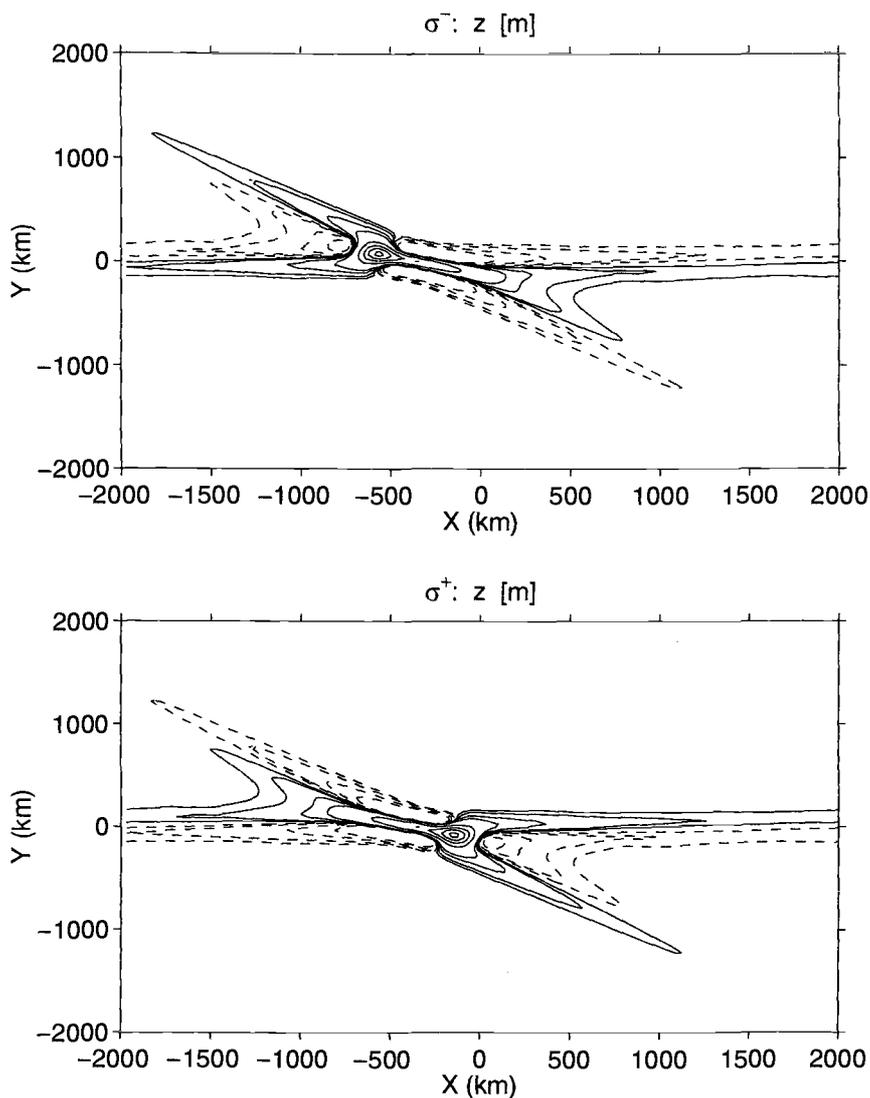


FIGURE 3.5. A snap shot of the time-dependent interface displacement of the middle layer (z in m) as a function of x and y after 150 days in the presence of the SW background flow: σ^- (upper panel) and σ^+ (lower panel) parts of the solution correspond to the first and the second baroclinic Rossby mode, respectively, in the case of no background flow; the largest amplitude is at that position in the basin where the first (second) baroclinic mode was centered in the case of no background circulation. Due to the presence of the background flow initially symmetric Gaussian disturbance has acquired very anisotropic shape. The wave amplitude is different from zero along the directions completely determined by the gradient of the background potential vorticity. Positive (solid line) and negative (dashed line) displacements have the orientation of the fastest growing waves in Figure 3.4. Contour levels are 0.5 cm, 1 cm, 2.5 cm, 5 cm, 0.25 m, 0.5 m, 0.75 m and 1m.

Figure 3.6 shows the sum of the two components σ^+ and σ^- of the solution of the initial value problem. This sum satisfies both of the initial conditions (3.20). Notice that the diagonal ray-like features noted above in each of the two components σ^+ and σ^- are no longer present in the sum.

*3.5.1.3. Gaussian disturbance centered at k_0, l_0 ;
spatially homogeneous background flow towards SW*

We next use the same method to solve the IVP with an initial condition which is a plane wave $e^{i(k_0x+l_0y-\sigma t)}$ modulated by a Gaussian envelope. In Fourier space it is a Gaussian disturbance centered at specified k_0, l_0 . If k_0 and l_0 are such that the dispersion relation gives a frequency that has a large imaginary component, then one part of the solution grows much more rapidly than the other.

Figure 3.7 shows the dispersion relation plot: the contours of the real part of the frequency (left panel) and the contours of the imaginary part of the frequency (right panel) for σ^+ part of the solution. The initial condition is a wave packet with central meridional wavelength 125 km ($k_0=0, l_0=0.8 \text{ km}^{-1}$) modulated by a radially symmetric Gaussian with the width of 94 km. The σ^+ part of the solution grows in time and the σ^- part of the solution decays in time for these wavenumbers. Figure 3.8 shows a zoom of the snapshot of the two separate components of the time-dependent interface displacement of the middle layer (z) as a function of x and y after 75 days for the parameters of Figure 3.7. As was the case in Figure 3.5 for similar initial condition, the two components propagate westward at different speeds. As Figure 3.7 suggests, one component (σ^+) grows westward and one (σ^-) decays. With this choice of parameters, the shortest resolved wavelengths are not strongly energized, so that the spatial shape of the disturbance is well represented, but if the solution is run longer or the initial wavelength is shorter, the shortest

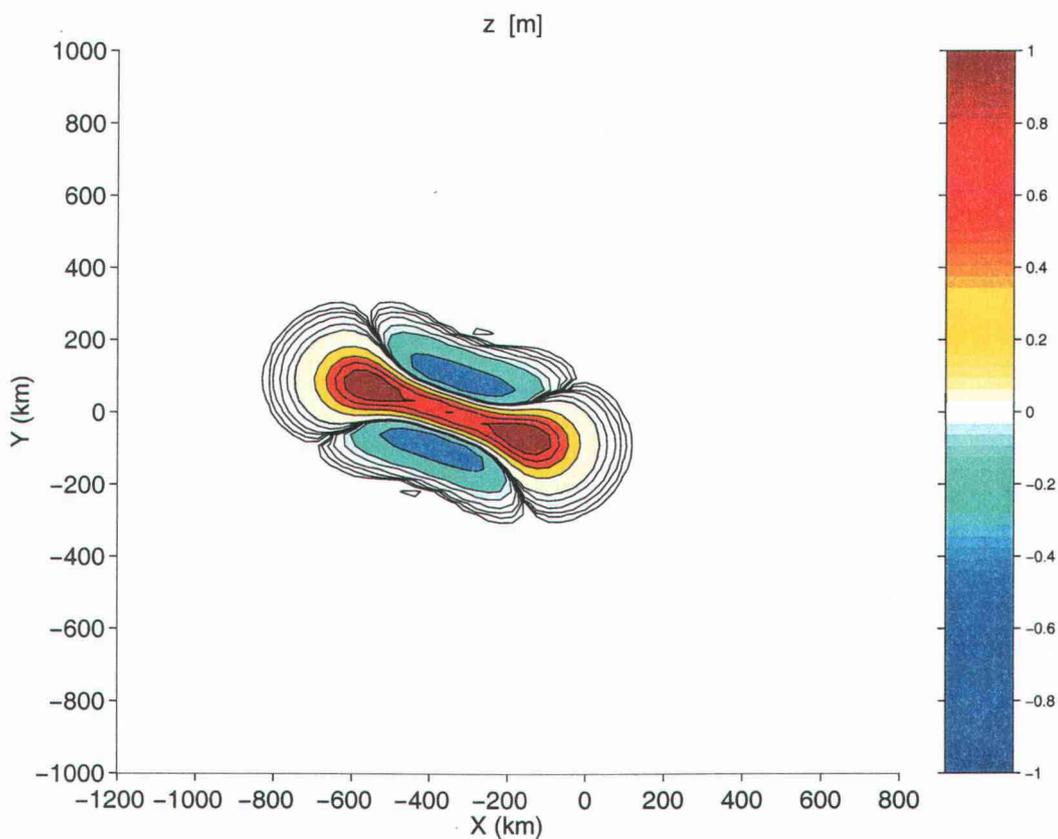


FIGURE 3.6. A zoom of the sum of the σ^+ and σ^- components of the time-dependent interface displacement of the middle layer (z in m) shown individually in Figure 3.5. Contour levels are the same as in Figure 3.5 with additional levels 0.25 cm and 0.125 cm.

resolved wavelengths are strongly energized. Because immediately shorter wavelengths are however not resolved and so cannot interfere correctly with the shortest resolved wavelengths, the spatial shape of the disturbance is poorly represented and is dominated by spurious short wavelength features. The maximum amplitude of the decaying (σ^+) part of the solution is 3.3 m and that of the growing (σ^-) part of the solution is 0.5 m; for both the initial maximum amplitude was 1.0 m.

Figure 3.8 illustrates the limitations of the method of numerically summing Fourier components, which are: 1) lack of fine resolution 2) inability to go to long times because the solution reenters the periodic domain 3) no analytical representation of the solution, so that we cannot readily identify some features of the solution such as caustics.

3.6. Transient solution obtained by the method of steepest descents with modifications

3.6.1. Introduction

We seek to solve the initial value problem asymptotically for large time using the method of steepest descents. The classical asymptotic solution of the IVP as $t \rightarrow \infty$ is given by a sum of terms each one of the form

$$z \propto \frac{e^{i(k_0 \frac{x}{t} + l_0 \frac{y}{t} - \sigma_0)t}}{t[\det(\frac{\partial^2 \Omega}{\partial k_i \partial k_j})]^{1/2}} \quad (3.23)$$

where $\sigma_0 = \Omega(k_0, l_0)$ denotes the dispersion relation (3.18) and $\nabla_k \Omega_0 = (\frac{x}{t}, \frac{y}{t})$ defines the stationary phase points (e.g. Carrier et al., 1966). But it can easily be shown from (3.18) that $\det(\frac{\partial^2 \Omega}{\partial k_i \partial k_j}) = 0$ everywhere in wavenumber space and the form (3.23) fails. (This comes about because the long wave assumption has been made.) Nevertheless, there is an orientation of coordinates that makes the matrix $(\frac{\partial^2 \Omega}{\partial k_i \partial k_j})$

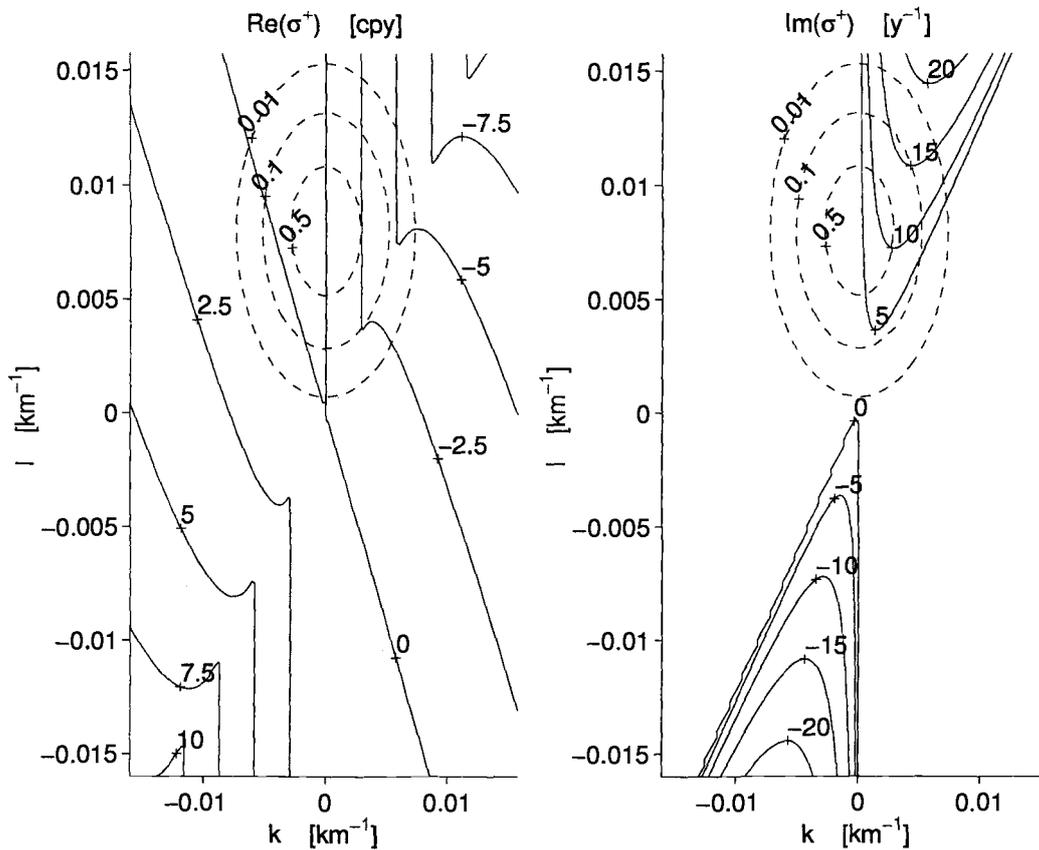


FIGURE 3.7. The dispersion relation plot: contours of the real part of the frequency (left panel) and contours of the imaginary part of the frequency (right panel) for σ^+ part of the solution. Dashed lines show the contours of the amplitude of the initial disturbance which is a group with central wavelength 125 km modulated by a radially symmetric Gaussian of width of 94 km. σ^+ part of the solution grows in time.

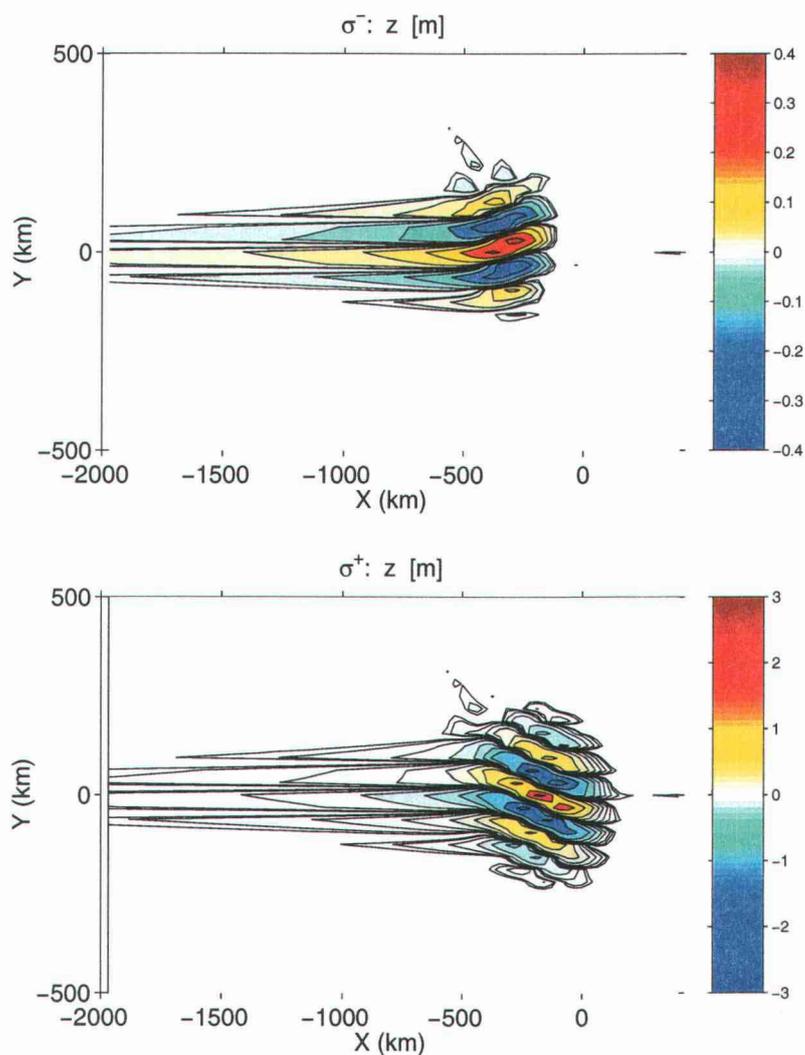


FIGURE 3.8. A zoom of the snapshot of the two separate components of the time-dependent interface displacement of the middle layer (z in m) as a function of x and y after 75 days for the parameters of Figure 3.7. The two components propagate westward at different speeds; σ^+ grows westward and σ^- decays. Contour intervals range from .0085 m to 3 m. In both panels the signal in the easternmost part of the basin (with the amplitude .0085 m) is caused by wrap-around since the solution reenters the periodic domain.

diagonal and one of the eigenvalues is nonzero; we thus integrate in the direction of the nonzero eigenvalue.

Therefore we transform into cylindrical polar coordinates in the wavenumber space using (3.19); the wavenumber vector components k and l are then described in terms of the magnitude of the wavenumber vector κ and the angle α that the wavenumber vector makes with the zonal direction. In cylindrical polar coordinates the effects of κ and α are completely separated. We first perform the integration over the angle α by the method of steepest descents. After that, for a Gaussian initial disturbance we perform the integration over the wavenumber magnitude κ by the method of steepest descents. The result of the latter integration thus depends on the shape of the initial disturbance.

3.6.2. Transient solution for the Gaussian shaped initial disturbance

We analytically solve the problem (i) described at the beginning of section 3.5. The governing equations (3.12) for the time-dependent interface displacements w and z of the upper and middle layer can be written as

$$z_t + Uz_x + Vz_y = w_t \quad (3.24)$$

$$-z_t + cz_x + Cw_x = 0 \quad (3.25)$$

where all the coefficients (3.13) and (3.16) are constant. The initial conditions are the Gaussian displacement of the interface between the middle and the lowest layer and the interface between the upper and the middle layer is assumed to be flat:

$$z_0(x, y, 0) = e^{-\frac{x^2+y^2}{L^2}} \quad w_0(x, y, 0) = 0 \quad (3.26)$$

where L is the width of a Gaussian disturbance.

To solve the initial value problem, it is convenient to Fourier transform equations (3.24) and (3.25) in space, and Laplace transform in time as

$$\begin{aligned}\bar{z}(k, l, s) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^{\infty} z(x, y, t) e^{-i(kx+ly)} dx dy e^{-st} dt \\ z(x, y, t) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{c-i\infty}^{c+i\infty} \bar{z}(k, l, s) e^{i(kx+ly)} \frac{dk}{2\pi} \frac{dl}{2\pi} \frac{e^{st}}{2\pi i} ds.\end{aligned}\quad (3.27)$$

Here $\bar{z}(k, l, s)$ denotes the Fourier-Laplace transform of $z(x, y, t)$; there are analogous expressions for w . Fourier transforms of the initial conditions z_0 and w_0 are denoted by \tilde{z}_0 and \tilde{w}_0 . With $s = -i\sigma$, the Fourier transformed governing equations (3.24) and (3.25) can be written as

$$\begin{aligned}\begin{pmatrix} \bar{z}(k, l, s) \\ \bar{w}(k, l, s) \end{pmatrix} &= \frac{i}{\sigma^2 - CUk^2 - CVkl + \sigma(c + C)k} \\ &\begin{pmatrix} kC & -\sigma \\ -\sigma - ck & -\sigma + kU + lV \end{pmatrix} \begin{pmatrix} \tilde{z}_0 - \tilde{w}_0 \\ -\tilde{z}_0 \end{pmatrix}.\end{aligned}\quad (3.28)$$

The expression in the denominator is the dispersion relation (3.18). Interface displacements z and w can be obtained by inverse Fourier transformation of equations (3.28)

$$\begin{aligned}\begin{pmatrix} z(x, y, t) \\ w(x, y, t) \end{pmatrix} &= -iL^2\pi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty+ic}^{\infty+ic} \frac{e^{-\frac{(k^2+l^2)L^2}{4}} e^{i(kx+ly-i\sigma t)}}{\sigma^2 + \sigma(c + C)k - CUk^2 - CVkl} \\ &\begin{pmatrix} kC + \sigma \\ -ck - kU - lV \end{pmatrix} \frac{dk}{2\pi} \frac{dl}{2\pi} \frac{d\sigma}{2\pi}.\end{aligned}\quad (3.29)$$

Using cylindrical polar coordinates as introduced in (3.19), integral (3.29) can be written as:

$$\begin{aligned}\begin{pmatrix} z(x, y, t) \\ w(x, y, t) \end{pmatrix} &= -iL^2\pi \int_0^{\infty} \int_0^{2\pi} \int_{-\infty+ic}^{\infty+ic} C\kappa \cos\alpha e^{-\frac{\kappa^2 L^2}{4}} e^{i\kappa C t \cos\alpha [\xi + \eta\theta - H]} \\ &\begin{pmatrix} 1 + H - \frac{a}{2} \\ -\frac{c}{C} - u - v\theta \end{pmatrix} \frac{1}{[H^2 - (\frac{a^2}{4} + u + v\theta)]} \kappa \frac{d\kappa}{2\pi} \frac{d\alpha}{2\pi} \frac{d\sigma}{2\pi}\end{aligned}\quad (3.30)$$

where we have introduced a scaled and shifted x coordinate and a scaled y coordinate as

$$\xi = \frac{x}{Ct} + \frac{a}{2} \quad \eta = \frac{y}{Ct}, \quad (3.31)$$

and the following abbreviations

$$H = \frac{\sigma}{C\kappa \cos \alpha} + \frac{a}{2}, \quad (3.32)$$

$$u = \frac{U}{C} \quad v = \frac{V}{C}, \quad (3.33)$$

$$a = 1 + \frac{c}{C} \quad \theta = \tan \alpha. \quad (3.34)$$

The σ integral is evaluated using the residue theorem. By $H^+(\alpha)$ and $H^-(\alpha)$ we denote the positive and negative root of the quadratic dispersion relation in the denominator. The latter can be compactly written as

$$H^2 = b + v\theta \quad (3.35)$$

with

$$b = \frac{a^2}{4} + u. \quad (3.36)$$

Equations (3.30) then become

$$\begin{aligned} \begin{pmatrix} z(x, y, t) \\ w(x, y, t) \end{pmatrix} &= \frac{L^2 \pi}{2} \int_0^\infty \int_0^{2\pi} e^{-\frac{\kappa^2 L^2}{4}} \left\{ \frac{e^{i\kappa C t \cos \alpha [\xi + \eta \theta - H^+]}}{H^+} \begin{pmatrix} 1 + H^+ - \frac{a}{2} \\ -\frac{c}{C} - u - v\theta \end{pmatrix} \right. \\ &\quad \left. + \frac{e^{i\kappa C t \cos \alpha [\xi + \eta \theta - H^-]}}{H^-} \begin{pmatrix} 1 + H^- - \frac{a}{2} \\ -\frac{c}{C} - u - v\theta \end{pmatrix} \right\} \frac{\kappa d\kappa}{2\pi} \frac{d\alpha}{2\pi}. \end{aligned} \quad (3.37)$$

In cylindrical polar coordinates the effects of κ and α are completely separated. We first perform the integration in α , using the method of steepest descents. The α integrals in (3.37) are of the form

$$I = \int_0^{2\pi} e^{i\kappa Ct[\xi \cos \alpha + \eta \sin \alpha - H \cos \alpha]} F(\alpha) d\alpha. \quad (3.38)$$

To integrate (3.38) by the method of steepest descents, Taylor expand the phase

$$\Phi = \xi \cos \alpha + \eta \sin \alpha - H \cos \alpha \quad (3.39)$$

around the angle α_0 , to obtain

$$\begin{aligned} \Phi = \xi \cos \alpha_0 + \eta \sin \alpha_0 - H_0 \cos \alpha_0 + (\alpha - \alpha_0) [-\xi \sin \alpha_0 + \eta \cos \alpha_0 - (H \cos \alpha)'_0] \\ + \frac{1}{2} (\alpha - \alpha_0)^2 [-\xi \cos \alpha_0 - \eta \sin \alpha_0 - (H \cos \alpha)''_0] + \dots \end{aligned} \quad (3.40)$$

where $(H \cos \alpha)'_0 = (H \cos \alpha)'(\alpha_0)$, $(H \cos \alpha)''_0 = (H \cos \alpha)''(\alpha_0)$. The angle α_0 is chosen to make the phase stationary,

$$\Phi'(\alpha_0) = -\xi \sin \alpha_0 + \eta \cos \alpha_0 - (H \cos \alpha)'_0 = 0. \quad (3.41)$$

We obtain $(H \cos \alpha)'_0$ from the dispersion relation (3.18) or (3.35), which can be written in the form

$$\gamma^2 + \gamma a \cos \alpha - u \cos^2 \alpha - v \sin \alpha \cos \alpha = 0 \quad (3.42)$$

where

$$\gamma(\alpha) \equiv \frac{\sigma}{C\kappa} = H \cos \alpha - \frac{a}{2}. \quad (3.43)$$

The dispersion relation (3.42) shows that $\frac{\sigma}{\kappa}$ is only a function of the direction of the wavenumber vector α , and not its magnitude (similar to the case of the internal gravity waves where σ is only a function of the direction α). The stationary phase requirement for α_0 thus becomes

$$(H \cos \alpha)' = \gamma' = \frac{\gamma a \sin \alpha - 2u \sin \alpha \cos \alpha + v(\cos^2 \alpha - \sin^2 \alpha)}{2\gamma + a \cos \alpha} = -\xi \sin \alpha + \eta \cos \alpha. \quad (3.44)$$

(Which is the same as (3.41)). Rewrite the equation above in terms of variables (3.32), (3.33) and (3.34) to obtain

$$(2H\xi - v\theta - 2b)\theta + v - 2H\eta = 0. \quad (3.45)$$

Eliminate θ between (3.45) and the dispersion relation (3.35) to get a fourth order polynomial in H

$$H^4 - 2H^3\xi + 2H(b\xi + v\eta) - (b^2 + v^2) = 0 \quad (3.46)$$

whose solutions give the four values of the angle α_0 that make the phase stationary. Once the values of H , hence α_0 have been obtained by solving (3.46), integral (3.38) can be written as

$$I(\alpha_0) = e^{i\kappa Ct\Phi(\alpha_0)} \int_0^{2\pi} e^{-\frac{1}{2}i\kappa Ct(\alpha - \alpha_0)^2 \left[\frac{x}{Ct} \cos \alpha_0 + \frac{y}{Ct} \sin \alpha_0 + \gamma_0'' \right]} F(\alpha_0) d\alpha. \quad (3.47)$$

For large times the remaining integral is Gaussian over the infinite interval, so the final solution of (3.38) is

$$I(\alpha_0) = \frac{\sqrt{2\pi}}{\sqrt{i\kappa Ct \left[\frac{x}{Ct} \cos \alpha_0 + \frac{y}{Ct} \sin \alpha_0 + \gamma_0'' \right]}} e^{i\kappa Ct \left[\frac{x}{Ct} \cos \alpha_0 + \frac{y}{Ct} \sin \alpha_0 - \gamma_0 \right]} F(\alpha_0). \quad (3.48)$$

We next perform κ integration in expression (3.37) using again the method of steepest descents. Therefore consider

$$J = \int_0^{\infty} e^{-\frac{\kappa^2 L^2}{4}} e^{i\kappa Ct\Phi(\alpha_0)} I(\alpha_0) \kappa \frac{d\kappa}{2\pi}. \quad (3.49)$$

Denote the exponent in (3.49) by

$$W(\kappa) = -\frac{\kappa^2 L^2}{4} + i\kappa Ct\Phi_0$$

where $\Phi_0 = \Phi(\alpha_0)$ is independent of κ . Taylor expand $W(\kappa)$ around κ_0 to obtain

$$W(\kappa) = -\frac{\kappa_0^2 L^2}{4} + i\kappa_0 Ct\Phi_0 + (\kappa - \kappa_0) \left(-\frac{\kappa_0 L^2}{2} + iCt\Phi_0 \right) - \frac{L^2}{4} (\kappa - \kappa_0)^2.$$

The choice

$$\kappa_0 = \frac{2itC\Phi_0}{L^2} \quad (3.50)$$

makes the phase stationary in (3.49) and the integral becomes

$$J = \frac{2\sqrt{\pi}}{L} \kappa_0 e^{-\frac{c^2 t^2 \Phi_0^2}{L^2}}. \quad (3.51)$$

In performing κ integration by the method of steepest descents, the result depends on the shape of the initial disturbance (which needs to be described by a twice differentiable function). Using the solutions (3.48) and (3.51), the expressions (3.37) for the time-dependent interface displacements z and w of the middle layer and the upper layer become

$$\begin{pmatrix} z(x, y, t) \\ w(x, y, t) \end{pmatrix} = \sum_{\alpha_0} \frac{L\sqrt{2}}{4} \kappa_0 \frac{e^{-\frac{c^2 t^2 \Phi_0^2}{L^2}}}{\sqrt{b + v\theta_0}} \frac{1}{\sqrt{i\kappa_0 C t \left[\frac{x}{Ct} \cos \alpha_0 + \frac{y}{t} \sin \alpha_0 + \gamma_0'' \right]}} \begin{pmatrix} 1 + \sqrt{b + v\theta_0} - \frac{a}{2} \\ -\frac{c}{C} - u - v\theta_0 \end{pmatrix} \quad (3.52)$$

where $\theta_0 = \tan(\alpha_0)$. To obtain the final solution we sum the contributions from all four angles α_0 that make the phase stationary and which have been obtained as the solutions to (3.46).

Figure 3.9 shows the time-dependent interface displacement of the middle layer plotted in moving coordinates ξ and η . This solution is analogous to the solution shown in Figure 3.6, which was obtained by numerically summing Fourier components. One large amplitude region is associated with the faster wave and the other large amplitude region is associated with the slower wave. The whole disturbance has propagated westward and it has acquired very anisotropic shape. The difference between the solutions obtained by two different methods is that the asymptotic solution has much finer resolution. The solution shows "jumps" in the

contours, concentrated along certain lines. These will be shown to be associated with caustics, where the method of steepest descents fails (e.g. Lighthill, 1978).

One respect in which the solution shown in Figure 3.9 will require modification is that Figure 3.9 shows the dispersive "wings" present in each of the individual σ^+ and σ^- parts of the Fourier solution of Figure 3.5 but not present in the full Fourier solution of Figure 3.6. The reason that these "wing" regions incorrectly appear in Figure 3.9 is that the solutions of Figure 3.9 include contributions from all four stationary phase points solutions of equation (3.46). Inspection of the phase (3.39) in the complex α plane shows that the path of steepest descents (PSD) indeed threads all four of these stationary phase points except at ξ, η points clearly within the "wings". There, the PSD threads only two of the four stationary phase points. When only the contributions of the stationary phase points threaded by the PSD are manually included at points ξ, η clearly inside the regions of the "wings", the "wings" then disappear. This fact has however not been taken into account in constructing Figure 3.9 because an automatic method of selecting the correct stationary phase points at each value of ξ, η has not been found.

3.6.3. Front

The solution (3.52) shows that the peak of the initial Gaussian, at $x = y = 0$, is subsequently found at $\Phi_0 = 0$. We call this the disturbance front. It is useful to nondimensionalise the variables as:

$$H = |v|^{\frac{1}{2}} H' \quad b = |v|^{\frac{1}{2}} b' \quad (3.53)$$

$$\eta = |v|^{\frac{1}{2}} \eta' \quad \xi = |v|^{\frac{1}{2}} \xi'. \quad (3.54)$$

Together with (3.41), this gives the curve described by

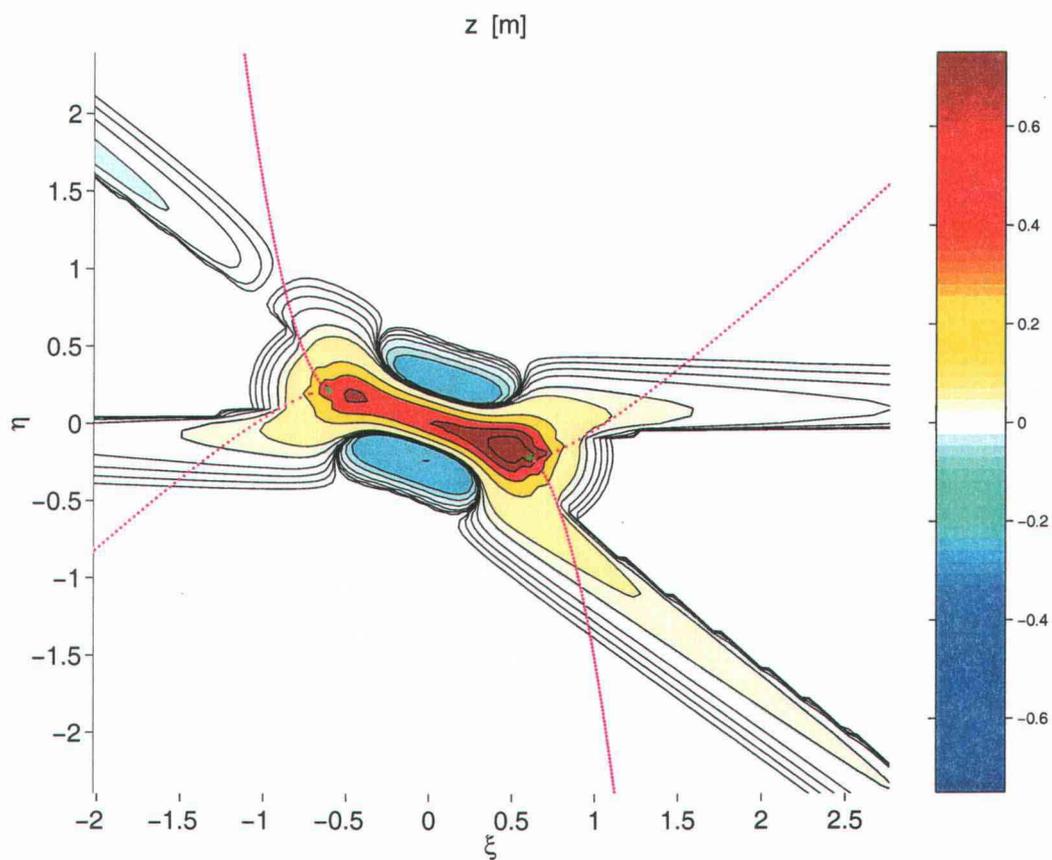


FIGURE 3.9. A snapshot of the time-dependent interface displacement of the middle layer of the transient solution obtained as the asymptotic solution for large time by the method of steepest descents, plotted in moving coordinates ξ and η . The "jumps" in the contour lines indicate that the flow develops caustics. Caustics are shown in magenta. Two green stars show the double caustic. The contour levels and the axis are the same as in Figure 3.6.

$$\xi + b\theta_0 = H_0, \quad \eta = \frac{v}{2H_0}. \quad (3.55)$$

Eliminating θ_0 and H_0 by means of

$$H_0 = \frac{v}{2\eta}, \quad \theta_0 = v\left(\frac{1}{4\eta^2} - b\right), \quad (3.56)$$

one obtains

$$\xi = v\left(\frac{1}{4\eta} + b\eta\right). \quad (3.57)$$

This simple relation gives the location of the front. The front is visible in σ^+ and σ^- components plotted in Figure 3.5.

3.6.4. Caustics

A caustic is a boundary between a region with two (or more) groups of waves, and a neighboring region with two groups of waves fewer than in the former (Lighthill, 1978). The condition for the caustic is that the first and the second derivatives of the phase (given by (3.39)) vanish:

$$\Phi'(\alpha) = 0, \quad \Phi''(\alpha) = 0. \quad (3.58)$$

The first of equations (3.58) may be solved to find the stationary phase values of α as in the previous section yielding equation (3.44). For each stationary phase value of α , equations (3.58) constitute two equations for ξ and η whose solutions define the caustic curves. We insert ξ obtained from (3.58) into second of equations (3.58) to obtain η as

$$\eta = \frac{3v^2\theta + v^2\theta^3 + 2vb}{4H^3} \quad (3.59)$$

where H is defined by the dispersion relation (3.35). From (3.59) and (3.58) we get

$$\xi = \frac{v^2(1 + \theta^2)^2}{4H^3} + H - \eta\theta. \quad (3.60)$$

The caustics are plotted in Figure 3.9. The contours of the interface displacement indeed "jump" across the caustic at the points where the stationary phase solution fails. One could use the Airy functions to "heal" the asymptotic solution (e.g. Lighthill, 1978), but that calculation will not be presented here.

3.6.5. Double Caustic

A double caustic is the set of points that satisfy the requirement that the first three derivatives of phase with respect to the wavenumber vector direction α are zero:

$$\Phi'(\alpha) = 0, \quad \Phi''(\alpha) = 0, \quad \Phi'''(\alpha) = 0. \quad (3.61)$$

There are two points that satisfy (3.61) for which θ is given by the roots of the quadratic

$$\theta = \frac{-b \pm \sqrt{b^2 + v^2}}{v}, \quad (3.62)$$

where v and b are defined by (3.33) and (3.36). The coordinates of these two points ξ and η can be obtained by inserting θ defined in (3.34) into equations (3.59) and (3.60). The two points that satisfy (3.62) are shown in Figure 3.9 as the green stars. The caustics meet in cusps at double caustics. The branches of the caustics, when extended to large distances asymptote to $\frac{\pi}{2}$ and $\arctan(-\frac{b}{v})$. The asymptotic expressions are worked out analytically in the next section.

3.6.6. Approximate solutions for stationary phase points far from the center of the disturbance

We seek the approximate analytic solutions of the quartic equation (3.46) that are valid at large distances away from the center of the disturbance; the results give the four values of the stationary phase angle α_0 .

In terms of the nondimensional variables (3.53) and (3.54) (where we have dropped the primes for the convenience) the equation (3.46) becomes:

$$(H^2 - b)(H^2 - 2H\xi + b) + 2H\eta v - 1 = 0 \quad (3.63)$$

where the nondimensional v can have values ± 1 . Write ξ and η in cylindrical polar coordinates R and β as

$$\xi = R \cos\beta \quad \eta = R \sin\beta. \quad (3.64)$$

Equation (3.63) then becomes

$$H[H^2 - (b + v \tan\beta)] + \frac{H^4 - (b^2 + 1)}{-2R \cos\beta} = 0. \quad (3.65)$$

We seek approximate solutions for large R . First seek a solution such that $H = O(1)$ and therefore neglect much smaller terms of the order R^{-1} ; one iteration on the two roots thus obtained yields

$$H_{1,2} = \pm \left[b + v \tan\beta + \frac{(b + v \tan\beta)^2 - (b^2 + 1)}{\pm 2R \cos\beta (b + v \tan\beta)^{\frac{1}{2}}} \right]^{\frac{1}{2}}. \quad (3.66)$$

These are two approximate $O(1)$ solutions for $R \rightarrow \infty$. The values of α are obtained from the dispersion relation (3.35) which gives

$$\tan\alpha_{1,2} = \tan\beta + \frac{1}{v} \left[\frac{(b + v \tan\beta)^2 - (b^2 + 1)}{\pm 2R \cos\beta (b + v \tan\beta)^{\frac{1}{2}}} \right], \quad (3.67)$$

so the two values of α are given by

$$\alpha_{1,2} = \beta \pm O(R^{-1}). \quad (3.68)$$

Thus for large R , α is close to β , but the correction term in (3.67) blows up at $\beta = \frac{\pi}{2}$ and $\beta = \arctan(-\frac{b}{v})$. We have denoted these two roots of (3.68) by subscripts 1 and 2. Next seek a large solution such that $H = O(R)$. The two highest order terms in equation (3.65) are of order R^4 . Neglecting terms of lower order yields $H^4 - 2H^3R \cos\beta = 0$, so the single approximate $O(R)$ solution is $H = 2R \cos\beta$. Define β_0 as the angle whose

$$\sin\beta_0 = \frac{b}{\sqrt{b^2+1}} \quad \cos\beta_0 = -\frac{v}{\sqrt{b^2+1}}, \quad (3.69)$$

so that the once iterated $O(R)$ solution becomes

$$H_3 = 2R \cos\beta + \frac{\sqrt{b^2+1}}{2R \cos^2\beta} \sin(\beta - \beta_0). \quad (3.70)$$

The correction of this root blows up at $\beta = \frac{\pi}{2}$. the solution α for (3.70) is

$$\alpha_3 = \frac{\pi}{2}. \quad (3.71)$$

We denote this root by subscript 3. Finally, seek a small solution such that $O(R^{-1})$.

There is one root approximated by

$$H_4 = \frac{\sqrt{b^2+1}}{2R \sin(\beta_0 - \beta)} \quad (3.72)$$

which becomes, after one iterative correction,

$$H_4 = -\frac{\sqrt{b^2+1}}{2R \sin(\beta - \beta_0)} + \frac{(b^2+1)^{\frac{3}{2}} \cos\beta}{(2R)^3 \sin^4(\beta_0 - \beta)}. \quad (3.73)$$

This correction blows up for $\beta = \beta_0$. The solution α corresponding to this root is

$$\alpha_4 = \beta_0. \quad (3.74)$$

We denote this root by subscript 4. The four roots of α given by (3.68), (3.71) and (3.74) are the approximations of the roots over which (3.52) must be summed.

When the approximations (3.68), (3.71) and (3.74) are inserted in (3.52), the result shows that the asymptotic solutions for large R decay transcendentally away from the center of the front.

3.7. Solutions for the stationary phase points uniformly valid across the caustics

We have shown in the previous section that for large distances the caustics are near $\beta = \frac{\pi}{2}$ and $\beta = \beta_0 = \arctan(-\frac{b}{v})$. Since the expressions obtained in the previous section are the crude approximations, here we work out the more accurate expressions for the position of the caustics at large distances from the center of the disturbance. In this section we also work out the approximate solutions for large distances which are uniformly valid across the caustics; to obtain them one should keep all the terms of the quartic equation (3.65) that interact at the caustic considered.

3.7.1. Solutions for stationary phase points uniformly valid across the caustic approximated by $\beta = \beta_0 = \arctan(-\frac{b}{v})$

We have shown in the previous section that a balance between the terms of the order H^3 and H^1 gives the first two approximate roots (3.68). The balance between the terms of the order H^1 and H^0 gives the last root (3.74). All three approximate solutions break down at $\beta = \beta_0$. To obtain a solution uniformly valid across the caustic close to $\beta = \beta_0$, neglect H^4 and solve the cubic equation

$$H[H^2 - (b + v \tan\beta)] + \frac{(b^2 + 1)}{2R \cos\beta} = 0. \quad (3.75)$$

Consider an angle β close to β_0 and write $\beta = \beta_0 + \delta$, where δ is a small angle. Rewrite the equation (3.75) using the expressions (3.69) as

$$2RH^3 \cos\beta + 2RH\sqrt{b^2 + 1} \sin(\beta - \beta_0) + b^2 + 1 = 0. \quad (3.76)$$

Use the approximations $\cos\beta \approx \cos\beta_0$ and $\sin(\beta - \beta_0) \approx \delta$ together with (3.69) to obtain the equation above in the form

$$-2RH^3 v + 2RH(b^2 + 1)\delta = -(b^2 + 1)^{\frac{3}{2}}. \quad (3.77)$$

We solve this equation by substitution $H = mM$ and determine m which transforms the lhs of the equation into the form $4M^3 + 3M$, which we solve analytically; with

$$m = \frac{2}{\sqrt{3}} \sqrt{\delta(1 + b^2)} \quad (3.78)$$

the equation (3.77) becomes

$$4M^3 - 3Mv = \frac{3^{\frac{3}{2}}v}{4R\delta^{\frac{3}{2}}}. \quad (3.79)$$

As we are modeling spatially homogeneous background flow towards the southwest, we choose $v = -1$ and using the substitution $M = \sinh(t)$, the lhs of the equation becomes $\sinh(3t)$. For $\delta > 0$ and $\frac{3^{\frac{3}{2}}}{4R|\delta|^{\frac{3}{2}}} < 1$ the solution can be written as

$$H = \frac{2\sqrt{\delta(1 + b^2)}}{\sqrt{3}i} \left[\sin\left(\frac{i}{3} \operatorname{arcsinh}\left(\frac{-3^{\frac{3}{2}}}{4R\delta^{\frac{3}{2}}}\right)\right) \left(1, -\frac{1}{2}, -\frac{1}{2}\right) - \cos\left(\frac{i}{3} \operatorname{arcsinh}\left(\frac{-3^{\frac{3}{2}}}{4R\delta^{\frac{3}{2}}}\right)\right) \left(0, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2}\right) \right] \quad (3.80)$$

indicating that one root is real and the other form a complex conjugate pair. For $\delta < 0$ and $\frac{3^{\frac{3}{2}}}{4R|\delta|^{\frac{3}{2}}} < 1$, the equation (3.79) has three real roots given by

$$H = \frac{2\sqrt{|\delta|(1 + b^2)}}{\sqrt{3}} \left[\sin\left(\frac{1}{3} \arcsin\left(\frac{3^{\frac{3}{2}}}{4R|\delta|^{\frac{3}{2}}}\right)\right) \left(1, -\frac{1}{2}, -\frac{1}{2}\right) - \cos\left(\frac{1}{3} \arcsin\left(\frac{3^{\frac{3}{2}}}{4R|\delta|^{\frac{3}{2}}}\right)\right) \left(0, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2}\right) \right]. \quad (3.81)$$

Finally, for $\delta < 0$ and $|\delta| < \frac{3}{(4R)^{\frac{2}{3}}}$ the solutions are given by

$$H = \frac{2\sqrt{|\delta|(1+b^2)}}{\sqrt{3}} \left[\cosh\left(\frac{1}{3}\operatorname{arcosh}\left(\frac{3^{\frac{3}{2}}}{4R|\delta|^{\frac{3}{2}}}\right)\right)\left(\frac{1}{2}, -1, -\frac{1}{2}\right) + i \sinh\left(\frac{1}{3}\operatorname{arcosh}\left(\frac{3^{\frac{3}{2}}}{4R|\delta|^{\frac{3}{2}}}\right)\right)\left(\frac{\sqrt{3}}{2}, 0, -\frac{\sqrt{3}}{2}\right) \right], \quad (3.82)$$

showing that one root is real and the other form a complex conjugate pair. This indicates that for large distances R the caustic is at $\beta = \beta_0 - \frac{3}{(4R)^{\frac{2}{3}}}$, where the solution H takes the form

$$H = \frac{2\sqrt{1+b^2}}{(4R)^{\frac{1}{3}}}\left(\frac{1}{2}, -1, \frac{1}{2}\right). \quad (3.83)$$

At the caustic all three roots are real and of order $O(R^{-\frac{1}{3}})$.

It can easily be shown that in the limit for the small argument $\frac{3^{\frac{3}{2}}}{4R|\delta|^{\frac{3}{2}}} \ll 1$ (which is satisfied for R large and δ small), the three roots of both expressions ((3.81), for $\delta < 0$ and (3.82) for for $\delta > 0$) reduce to the approximate solutions for R large (given by (3.66) and (3.73)).

3.7.2. Solution for stationary phase points uniformly valid across the caustic approximated by $\beta = \frac{\pi}{2}$

We have seen from the crude approximate solutions for the stationary phase points valid at large distances (section 3.6.6) that one caustic is close to $\beta = \frac{\pi}{2}$. The three solutions that interact at this caustic are given by balances between terms of the order H^3 and H^1 (which gives two roots (3.68)) and between terms of the order H^4 and H^3 , which gives the large root (3.71). Corrections of all of these solutions blow up at $\frac{\pi}{2}$. Thus, in order to get a solution which is uniformly valid across $\frac{\pi}{2}$, one needs to consider the equation

$$H[2RH^2\cos\beta + 2R\sqrt{b^2+1}\sin(\beta - \beta_0) - H^3] = 0. \quad (3.84)$$

One solution of this equation is $H = 0$. This is the fourth solution of the quartic equation (3.65), whose more correct form is given by (3.74). We divide the equation out by this root and solve the expression in the square bracket. Consider an angle β close to the caustic given by $\frac{\pi}{2}$, so $\beta = \frac{\pi}{2} + \delta$. Equation (3.84) then becomes

$$-2RH^2 \sin \delta + 2R\sqrt{b^2 + 1} \cos \beta_0 - H^3 = 0. \quad (3.85)$$

Approximate $\sin \delta \approx \delta$, use the definition of β_0 given by (3.69) and again consider $v = -1$ to get

$$2RH^2 \delta - 2R + H^3 = 0. \quad (3.86)$$

To obtain the analytical solution of (3.86), use the same method as in the previous section: introduce $H' = H - p$ (with $p = -\frac{2}{3}R\delta$) to eliminate the quadratic term in (3.86) and introduce $H'' = rH'$ which (with $r = 2p$) transforms the equation (3.86) in the form

$$4H''^3 - 3H'' = 1 - \frac{27}{8R^2\delta^3}. \quad (3.87)$$

With the substitution $H'' = \cosh(t)$, the solution of (3.86) becomes

$$\begin{aligned} \frac{H - p}{r} = \cosh \left[\frac{1}{3} \operatorname{arcosh} \left(1 - \frac{27}{8R^2\delta^3} \right) \right] & \left(1, -\frac{1}{2}, -\frac{1}{2} \right) + \\ i \sinh \left[\frac{1}{3} \operatorname{arcosh} \left(1 - \frac{27}{8R^2\delta^3} \right) \right] & \left(0, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2} \right). \end{aligned} \quad (3.88)$$

For δ negative and small, the following approximations

$$\begin{aligned} \cosh \left[\frac{1}{3} \operatorname{arcosh} \left(1 + \frac{27}{8R^2|\delta|^3} \right) \right] & \cong \sinh \left[\frac{1}{3} \operatorname{arcosh} \left(1 + \frac{27}{8R^2|\delta|^3} \right) \right] \\ & \cong \frac{1}{2} \left(2 + \frac{27}{4R^2|\delta|^3} \right)^{\frac{1}{3}} \cong \frac{1}{2} \left(\frac{27}{4R^2|\delta|^3} \right)^{\frac{1}{3}} \end{aligned} \quad (3.89)$$

simplify the solution (3.88) to

$$H = \frac{2R|\delta|}{3} + \frac{2R|\delta|}{3} \left(\frac{27}{4R^2|\delta|^3} \right)^{\frac{1}{3}} \left(1, -\frac{1}{2} + i\frac{\sqrt{3}}{2}, -\frac{1}{2} - i\frac{\sqrt{3}}{2} \right). \quad (3.90)$$

Neglect the first term on the rhs of (3.90) and write the solution close to $\beta = \frac{\pi}{2}$, valid for $\delta < 0$ and small, as

$$H = (2R)^{\frac{1}{3}} \left(1, -\frac{1}{2} + i\frac{\sqrt{3}}{2}, -\frac{1}{2} - i\frac{\sqrt{3}}{2} \right). \quad (3.91)$$

All three roots are of the order $R^{\frac{1}{3}}$, one of which is real and the other form a complex conjugate pair.

For $\delta > 0$ and $|\delta| > \frac{3}{2}R^{-\frac{2}{3}}$ the solution of (3.87) becomes

$$\begin{aligned} \frac{H-p}{r} = \cos \left[\frac{1}{3} \arccos \left(1 - \frac{27}{8R^2\delta^3} \right) \right] \left(1, -\frac{1}{2}, -\frac{1}{2} \right) + \\ \sin \left[\frac{1}{3} \arccos \left(1 - \frac{27}{8R^2\delta^3} \right) \right] \left(0, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2} \right) \end{aligned} \quad (3.92)$$

which shows that for $\delta > 0$, $|\delta| > \frac{3}{2}R^{-\frac{2}{3}}$ all three solutions are real.

For $\frac{3}{4}R^{-\frac{2}{3}} \leq \delta \leq \frac{3}{2}R^{-\frac{2}{3}}$, so that $-1 \leq \cosh \left(1 - \frac{27}{8R^2\delta^3} \right) \leq 0$ equation (3.88) has the same form as (3.92). For $\delta = \frac{3}{4}R^{-\frac{2}{3}}$ equation (3.92) reduces to

$$\frac{H-p}{r} = \cos \left(\frac{\pi}{3} \right) \left(1, -\frac{1}{2}, -\frac{1}{2} \right) + \sin \left(\frac{\pi}{3} \right) \left(0, \frac{\sqrt{3}}{2}, -\frac{\sqrt{3}}{2} \right) = \left(\frac{1}{2}, \frac{1}{2}, -1 \right). \quad (3.93)$$

So, $\delta = \frac{3}{4}R^{-\frac{2}{3}}$ describes the position of the caustic for large distances and thus the two solutions in (3.93) become equal. On the other side of the caustic, for $\delta < \frac{3}{4}R^{-\frac{2}{3}}$, the equation for H is

$$\begin{aligned} \frac{H-p}{r} = \cosh \left[\frac{1}{3} \operatorname{arcosh} \left(-1 + \frac{27}{8R^2\delta^3} \right) \right] \left(\frac{1}{2}, \frac{1}{2}, -1 \right) + \\ i \sinh \left[\frac{1}{3} \operatorname{arcosh} \left(-1 + \frac{27}{8R^2\delta^3} \right) \right] \left(-i\frac{\sqrt{3}}{2}, i\frac{\sqrt{3}}{2}, 0 \right), \end{aligned} \quad (3.94)$$

which has one real root and a complex conjugate pair.

3.7.3. Discussion of the solutions of the quartic equation

Figure 3.10 shows the roots of the quartic equation (3.65) as a function of β for a large value of R . In Figure 3.11 they are compared to the roots of the cubic equations corresponding to the approximations (3.75) and (3.84) in which, respectively, the quartic and the constant term are neglected in the quartic. For large R the caustics are at $\beta = \arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}}$ and $\beta = -\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$ as well as at these values plus π . We may thus distinguish four different regions in the ξ, η plane: $-\pi < \beta < \arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}}$, $\arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}} < \beta < -\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$, $-\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}} < \beta < \arctan(\frac{-b}{v}) - 3/(4R)^2/3 + \pi$, $\arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}} + \pi < \beta < \frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$, $\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}} < \beta < \pi$. The first and last are really part of the same region. In the first and third intervals all the roots of the quartic are real, in the second and fourth intervals two roots are real and two form a complex conjugate pair: in the second interval roots 1 and 2 form a complex conjugate pair and in the fourth interval roots 2 and 3 are a complex conjugate pair. The cubic approximation (3.75) is clearly uniformly valid across $\arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}}$ and $\arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}} + \pi$ (Figure 3.11, middle panel); the cubic approximation (3.84) is clearly uniformly valid across $-\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$ and $\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$ (Figure 3.11, lower panel). The coefficient of H^3 in the quartic vanishes at $-\frac{\pi}{2}, \frac{\pi}{2}$ and that of the linear term vanishes at $\arctan(\frac{-b}{v})$ and $\arctan(\frac{-b}{v}) + \pi$. This occasions some singular behavior of the solutions of the cubic approximate equations (3.75) and (3.84) but that behavior is an artifact of the approximations, and not a property of the solutions of the full quartic (3.65).

At the caustic given by $\beta = \arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}}$ all the roots are of the order $R^{-\frac{1}{3}}$; at the caustic given by $\beta = \frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$ they are of the order $R^{\frac{1}{3}}$.

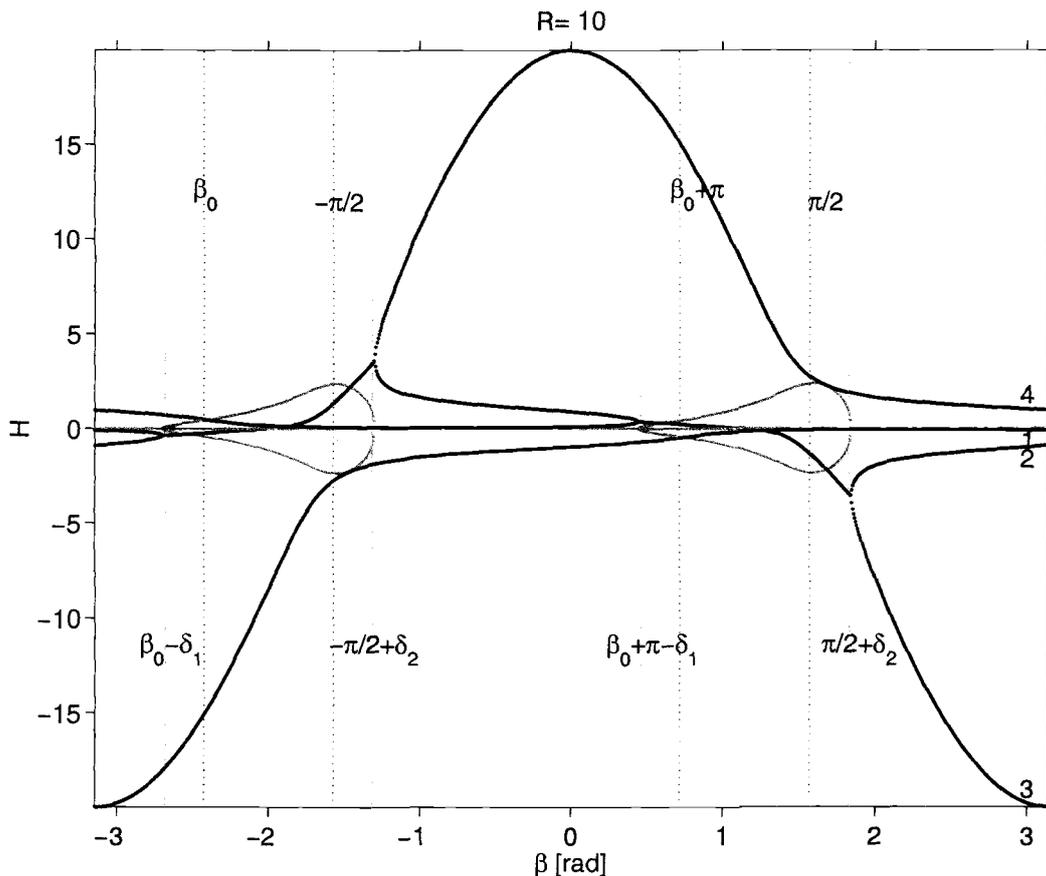


FIGURE 3.10. Four roots H_i of the quartic equation (3.65) for $R=10$. The $O(R^{-1})$ solution is labeled 3, the $O(R^1)$ solution is labeled 4 and two solutions of the order 1 close together are labeled 1 and 2. Dotted lines show the approximate caustics; dashed lines show the more accurate expressions for the caustics where $\beta_0 = \arctan(\frac{-b}{v})$, $\delta_1 = \frac{3}{(4R)^{\frac{2}{3}}}$ and $\delta_2 = \frac{3}{2^{\frac{2}{3}}} R^{-\frac{2}{3}}$. Dark solid line shows the real part of H and the light solid line shows the imaginary part of H .

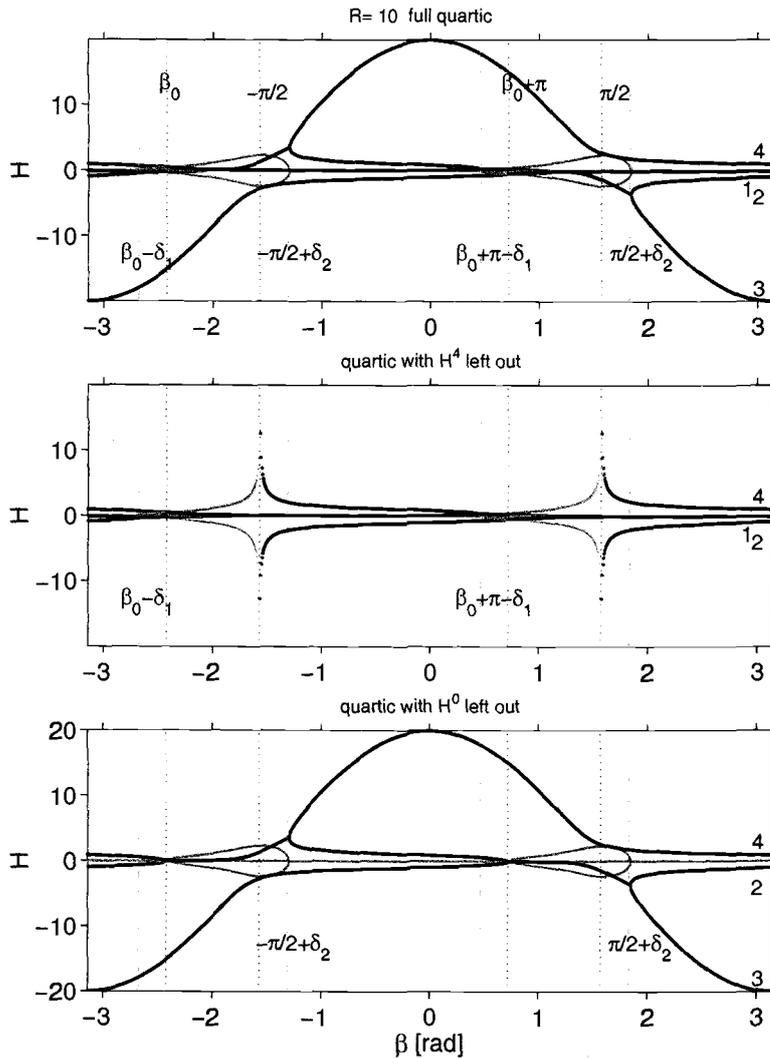


FIGURE 3.11. The roots of the quartic equation (3.65) as a function of β (upper panel) compared to the roots of the cubic equations (3.75) (middle panel) and (3.84) (lower panel) in which, respectively, the quartic and the constant term are neglected in the quartic. Dotted and dashed lines indicate the position of the caustics as described in Figure 3.10. The cubic approximation (3.75) is uniformly valid across $\arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}}$ and $\arctan(\frac{-b}{v}) - \frac{3}{(4R)^{\frac{2}{3}}} + \pi$; the cubic approximation (3.84) is uniformly valid across $-\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$ and $\frac{\pi}{2} + \frac{3}{2^{\frac{2}{3}}}R^{-\frac{2}{3}}$. The coefficient of H^3 in the quartic vanishes at $-\frac{\pi}{2}, \frac{\pi}{2}$ and that of the linear term vanishes at $\arctan(\frac{-b}{v})$ and $\arctan(\frac{-b}{v}) + \pi$ which causes some singular behavior of the solutions of the cubic approximate equations (3.75) and (3.84).

3.8. Nature of the instability

Figure 3.12 shows a time series of the logarithm of the stationary phase approximation to the middle layer interface displacement, moving with the disturbance (upper panel); it grows more rapidly than linearly as time becomes large. Initially the energy has been put into Fourier components associated with smaller temporal growth, but as the time increases shorter wavelengths, associated with faster temporal growth get energized (Figure 3.4). The lower panel of Figure 3.12 shows that the amplitude of the stationary phase approximation to the disturbance at the origin $x = 0, y = 0$, decays with time.

3.9. Transient solution for a localised wavemaker switched on at the initial time $t = 0$

Consider an initial value problem where at the initial time localized forcing of a Gaussian shape has been turned on which thereafter oscillates at the fixed frequency σ_0 . The governing equation for the time-dependent interface displacement of the the middle layer (z) is

$$CUz_{xx} + CVz_{xy} + (c + C)z_{xt} - z_{tt} = E(x, y)e^{-i\sigma_0 t} \quad (3.95)$$

where

$$E(x, y) = e^{-\frac{x^2 + y^2}{L^2}}, \quad (3.96)$$

describes the shape of the Gaussian disturbance, L is the width of the Gaussian, σ_0 is the forcing frequency and all the coefficients in the equation (3.95) are constant. The initial conditions are

$$z(x, y, 0) = 0 \quad z_t(x, y, 0) = 0. \quad (3.97)$$

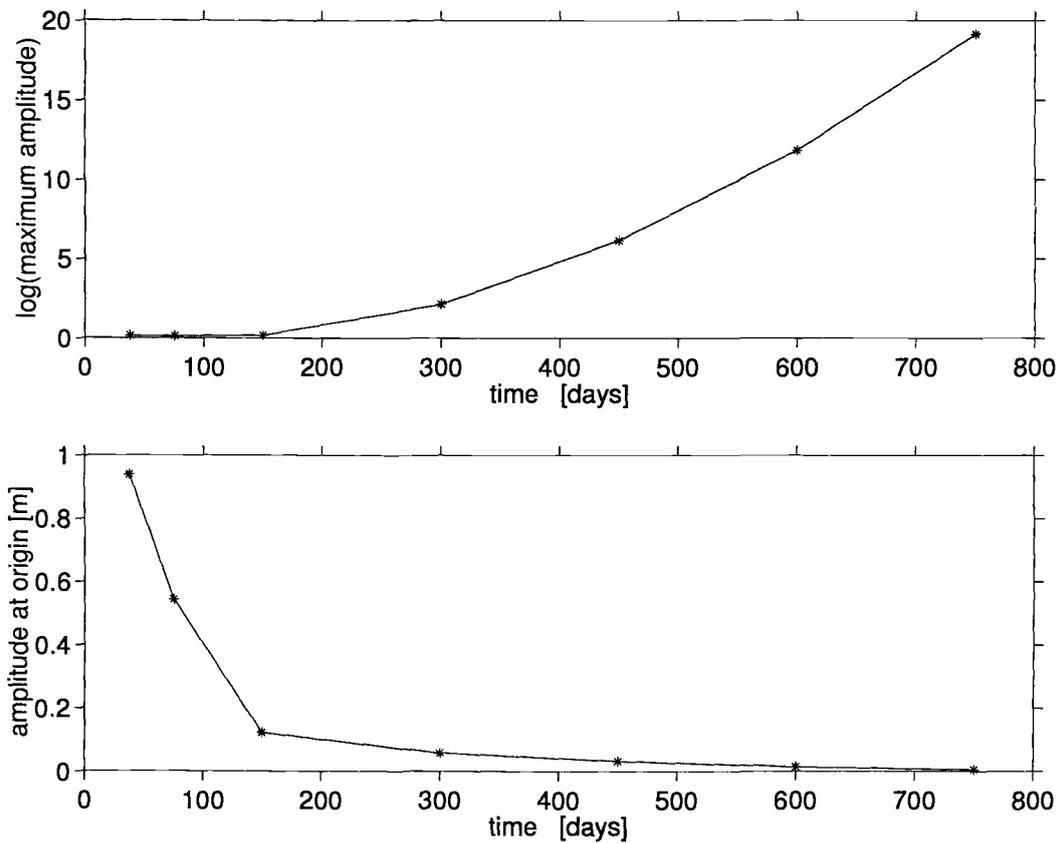


FIGURE 3.12. A time series of the logarithm of the middle layer interface displacement, moving with the disturbance (upper panel); it grows more rapidly than linearly as the time becomes large. The lower panel shows that the amplitude of the disturbance at the origin decays with time. The parameters used here are the same as in Figures 3.5, 3.6 and 3.9.

We Fourier transform equation (3.95) in space, using the convention

$$\hat{z}(k, l, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z(x, y, t) e^{-i(kx+ly)} dx dy \quad (3.98)$$

$$z(x, y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{z}(k, l, t) e^{i(kx+ly)} \frac{dk}{2\pi} \frac{dl}{2\pi}. \quad (3.99)$$

The Fourier transformed equation (3.95) becomes

$$[-CUk^2 - CVkl + (c + C)ik\partial_t - \partial_{tt}]\hat{z} = \hat{E}(k, l)e^{-i\sigma_0 t} \quad (3.100)$$

where $\hat{E}(k, l)$ is the Fourier transform of the Gaussian disturbance defined by (3.96)

$$\hat{E}(k, l) = L^2 \pi e^{-\frac{L^2(k^2+l^2)}{4}}. \quad (3.101)$$

The Fourier transformed initial conditions (3.97) are

$$\hat{z}(k, l, 0) = 0 \quad \hat{z}_t(k, l, 0) = 0. \quad (3.102)$$

The solution to (3.100) consists of a transient and sympathetically oscillating part:

$$\hat{z}(k, l, t) = \hat{z}_T(k, l, t) + \hat{z}_S(k, l, t) \quad (3.103)$$

where we write

$$\hat{z}_S(k, l, t) = \hat{z}_S(k, l) e^{-i\sigma_0 t}.$$

The sympathetic part of the solution is of the form

$$\hat{z}_S = \frac{\hat{E}(k, l)}{(\sigma_0 - \sigma^+)(\sigma_0 - \sigma^-)} \quad (3.104)$$

where σ^+ and σ^- are the two solutions of the dispersion relation (3.18). The transient solution satisfies the homogeneous part of the equation (3.100) and its solution can be written in the form:

$$\hat{z}_T(k, l, t) = Ae^{-i\sigma^+t} + Be^{-i\sigma^-t}. \quad (3.105)$$

It follows from (3.101)

$$A = \frac{-\hat{E}(k, l)}{(\sigma_0 - \sigma^+)(\sigma_0 - \sigma^-)}, \quad B = -A - \hat{z}_S. \quad (3.106)$$

Finally, using (3.99), (3.105) and (3.106), the transient and sympathetic solution can be written as

$$z_T(x, y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\hat{E}(k, l)}{(\sigma^+ - \sigma^-)} e^{i(kx+ly)} \left[\frac{e^{-i\sigma^+t}}{\sigma_0 - \sigma^+} - \frac{e^{-i\sigma^-t}}{\sigma_0 - \sigma^-} \right] \frac{dk}{2\pi} \frac{dl}{2\pi} \quad (3.107)$$

$$z_S(x, y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}(k, l) e^{i(kx+ly)} e^{-i\sigma_0 t} \frac{1}{(\sigma_0 - \sigma^+)(\sigma_0 - \sigma^-)} \frac{dk}{2\pi} \frac{dl}{2\pi} \quad (3.108)$$

where $\hat{E}(k, l)$ is given by (3.101). The importance of this development is that it shows that the transient solution (3.107) may grow spatially, but as will now be seen the sympathetic part (3.108) will not.

Equation (3.107) describes the transient solution, whose most important characteristics have been previously elucidated. In particular the amplitude of the transient solution increases without limit as $t \rightarrow \infty$. Equation (3.108) describes the sympathetically oscillating part of the solution. It oscillates as $e^{-i\sigma_0 t}$ and so does not grow without limit as $t \rightarrow \infty$. If all free motions were stable waves, then (3.108) would describe a wave train radiating from the region of forcing. The amplitude of the wave train would not grow with distance away from the forcing region. But some of the free solutions are unstable waves. The question therefore arises as to whether or not the sympathetic part of the solution described by (3.108) increases in amplitude far from the region of forcing: as $|x|, |y| \rightarrow \infty$, does $|z_S(x, y, t)| \rightarrow \infty$? To answer this question rewrite (3.108) as the sum of two expressions

$$z_S(x, y, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}(k, l) e^{i(kx+ly)} e^{-i\sigma_0 t} \left(\frac{1}{(\sigma^-(k, l) - \sigma^+(k, l))} \frac{1}{(\sigma^+(k, l) - \sigma_0)} - \frac{1}{(\sigma^-(k, l) - \sigma^+(k, l))} \frac{1}{(\sigma^-(k, l) - \sigma_0)} \right) \frac{dk dl}{2\pi 2\pi}.$$

The integrand of each of these two double integrals is finite except at the curves in the real k, l plane at which $\sigma^+(k, l) = \sigma_0$ or $\sigma^-(k, l) = \sigma_0$. For the first case, call this line $k = kp(l)$. Consider the first double integral. It may be rewritten as the sum of two double integrals

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{E}(k, l) e^{i(kx+ly)} e^{-i\sigma_0 t} \left(\frac{1}{(\sigma^-(k, l) - \sigma^+(k, l))} \left(\frac{1}{(\sigma^+(k, l) - \sigma_0)} - \frac{1}{\frac{\partial \sigma^+}{\partial k} |_{kp} (k - kp(l))} \right) + \frac{1}{(\sigma^-(kp(l), l) - \sigma^+(kp(l), l))} \frac{1}{\frac{\partial \sigma^+}{\partial k} |_{kp} (k - kp(l))} \right) \frac{dk dl}{2\pi 2\pi}.$$

The integrand of the first double integral is everywhere finite. The integral itself is bounded by the integral of the absolute value of its integrand, which is finite on account of the decay of the Gaussian $\hat{E}(k, l)$ and is independent of x, y , and so cannot grow indefinitely as $|x|, |y| \rightarrow \infty$.

The integrand of the k integral of the second double integral has a pole at $k = kp(l)$. If that k integral can be evaluated by the method of residues, then the remaining l integral will be of the form

$$\int_{-\infty}^{\infty} \hat{E}(kp(l), l) e^{i(kp(l)x+ly-\sigma_0 t)} 2\pi i \frac{1}{(\sigma^-(kp(l), l) - \sigma^+(kp(l), l))} \frac{1}{\frac{\partial \sigma^+}{\partial k} |_{kp}} \frac{dl}{2\pi}.$$

This integral itself is bounded by the integral of the absolute value of its integrand, which is finite on account of the decay of the Gaussian $\hat{E}(k, l)$ and is independent of x, y , and so cannot grow indefinitely as $|x|, |y| \rightarrow \infty$. Thus the entire expression (3.108) for $z_S(x, y, t)$ is finite as $|x|, |y| \rightarrow \infty$ and so does not grow indefinitely as $|x|, |y| \rightarrow \infty$. The same argument and conclusion holds for the second double integral.

3.10. The sympathetic solution obtained by numerically summing Fourier components

The sympathetic solution is driven by Gaussian shaped forcing with a fixed frequency representing a localised wavemaker oscillating at the fixed frequency. The governing equation is:

$$CUz_{xx} + CVz_{xy} - i\sigma_0(c + C)z_x + \sigma_0^2 z = e^{-\frac{(x^2+y^2)}{L^2}} e^{-i\sigma_0 t}, \quad (3.109)$$

where σ_0 is the fixed forcing frequency. We chose a Gaussian disturbance oscillating at the annual period. The solution is obtained by numerically summing Fourier components. Since the numerical solutions are periodic in space it is necessary to add some damping, otherwise the individual transient and sympathetic solutions are dominated by distant images of the forcing; in our examples damping is 0.02·annual frequency.

Consider two cases that differ only in the size of the initial Gaussian disturbance: in the first case the Gaussian width is 32 km and in the second it is 125 km. Figure 3.13 shows the dispersion relation for the first case. Dashed lines are the contours of the absolute value of the initial Gaussian disturbance which was very localised in space, therefore it is large in the wavenumber space. The dispersion relation (solid line) shows that the energy has been put into both stable wave branches as well as into the unstable wavenumbers. This shows that the solution will consist of two wavetrains: longer waves are moving westward and the shorter set of waves is moving towards SW. They are visible in the snap shot of the time-dependent interface displacement of the middle layer shown as a function of x and y for the sympathetic solution shown in Figure 3.14.

The sympathetic solution of Figure 3.14 extends over the entire x, y plane, but the sum of sympathetic and transient solutions plotted in Figure 3.15 is different

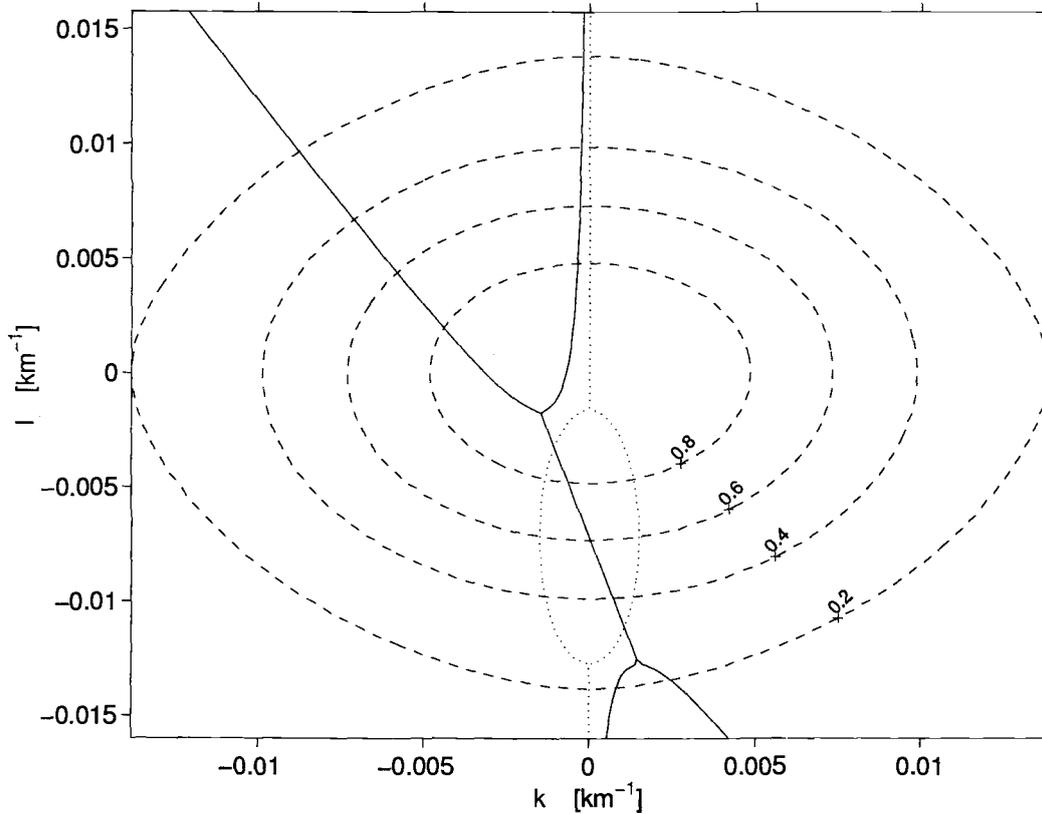


FIGURE 3.13. The dispersion relation for the annual frequency forcing: the real part of k is plotted as a solid line and the imaginary part of k (representing the spatial growth) is plotted as a dotted line. Dashed lines show the contours of absolute value of the initial Gaussian disturbance with the width of 32 km. Energy has been put into both stable branches of the dispersion relation as well as in the unstable wavenumbers.

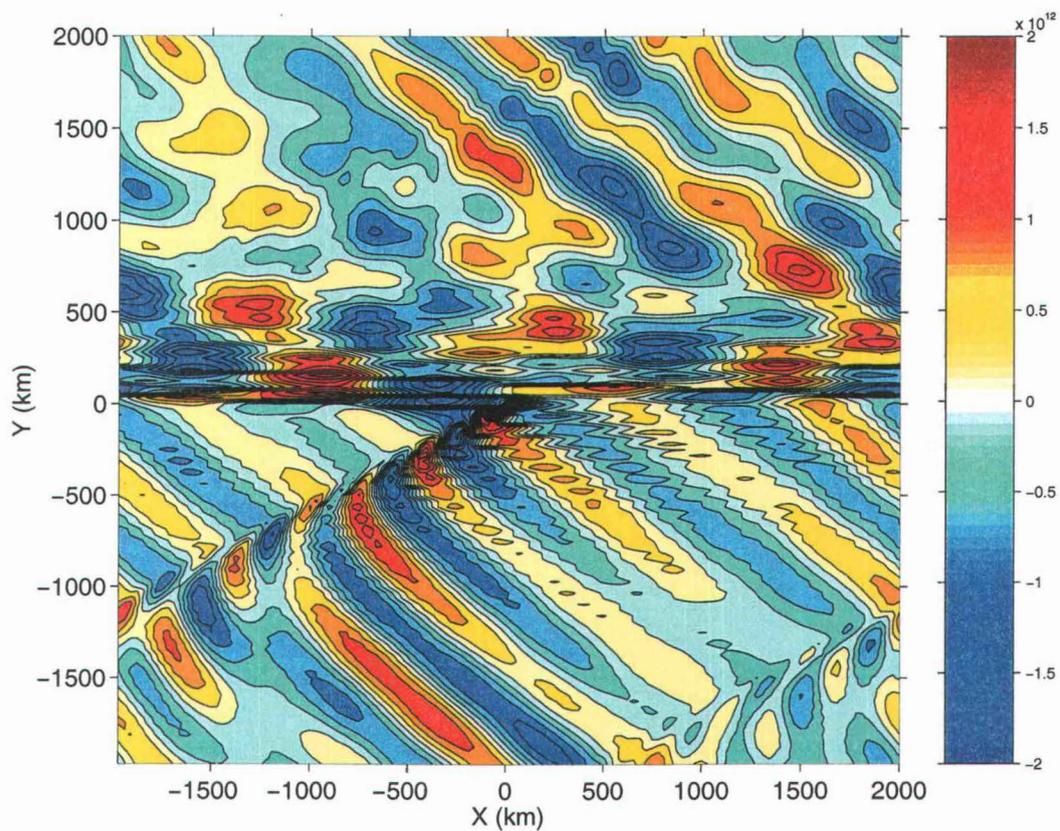


FIGURE 3.14. A snap shot of the time-dependent interface displacement of the middle layer as a function of x and y for the sympathetic solution forced with the wavemaker shown in Figure 3.13. The solution shows two wavetrains: the longer waves are moving westward and the shorter are moving towards SW. The amplitudes are very big.

from zero only in the vicinity of a localized, growing disturbance that propagates westward away from the region of forcing.

Figure 3.16 shows the dispersion relation for the Gaussian disturbance of the width 125 km. Now the energy has dominantly been put in the long wave branch, so we expect to see mostly westward propagation. Figure 3.17 shows a snap shot of the time-dependent interface displacement of the middle layer as a function of x and y for the sympathetic solution for forcing shown in Figure 3.16. Waves propagate mostly westward because more energy has been put into the long wave. In both cases, the damping is not sufficient to completely avoid wrap-around.

3.11. Conclusions

We consider small amplitude perturbations of steady subtropical gyre flow. As we restrict our attention to perturbations whose horizontal length scales are much smaller than the basin-wide scales of variation of the large scale circulation, we consider – and simply prescribe ab initio – spatially homogeneous steady upper layer flow which we assume to have been generated by solving the planetary geostrophic equations with a specified wind forcing, and then study time dependent perturbations of that flow that are excited by a specified initial perturbation and/or by harmonic forcing initiated at some initial time.

In a two-and-a-half layer planetary geostrophic model with nonzonal mean flow, long planetary waves are radially nondispersive and azimuthally dispersive (i.e. $\frac{\sigma}{\kappa}$ is only a function of the direction of the wavenumber vector α , somewhat as in the case of the internal gravity waves where σ is only a function of the direction α).

Even very gentle background flow dramatically changes the waves away from the usual simple β solutions. The presence of the nonzonal background flow couples

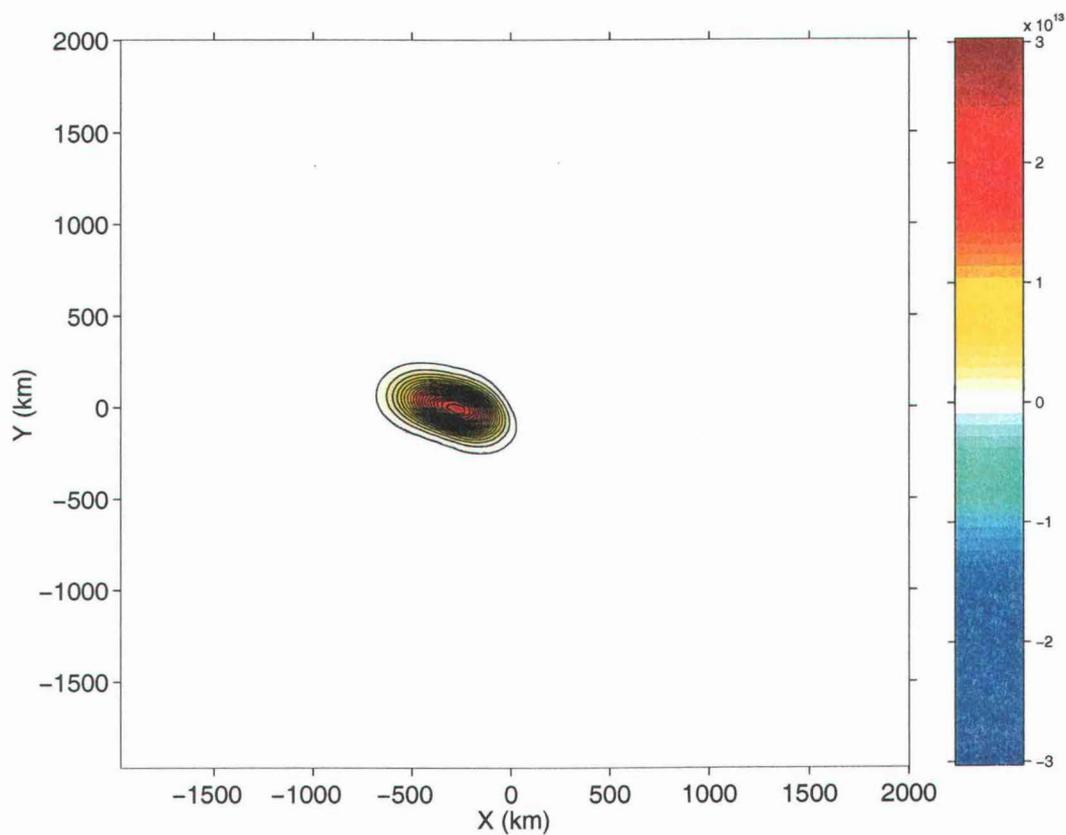


FIGURE 3.15. A snap shot of the time-dependent interface displacement of the middle layer as a function of x and y for the total (transient plus sympathetic) solution forced with the wavemaker shown in Figure 3.13.

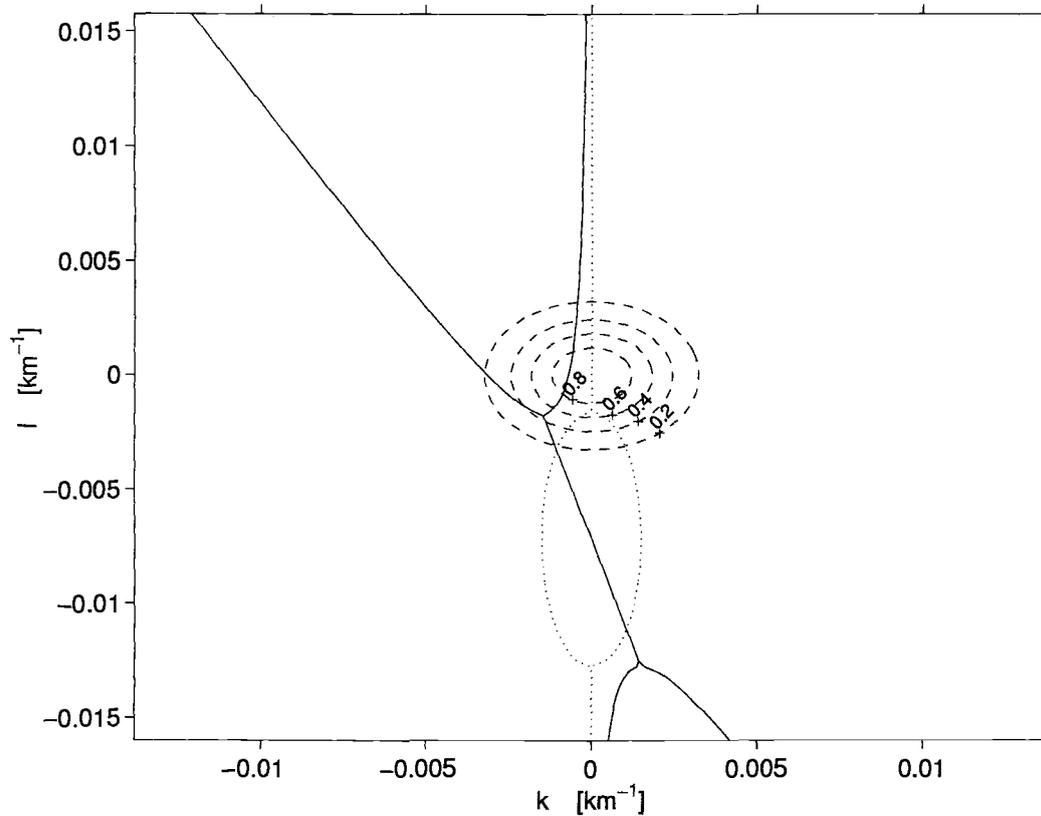


FIGURE 3.16. The dispersion relation for the annual frequency forcing: the real part of k is plotted as a solid line and the imaginary part of k (representing the spatial growth) is plotted as a dotted line. Dashed lines show the contours of absolute value of the initial Gaussian disturbance with the width of 125 km. Energy has been mostly put into the westward propagating long wave stable branch of the dispersion relation.

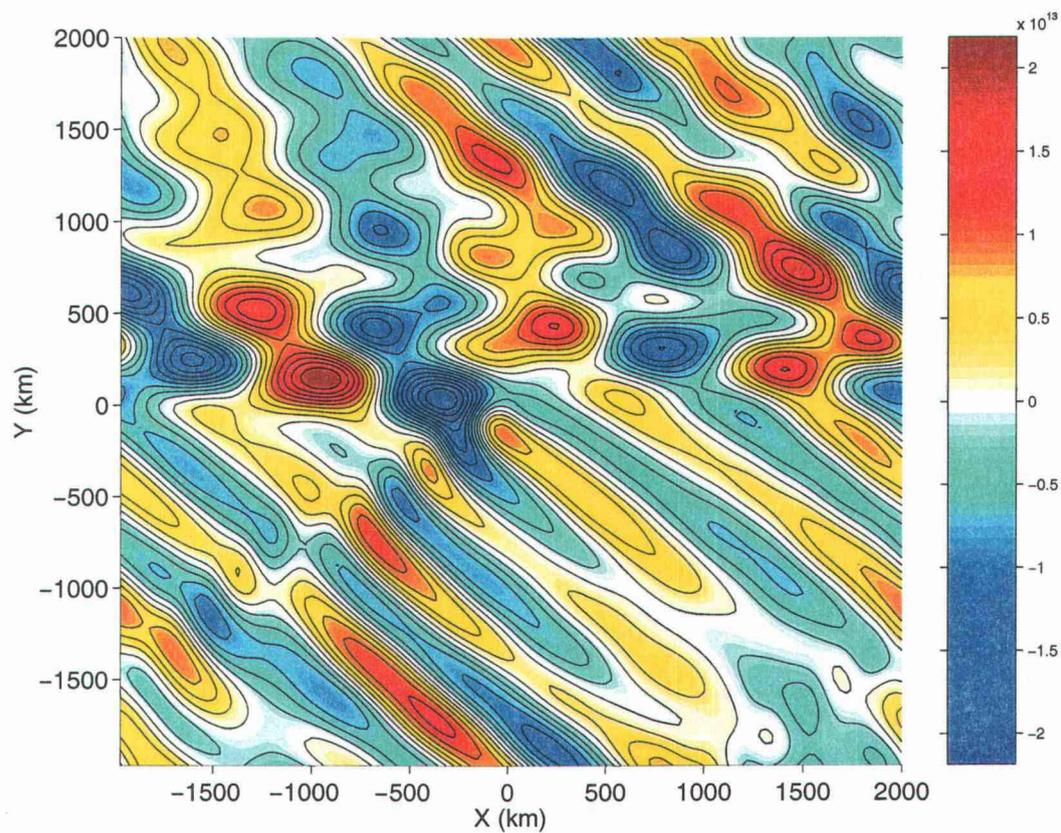


FIGURE 3.17. A snap shot of the time-dependent interface displacement of the middle layer as a function of x and y for the sympathetic solution forced with the wavemaker shown in Figure 3.16. The solution shows mostly the longer waves moving westward. The amplitudes are very big.

the two solutions of a two-and-a-half layer problem, and may cause the solutions to become unstable.

A group of long planetary waves in the presence of a nonzonal background flow evolves toward very anisotropic shape. The largest amplitudes are found at the position in the basin where the classical first and the second baroclinic mode would arrive at that time in the case of no background circulation. Only the transient solution may grow spatially, but the sympathetic part can not. The growth of the amplitude of the transient solution depends strongly on the central wavenumber and spectral width of the initial disturbance. The direction of the propagation of the sympathetic solution depends on the central wavenumber and spectral width of the forcing.

References

- Briggs, R. J., 1964: *Electron-Stream Interaction With Plasmas*. Cambridge, Mass MIT Press.
- Carrier, G. F., M. Krook, C. E. Pearson, 1966: *Functions of a complex variable: theory and technique*. McGraw-Hill, 438 pp.
- Cerovečki, I., R. A. de Szoeke: Long Planetary Waves in a Two-and-a-half Layer Model Ocean with a Wind-Driven Steady Circulation. (*submitted to J. Phys. Oceanogr.*)
- Gill, A. E., J. S. A. Green, Simmons, A. J., 1974: Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies. *Deep-Sea Research*, **21**, 499–528.

- Huerre P., and P. A. Monkewitz, 1990: Local and global instabilities in spatially developing flows. *Annu. Rev. Fluid Mech.*, **22**, 473–537.
- Lighthill, M. J., 1978: *Waves in Fluids*. Cambridge University Press, 504 pp.
- Luyten, J., and H. Stommel, 1986: Gyres driven by combined wind and buoyancy flux. *J. Phys. Oceanogr.*, **16**, 1551–1560.
- Pedlosky, J., 1987: *Geophysical Fluid Dynamics*. 2 d ed. Springer-Verlag, 710 pp.
- Pedlosky, J., 2002: The destabilization of Rossby normal modes by meridional baroclinic shear. *J. Phys. Oceanogr.*, **32**, 2418–2423.
- Samelson, R. M., 1999: Note on a baroclinic analogue of vorticity defects in shear. *J. Fluid Mech.*, **382**, 367–373.

4. SUMMARY

Long planetary waves are a mechanism by which the ocean adjusts to forcing on long time scales. Wavelike annual period signals are prominent in satellite altimetry of the ocean surface as well as in sufficiently detailed hydrographic data. In order to understand how the structure of the annual period signal is affected by the underlying large scale circulation, we consider a model in which long planetary waves are linearized about a schematic, wind-driven, steady subtropical gyre in a three layer ocean. The steady circulation divides the gyre in an east zone (EZ) where only the uppermost layer is in steady motion, and a west zone (WZ) where the uppermost and middle layers are in steady motion because the potential vorticity of the middle layer is assumed homogeneous. In both zones the upper two layers participate in the wave motion, which is driven by a wind stress fluctuating at the annual period, but the lowermost layer is quiescent because it is assumed to be infinitely deep.

The waves are governed by spatially hyperbolic equations having two families of characteristics. One family is zonal and carries information from the eastern boundary into the WZ; characteristics of the other family emanate from the eastern or western boundaries, but do not cross the boundary between EZ and WZ. The EZ wave equations are numerically integrated westward away from the eastern boundary along their characteristics, though only the westward integration along the family of zonal characteristics can continue into the WZ. Because the model domain excludes the western boundary current region, the part of the solution that propagates along characteristics emanating from the western boundary is neglected.

In the EZ two sets of baroclinic waves are apparent in the solutions. They resemble sinuous and varicose vertical modes in certain limits. The wavelengths of

one set of waves are much smaller than those characterizing the basin or the wind field. Each is a combination, coupled by the background flow, of the classical first and second mode baroclinic Rossby waves with quiescent background flow, but in marked contrast with the case of quiescent background flow, each is strongly dispersive. WKB solutions agree well with the numerical waves with regards to local wavelength and direction of propagation and to the overall basin wide distribution of wave energy. The WKB solutions allow construction of each set of waves independently of constraints on resolution imposed by the numerics. It is important to note that, on account of the way in which the background flow modifies the wave dispersion relations, the short waves can become so short in the vicinity of the boundary x_B between EZ and WZ that the neglect of relative vorticity common to all planetary geostrophic calculations is no longer self consistent.

In the southwest portion of the subtropical gyre the annual period numerical solutions may become baroclinically unstable and grow exponentially westward. The WKB solutions additionally confirm the possibility of this instability in the southwestern subtropical gyre. In the WKB solutions, instability occurs near the point in the horizontal wavenumber plane where the two stable wave branches, extrapolated across the unstable wavelengths, intersect.

The numerical procedure used for solving across the coupled characteristics in the EZ is examined using an extension of the classical von Neumann stability analysis. It is found that the procedure gives rise to spurious instabilities when finite resolution aliases one of the stable wave branches so that it crosses the other at wavenumbers that are stable in the analytic WKB solution. An important result is that naive refinement of the resolution does not necessarily improve the accuracy of the solution. The numerical instabilities may be suppressed by a moderate amount of averaging over adjacent characteristics, corresponding to a lateral diffusivity of

about $14 \text{ m}^2 \text{ s}^{-1}$ in our examples. Careful analysis is needed to distinguish between true and numerical instability in the southwestern subtropical gyre.

Numerical time-longitude plots of sea-level qualitatively resemble those of Chelton and Schlax (1996) but there are some important differences. To begin with, our time-longitude plots show a gradual westward increase of wave amplitude (except for sections that penetrate the baroclinically unstable region of the southwestern subtropical gyre) whereas all the sections of CS, at 21° N , 32° N and 39° N , as well as the section of Roemmich and Gilson (2001) show an abrupt increase in amplitude at roughly the longitude of the Hawaiian Island Chain. Tailleux and McWilliams (2000) suggest that mid basin ridge bottom topography may result in propagating Rossby waves west of the topography. The only similarly abrupt westward change in our computations is on time-longitude sections that penetrate into the baroclinically unstable region of the southwestern subtropical gyre. Our time-longitude plots also differ from those of CS and of Roemmich and Gilson (2001) in the presence of clearly defined stationary nodes through which the waves propagate, indicating that in addition to the westward traveling waves our solution also contains a very large scale annually varying component. Such a situation is however found in the analysis of Wang et al. (2001) of sealevel in the Indian Ocean. These authors attribute the stationary node to destructive interference between the response of the ocean to local annual period winds and eastern boundary radiation of Rossby waves. Finally, the similarity between the annual period time-longitude plots of CS and the time-longitude plots of "eddy locations" made by Roemmich and Gilson (2001) for both altimetry and XBT data suggest to us that much of the eddy signal of Roemmich and Gilson is the annual Rossby wave field studied here. They construct a composite eddy from all their data. On a zonal transect through the composite eddy, each isotherm is offset westward relative to the one immediately below, a structure

consonant both with baroclinic instability and with poleward heat transport by the eddy. If indeed much of the eddy signal of Roemmich and Gilson is the annual Rossby wave field studied here, then studies such as those of the present paper may be useful in studying interannual variation of eddy heat transport in the major ocean gyres.

For sufficiently strong background flow annual period numerical solutions showed westward growth in the southwestern part of the basin, though what excites this instability was not entirely clear. The three possibilities are: (i) perturbation solution generated by local winds indeed grows westward; (ii) perturbation solution that has propagated westward through inhomogeneous background flow from more easterly regions of wind stress refracts into unstable wavenumbers and begins to grow westward; (iii) perturbation solution decays westward but the numerical procedure results in spurious westward growth.

We focus on the first of these possibilities. We consider small amplitude perturbations of steady subtropical gyre flow. As we restrict our attention to perturbations whose horizontal length scales are much smaller than the basin-wide scales of variation of the large scale circulation, we prescribe *ab initio* spatially homogeneous steady upper layer flow which is a solution of the steady model described above for a specified point in the basin, and then study time-dependent perturbations of that flow which are excited by a specified initial perturbation or by harmonic forcing initiated at some initial time.

The effect of the underlying steady circulation is of crucial importance when studying long planetary waves propagation. Even very gentle nonzonal background flow dramatically changes the waves away from the usual simple β -plane solutions. The presence of the background flow couples the two solutions of a two-and-a-half layer problem and in general may cause the solutions to become unstable.

With nonzonal mean flow long planetary waves in a two-and-a-half layer model are radially nondispersive and azimuthally dispersive (i.e. phase speed $\frac{\sigma}{\kappa}$ is only a function of the direction α of the wavenumber vector, analogous to the case of the internal gravity waves where σ is only a function of the direction α). An initially Gaussian shaped group of long planetary waves propagating in the presence of a nonzonal background flow thus evolves towards a very anisotropic shape. In our two-and-a-half layer model the largest perturbation amplitudes are found at the position in the basin where the classical first and the second baroclinic mode would have arrived at that time in the case of no background circulation.

The solution driven by periodic forcing started at some initial time consists of a transient solution and a sympathetic solution oscillating at the forcing frequency. The resulting transient solution may grow spatially, but the sympathetic solution does not. This suggests that the spatial growth observed in our study of the fixed frequency annual perturbations was generated by waves incident from more easterly regions and not directly by local wind forcing.

BIBLIOGRAPHY

- Bender, C. M. and S. A. Orszag, 1978: Advanced mathematical methods for scientists and engineers. McGraw-Hill, Inc., 593 pp.
- Briggs, R. J., 1964: *Electron-Stream Interaction With Plasmas*. Cambridge, Mass MIT Press.
- Carrier, G. F., M. Krook, C. E. Pearson, 1966: *Functions of a complex variable: theory and technique*. McGraw-Hill, 438 pp.
- Cerovečki, I., R. A. de Szoeke: Long Planetary Waves in a Two-and-a-half Layer Model Ocean with a Wind-Driven Steady Circulation. (*submitted to J. Phys. Oceanogr.*)
- Chelton, D. B., and M. G. Schlax, 1996: Global observations of oceanic Rossby waves. *Science*, **272**, 234–238.
- de Szoeke, R. A., 1992: Effects of Alongshore Wind Stress at the Eastern Boundary on Circulation in the Ocean Pycnocline. *J. Phys. Oceanogr.*, **29**, 500–511.
- de Szoeke, R. A., and D. Chelton, 1999: The modification of long planetary waves by homogeneous potential vorticity layers. *J. Phys. Oceanogr.*, **22**, 247–267.
- Dewar, W. K., 1998. On "too-fast" baroclinic planetary waves in the general circulation. *J. Phys. Oceanogr.*, **28**, 1739–1758.
- Dewar, W. K., M. Y. Morris, 2000: On the propagation of baroclinic waves in the general circulation. *J. Phys. Oceanogr.*, **30**, 2637–2649.
- Feddersen, F., 1998: Weakly nonlinear shear waves. *J. Fluid Mech.*, **372**, 71–91.
- Gill, A. E., Green, J. S. A., Simmons, A. J., 1974: Energy partition in the large-scale ocean circulation and the production of mid-ocean eddies. *Deep-Sea Research*, **21**, 499–528.
- Hallberg, R. 1997: Localized coupling between surface and bottom-intensified flow over topography. *J. Phys. Oceanogr.*, **27**, 977–998.
- Hayashi, Y.-Y., Young, W. R., 1987: Stable and unstable shear modes of rotating parallel flows in shallow water. *J. Fluid Mech.*, **184**, 477–504.
- Hendershott, M. C., 1989: The ventilated thermocline in quasigeostrophic approximation. *J. Mar. Res.*, **47**, 33–53.

- Huang, R. X., and J. Pedlosky, 2000: Climate variability induced by anomalous buoyancy forcing in a multilayer model of the ventilated thermocline. *J. Phys. Oceanogr.*, **30**, 3009–3021.
- Huang, R. X., and J. Pedlosky, 1999: Climate variability inferred from a layered model of the ventilated thermocline. *J. Phys. Oceanogr.*, **29**, 779–790.
- Huerre P., and P. A. Monkewitz, 1990: Local and global instabilities in spatially developing flows. *Annu. Rev. Fluid Mech.*, **22**, 473–537.
- Keffer, T., 1985: The ventilation of the world's oceans: Maps of the potential vorticity field. *J. Phys. Oceanogr.*, **15**, 509–523.
- Killworth, P. D., D. B. Chelton, and R. A. de Szoeke, 1997: The speed of observed and theoretical long extratropical planetary waves. *J. Phys. Oceanogr.*, **27**, 1946–1966.
- Lighthill, M. J., 1978: *Waves in Fluids*. Cambridge University Press, 504 pp.
- Liu, Z., 1999: Planetary wave modes in the thermocline: Non-Doppler-shift mode, advective mode and Green mode. *Quart. J. Roy. Meteor. Soc.*, **125**, 1315–1339.
- Luyten, J., and H. Stommel, 1986: Gyres driven by combined wind and buoyancy flux. *J. Phys. Oceanogr.*, **16**, 1551–1560.
- Luyten, J., J. Pedlosky, and H. Stommel, 1983: The ventilated thermocline. *J. Phys. Oceanogr.*, **13**, 293–309.
- Pedlosky, J., 1987: *Geophysical Fluid Dynamics*. 2 d ed. Springer-Verlag, 710 pp.
- Pedlosky, J., 1996: *Ocean Circulation Theory*. Springer-Verlag, 453 pp.
- Pedlosky, J., 2002: The destabilization of Rossby normal modes by meridional baroclinic shear. *J. Phys. Oceanogr.*, **32**, 2418–2423.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1994: *Numerical Recipes in FORTRAN*. Cambridge University Press, 963 pp.
- Qiu, B., W. Miao, and P. Mueller, 1997: Propagation and decay of forced and free baroclinic Rossby waves in off-equatorial oceans. *J. Phys. Oceanogr.*, **27**, 2405–2417.
- Rhines, P. B., and W. R. Young, 1982: A theory of the wind-driven ocean circulation. I. Mid-ocean gyres. *J. Mar. Res.*, **40**, 559–596.

- Roemmich, D. and Gilson, J., 2001: Eddy transport of heat and thermocline waters in the North Pacific: a key to interannual/decadal climate variability? *J. Phys. Oceanogr.*, **31**, 675–687.
- Samelson, R. M., 1999: Note on a baroclinic analogue of vorticity defects in shear. *J. Fluid Mech.*, **382**, 367–373.
- Samelson, R. M., and G. K. Vallis, 1997: Large-scale circulation with small diapycnal diffusion: the two-thermocline limit. *J. Mar. Res.*, **55**, 223–275.
- Talley, L. D., 1988: Potential vorticity distribution in the North Pacific. *J. Phys. Oceanogr.*, **18**, 89–106.
- Tailleux, R. and J. C. McWilliams, 2000: Acceleration, Creation and Depletion of Wind-Driven, Baroclinic Rossby Waves over an Ocean Ridge. *J. Phys. Oceanogr.*, **30**, 2186–2213.
- Young, W. R., and Rhines, P. B., 1982: A theory of the wind-driven ocean circulation. II. Gyres with western boundary layers. *J. Mar. Res.*, **40**, 849–872.
- Wang, L. P., C. J. Koblinsky, S. Howden., 2001: Annual Rossby wave in the southern Indian Ocean: Why does it "appear" to break down in the middle ocean? *J. Phys. Oceanogr.*, **31**, 54–74.
- White, W., 1977: Annual forcing of baroclinic long waves in the tropical North Pacific. *J. Phys. Oceanogr.*, **7**, 50–61.
- White, W., Y. Chao, and C.-K. Tai, 1998: Coupling of biennial oceanic Rossby waves with the overlying atmosphere in the Pacific basin. *J. Phys. Oceanogr.*, **28**, 1236–1251.