STUDIES OF WAVE–MEAN INTERACTION RELEVANT TO THE MIDDLE ATMOSPHERIC CIRCULATION

Ruping Mo

Magdalene College
Cambridge

A dissertation submitted for
the degree of Doctor of Philosophy
at the University of Cambridge

December 1994

子曰：“天何言哉？四时行焉，百物生焉，天何言哉？”

《论语·阳货篇》

To my father

Preface

The research described in this thesis, which was carried out between October 1990 and December 1994 in the Department of Applied Mathematics and Theoretical Physics (DAMTP), the University of Cambridge, represents my own work, except where explicit reference to other work is made. No part of it has been submitted for a degree or any similar qualification at any other university.

I wish to acknowledge a great debt to my supervisor, Professor Michael E. McIntyre, who has been a continual source of advice, encouragement and enthusiasm. Indeed, this thesis would not have been possible without his suggestion and criticism. I would like to thank Dr. P. H. Haynes, Dr. P. Linden, Dr. R. Saravanan, and Professor W. A. Robinson for their useful comments and suggestions. I have also benefited from discussions with other members of the Atmospheric Dynamics group in DAMTP. In particular, I would like to thank Ben Edgington and Oliver Bühler, who have given so generously of their time in reading the manuscript and suggesting improvements to it. Help offered by Wengsong Weng, Nicholas Pinhey, and Jennie Henderson is also highly appreciated. I am indebted to Professors Yan Zhang and Jinhai Miao at Beijing Institute of Meteorology for their continuing encouragement and support during the course of the present work. My thanks also go to Fenton Wiffen and Rebecca Wood in the British Council, Dr. H. A. Chase of Magdalene College, and all members of DAMTP, including research personnel, secretaries, and computer officers, who have helped me in some way.

I am grateful to the British Council and the State Education Commission of the People's Republic of China for a 4-year Sino-British Friendship Scholarship, and also to Lundgren Fund, Magdalene College, and Beijing Institute of Meteorology for various financial support.

Finally, I would like to thank my wife, Liyin, for her considerable patience, help and love.

SUMMARY

By Ruping Mo

The work described in this thesis can be divided into two parts. In the first part (Chapters 2 and 3) we demonstrate how the wave-induced mean motion can be described in terms of the dynamics of Rossby-Ertel potential vorticity (PV). In the second part (Chapters 4 and 5), Rossby waves and their mean effects in the middle atmosphere are investigated within the framework of quasi-geostrophic theory.

Chapter 2 describes a simple thought experiment to highlight the usefulness of the description of wave–mean interaction in terms of wave-induced transport of PV-substance (PVS). It is shown that the wave-induced irreversible PVS transport depends crucially on wave dissipation. When the invertibility principle for the mean PV anomaly field applies from a coarse-grain perspective, the resulting balanced mean motions are dissipation-dependent, and are equivalent to the $O(a^2)$ dissipation-dependent mean motions deduced from the momentum viewpoint (a is a dimensionless amplitude parameter). When the invertibility principle applies from a fine-grain perspective, the balanced mean motions include also the $O(a^2)$ mean motions induced by the effect of wave transience. In addition, the $O(a^2)$ dissipation-dependent mean motions are cumulatively much larger than the $O(a^2)$ dissipation-independent mean motions as time goes on. Thus, even from a coarse-grain perspective, the PV description can provide a key to understanding and characterizing the general nature of wave-induced mean motions.

Some general relationships between the wave-induced PVS transport and momentum transport are derived in Chapter 3. It is shown that the wave-induced contribution to the PVS transport is closely related to the rate of dissipation of quasimomentum. This result generalizes Taylor's well-known identity, which was derived for a two-dimensional, incompressible, non-rotating fluid (Taylor, 1915), to a stably stratified, rapidly rotating fluid. It also provides a physical basis for the description of wave-induced mean motions in terms of PVS transport.

Chapter 4 focuses on the dissipative nature of the Rossby waves and their mean effects in a Charney–Drazin model. It is shown that dissipative processes in the atmosphere not only act to damp the wave amplitude, but also affect significantly the wave phase structure. Moreover, our results suggest that the existence of anomalously-signed (positive) Eliassen–Palm (EP) flux divergences in the middle atmosphere may be physically possible, and the difference between the *transformed Eulerian-mean* and the *generalized Lagrangian-mean* meridional circulation is not always negligible.

A sharp-edge model on the polar γ-plane is introduced in Chapter 5 to study Rossby waves associated with the polar vortex. Results show that the vortex edge can support both free travelling and forced Rossby waves that have a horizontal structure decaying exponentially away from the vortex edge. When the polar night jet is strong enough, the free travelling Rossby waves with each zonal wavenumber tend to travel eastward with approximately the same zonal angular phase velocity, resembling many aspects of the behaviour of the *4-day waves* observed in the winter stratosphere (Randel and Lait, 1991). Of the waves forced by the topography, only those of planetary scale can exist under the typical parameter conditions of the winter stratosphere. Dependences of the EP flux divergence and the mean meridional circulation in the sharp-edge model are also examined. In particular, our result gives no support to the 'flowing processor' hypothesis (Tuck *et al.*, 1992, 1993), which requires a significant transport of chemically perturbed air from within the stratospheric polar vortex to mid-latitudes to explain the observed ozone depletion.

CONTENTS

1 **INTRODUCTION** **1**

1.1 General remarks 1

1.2 A brief historical review 5

1.3 Zonal mean temperature and wind distributions 10

1.4 Important wave motions in the middle atmosphere 14

1.5 Fundamental properties of PV and PV-substance 16

1.6 Scope of the thesis 20

2 **DESCRIPTION OF WAVE-INDUCED MEAN MOTIONS IN TERMS OF WAVE-INDUCED TRANSPORT OF PV-SUBSTANCE: A THOUGHT EXPERIMENT** **22**

2.1 Introduction 22

2.2 Description of the thought experiment 23

2.3 Approximations to the governing equations and nondimensionalization . 25

2.3.1 *The Boussinesq approximation* 26

2.3.2 *f-plane approximation* 27

2.3.3 *Non-dimensional equations and the small turbulent dissipation assumption* 27

2.4 The linearised inertia-gravity waves 31

2.5 PVS redistribution due to the wave dissipation 36

2.5.1 The generalised Lagrangian-mean PV distribution 37

2.5.2 The Eulerian-mean PV distribution 43

2.6 The wave-induced balanced mean motions 46

2.6.1 The mass distribution, balance condition and boundary conditions 46

2.6.2 The balanced mean motions from a coarse-grain perspective 47

2.6.3 The balanced mean motions from a fine-grain perspective 52

2.7 The wave-induced mean motions in the Eulerian-mean and GLM frameworks 54

2.7.1 The wave-induced Eulerian-mean motion 55

2.7.2 The wave-induced GLM motions 58

2.8 Summary and remarks 58

Appendix 2A: The Boussinesq approximation in a compressible atmosphere 60

Appendix 2B: An independent derivation of the GLM motion induced by inertia-gravity waves 63

3 WAVE-INDUCED TRANSPORT OF PV-SUBSTANCE AND THE RATE OF QUASIMOMENTUM DISSIPATION: THE GENERALIZED TAYLOR IDENTITY 67

3.1 Introduction 67

3.2 Some basic properties of the GLM theory 69

3.3 Wave-induced contribution to the PVS transport: Integral properties . . 72

3.4 Wave-induced contribution to the PVS transport: Differential properties 75

3.5 Discussion 77

4 ROSSBY WAVES AND THEIR MEAN EFFECTS IN A CHARNEY–DRAZIN MODEL ON THE MID-LATITUDE β-PLANE 79

4.1 Introduction 79

4.2 The β-plane approximation in the mid-latitudes 81

4.2.1 The primitive equations of atmosphere in spherical coordinates . . 81

4.2.2 The β-plane approximation and quasi-geostrophic theory 83

4.3 Forced Rossby waves on the β-plane 86

4.3.1 Linearised equations for Rossby waves 86

4.3.2 Wave solutions for constant $\overline{U}$ and N^2 90

4.3.3 Wave solutions for $\overline{U}$ and N^2 varying with z 96

4.4 The quasi-geostrophic EP flux 99

4.4.1 The TEM equations and the generalized Eliassen–Palm theorem . 100

4.4.2 The mean effect of wave transience: Linear versus nonlinear Rossby waves 102

4.4.3 EP flux divergences induced by steady, nonbreaking, dissipating Rossby waves 105

4.5 The difference between GLM and TEM meridional circulations 110

4.5.1 The particle displacement and Stokes correction for quasi-geostrophic motions 111

4.5.2 Comparisons of the TEM and GLM circulations 112

4.6 Discussion 117

Appendix 4A: Demonstration of the upward wave-energy flux due to Rossby waves in the case of constant $\overline{U}$ and N^2 118

Appendix 4B: An independent check of Eqs.(4.5.22) and (4.5.24) 120

5 STRATOSPHERIC ROSSBY WAVES ON AN IDEALIZED POLAR VORTEX **122**

5.1 Introduction 122

5.2 The γ-plane approximation in the polar region 123

5.3 Conservative interfacial Rossby waves associated with the polar vortex . 129

5.3.1 A sharp-edge model for the polar vortex 129

5.3.2 Free travelling interfacial Rossby waves 131

5.3.3 Topographically forced interfacial Rossby waves 136

5.4 The effect of weak dissipation on the forced interfacial Rossby waves . . . 142

5.5 The quasi-geostrophic EP flux on the γ-plane 147

5.5.1 The TEM equations and the generalized Eliassen–Palm theorem . 147

5.5.2 EP flux divergences induced by dissipating Rossby waves 148

5.6 Wave-induced mass transport across the edge of the polar vortex 152

5.7 Discussion 157

6 CONCLUSIONS **159**

CHAPTER 1

INTRODUCTION

> Within wave science as a whole, the nature of waves in fluids is characterised especially by their ability to interact with complex fluid flow fields.
>
> — Sir James Lighthill (1978)

1.1 General remarks

The interaction between waves and mean flow is a fundamental problem in the theory of fluid dynamics. It is well-known that a wave packet propagating in a fluid medium is capable of transporting momentum, thereby exerting forces on and inducing changes in the mean flow. On the other hand, the mean-flow configuration can strongly modify the structure of the waves.

In the meteorological context, many phenomena observed in the atmosphere cannot be properly understood without considering the complex interactions between the waves and the mean flow. The most dramatic example of such interaction is the so-called *quasi-biennial oscillation* (QBO) discovered independently by Reed *et al.* (1961) and by Veryard and Ebdon (1961). The phenomenon involves the reversal of the mean easterly or westerly winds with a variable period between about 24 and 32 months in the equatorial lower stratosphere throughout a belt encircling the globe (see Andrews *et al.*, 1987 and Holton, 1992). Our current theoretical understanding of this phenomenon is based on the work of Lindzen and Holton (1968) and Holton and Lindzen (1972), who argued that the QBO is an internal oscillation that results from the interaction between mean flow and vertically propagating Kelvin and Rossby-gravity waves from the troposphere[1]. The two basic mechanisms involved are first the wave-induced angular momentum transport, cumulatively changing the mean velocity profile as the waves dissipate; and second the effect of mean shear in Doppler shifting and refracting the waves (McIntyre, 1993). These mechanisms are simultaneous and highly interdependent. An idealized form of the same

[1] It is not clear, however, that observed Rossby-gravity waves have sufficient amplitude; and other waves may also be involved (Dickinson, 1968; Lindzen and Tsay, 1975; Andrews and McIntyre, 1976a; Dunkerton, 1985; Boville and Randel, 1991; Takahashi and Boville, 1992).

phenomenon — a periodically reversing mean flow with qualitatively the same space-time pattern, driven entirely by a steady input of waves — has been demonstrated in the laboratory by Plumb and McEwan (1978).

Another interesting example of wave–mean interaction in the atmosphere is the mid-winter stratospheric sudden warming[2]. Matsuno (1971) proposed the first successful theory to model this phenomenon. According to his theory, the stratospheric sudden warming results from the upward propagation from the troposphere of Rossby waves and their interaction with the mean stratospheric flow.

This thesis is concerned with the general nature of wave–mean interaction in the atmosphere, with dynamics of the middle atmosphere[3] particularly in mind, and with emphasis on the description of the problem in terms of the potential vorticity dynamics. One important reason for being interested in the middle atmosphere is the increasing scientific and public concern over the detectable depletion of atmospheric ozone, which is largely confined to the stratosphere. Since the pioneering study of Hartley (1881), it has been known for more than a century that ozone in the stratosphere is an essential filter of ultraviolet radiation that protects life on the Earth's surface (Rowland, 1991; Wayne, 1991). By its strong absorption of ultraviolet radiation, ozone also provides the heat source that is responsible for the observed temperature inversion between the tropopause and the stratopause (see Andrews *et al.*, 1987). Thus, the discovery of the spectacular decrease in October Antarctic total ozone starting at the end of 1970's, as first reported by scientists of the British Antarctic Survey (Farman *et al.*, 1985), has led to a blossoming of interest in the middle atmosphere.

Clearly, the distribution and evolution of ozone in the atmosphere depends not only on photochemical reactions, but also on dynamical transport by air motions. It is now widely recognised that there exists a systematic, persistent, global-scale mean mass circulation in

[2] The climatological zonal-mean temperature in the winter stratosphere generally decreases towards the winter pole. The thermal wind balance condition thus requires a polar vortex with strong westerly shear with height. During some winters, however, this zonal-mean configuration is dramatically disrupted with an accompanying large-scale warming of the polar stratosphere, which can quickly (within a few days) reverse the meridional temperature gradient and create a circumpolar easterly current. Such an event is called a ***stratospheric sudden warming***.

[3] The ***middle atmosphere*** is generally regarded as the region extending from the tropopause (about 10–16 km altitude depending on latitude) to about 100 km. The bulk of the middle atmosphere consists of two main layers, i.e., the stratosphere extending from the tropopause to the stratopause at about 50 km and the mesosphere extending from the stratopause to the mesopause at about 85 km. Below the stratosphere is the troposphere and above the mesosphere is the thermosphere. The stratosphere has large static stability associated with an overall increase of temperature with height, while in the troposphere and mesosphere the temperature generally decreases with height.

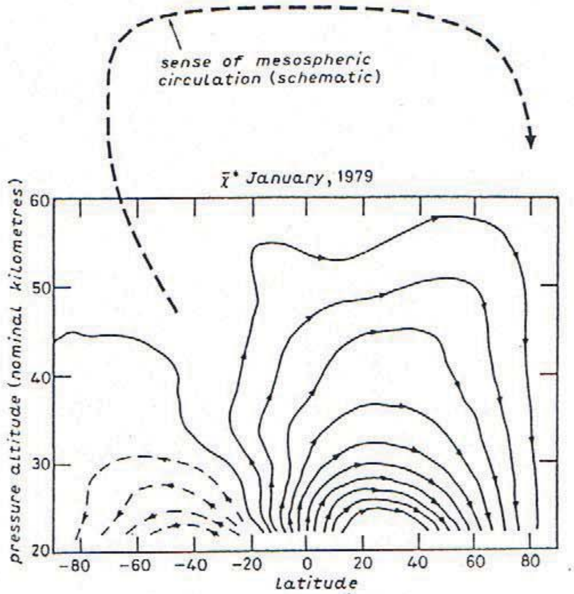


Figure 1.1: Mass transport streamlines of the global-scale mean circulation for January 1979 estimated using satellite data (light curves) from Solomon *et al.* (1986). The heavy dashed streamline was added by McIntyre (1992) to indicate schematically the mesospheric mean circulation deduced from other observational and theoretical evidences (e.g., WMO, 1985; Andrews *et al.*, 1987).

the middle atmosphere, as shown in Fig.1.1 (after McIntyre, 1992). This mass circulation is believed to make a significant contribution to the meridional transport of ozone from its source region in the tropical upper stratosphere to the high-latitude lower stratosphere. It may also be a crucial factor in the vertical transport of ozone-destroying chemicals (e.g., HO_x, NO_x, and ClO_x) from the polluted troposphere to the middle atmosphere; this is believed to be an important mechanism that causes the significant depletion of stratospheric ozone (Farman *et al.*, 1985; Wayne, 1991; Holton, 1992; McIntyre, 1992).

From a dynamical viewpoint, on the other hand, the mean meridional circulation in the middle atmosphere is primarily a wave-induced circulation (Andrews *et al.*, 1987; Holton, 1992). In the absence of wave motions the zonal mean temperature of the middle atmosphere would relax to a radiative equilibrium which can be calculated from a radiative-photochemical model of the middle atmosphere together with a radiative-convective model of the troposphere (Fels, 1985). The corresponding circulation would consist of a zonal-mean zonal flow in thermal wind balance with the meridional temperature gradient, and a negligible meridional circulation associated with the annually-varying, radiatively-determined temperature field. This implies that any significant departure from the radiatively determined state in the middle atmosphere, such as the systematic mean meridional mass circulation shown in Fig.1.1, must be maintained by various wave motions (Andrews *et al.*, 1987; Holton, 1992). Thus, the study of the waves and their interaction with the mean flow is a critical aspect of middle atmosphere science.

For a stably stratified, rapidly rotating fluid such as the middle atmosphere, it proves to be very useful and instructive for many different purposes to describe the motion in terms of the evolution of the Rossby-Ertel potential vorticity, hereafter PV (Hoskins *et al.*, 1985; Kurganskiy and Tatarskaya, 1987; McIntyre, 1992). The concept of PV first appeared in the works of Rossby (1936, 1940) in the development of the general theory of hydrostatic vortex motions as applied to problems in atmospheric and oceanic dynamics. In 1942 Ertel showed, for general (hydrostatic or nonhydrostatic) motion in the absence of diabatic heating and nonconservative forces, that the quantity

$$P = \rho^{-1}\boldsymbol{\zeta}_{\mathrm{a}} \cdot \nabla\theta \tag{1.1.1}$$

is materially conserved, where $\rho(\mathbf{x}, t)$ and $\theta(\mathbf{x}, t)$ denote respectively the density of the fluid and the potential temperature at each point $\mathbf{x}$ and instant t, and

$$\boldsymbol{\zeta}_{\mathrm{a}} = 2\boldsymbol{\Omega} + \nabla \times \mathbf{u} \tag{1.1.2}$$

is the absolute vorticity vector, with $\boldsymbol{\Omega}$ being the Earth's angular velocity vector and $\mathbf{u}(\mathbf{x}, t)$ the three-dimensional velocity vector relative to the Earth. Eq.(1.1.1) gives a general definition of PV. The significance and importance of PV for understanding the geophysical fluid motion stems from some fundamental properties of the PV, which are now briefly summarised as follows:

- the PV is materially conserved in the absence of diabatic heating and nonconservative forces — a result known in the general, nonhydrostatic case as *Ertel's theorem* (Ertel, 1942);
- the PV satisfies an exact conservation equation, which allows us to imagine PV as the mixing ratio of a generalised tracer, whose "molecules" are neither created nor destroyed away from boundaries — this generalised tracer may be referred to as the "PV-substance" or the "PVS" (Haynes and McIntyre, 1987, 1990);

- the isentropic surfaces behave as if they were completely impermeable to the PVS, even in the presence of friction and diabatic heating — the *impermeability theorem* (Haynes and McIntyre, 1987, 1990);

- the isentropic PV distributions contain nearly all the dynamical information of the large-scale motion in the sense that, if the total mass under each isentropic surface is specified, then, under a suitable balance condition, a knowledge of the distribution of PV on each isentropic surface and of potential temperature at the lower boundary is sufficient to deduce diagnostically all the other dynamical fields — the (PV) *invertibility principle* (Charney, 1948; Kleinschmidt, 1950a,b, 1951; Hoskins *et al.*, 1985 and references therein.)

These fundamental properties of PV, together with their mathematical and physical background, will be further discussed in Section 1.5.

1.2 A brief historical review

Early studies relevant to wave–mean interaction in fluids date from the late nineteenth century with the pioneering contributions by Lord Rayleigh. In his theoretical investigations of the stability of a parallel flow of an inviscid fluid, Rayleigh (1880) demonstrated that a necessary condition for an initially small-amplitude wave to grow with time is that the mean parallel-flow velocity must possess an inflection point. The application and generalisation of Rayleigh's stability theory to the problems of meteorological interest have been successfully treated by Charney (1947), Eady (1949), Kuo (1949, 1951,1953), Fjørtoft (1950), Charney and Stern (1962), among others. In 1896, Rayleigh published his second volume of "The Theory of Sound", in which he pointed out that certain mean flows can be induced by dissipating acoustic oscillations. This idea has been revitalized by recent evidence that the dissipative type of wave-induced mean flows plays an important role in the middle atmospheric circulation (e.g., Holton and Lindzen, 1972; WMO, 1985; McIntyre and Norton, 1990).

Probably the first to state explicitly that the atmospheric circulation involves the interaction between the waves and the mean flow was Helmholtz, as is evidenced by the following quotation from the introduction to a paper by him published in 1889:

> The calculations performed by me show further that for the observed velocities of the wind there may be formed in the atmosphere not only small waves, but also those whose

> wave-lengths are many kilometers which, when they approach the earth's surface to within an altitude of one or several kilometers, set the lower strata of air into violent motion and must bring about the so-called gusty weather. The peculiarity of such weather (as I look at it) consists in this, that gusts of wind often accompanied by rain are repeated at the same place, many times a day, at nearly equal intervals and nearly uniform order of succession.
>
> I think it may be assumed that this formation of waves in the atmosphere most frequently gives occasion to the mixture of atmospheric strata and, under favourable circumstances, when the ascending masses form mist, give opportunity for disturbances of an equilibrium that had already become nearly unstable. Under conditions, such as those where we see water waves breaking and forming white caps, thorough mixtures must form between the strata of air [Originally written in German, English transl. in Abbe (1893, pp.94–111)].

Although Helmholtz did not realise at that time that the dominant waves in the atmosphere may have horizontal wave-lengths of thousands of kilometers, he had a clear idea in mind that waves should play a role in the general circulation of atmosphere. Furthermore, the remark quoted above embodies clearly the basic concept of wave–mean interaction; in particular, Helmholtz correctly identified wave breaking as an important mechanism through which waves exert their influences on the atmospheric circulation.

The modern era in the study of the wave–mean interaction begins with two pioneering papers by Eliassen and Palm (1961) and Charney and Drazin (1961). Eliassen and Palm pointed out that the eddy flux of momentum is constant in an inviscid, adiabatic, and nonrotating fluid, except at a critical level where the mean velocity is equal to the horizontal phase velocity of the waves. This result had in fact been worked out already by Foote and Lin (1950) in the context of hydrodynamic stability theory (also see Lin, 1955). The matter becomes much complicated in the atmosphere when the effect of the Earth's rotation cannot be omitted, because under such circumstances both eddy heat and momentum transports may modify the mean flow. However, Eliassen and Palm (*op. cit.*) were able to demonstrate that the combined effects of eddy heat and momentum transports on the mean flow in a rotating fluid can be summarised into a vector, which now is generally referred to as the 'EP flux'. Charney and Drazin (*op. cit.*) first demonstrated that vertical propagation of stationary Rossby waves can occur only in the presence of westerly winds weaker than a critical value that depends on the horizontal scale of the waves. When regard was paid to the rectified nonlinear effects of the waves on the mean flow, they revealed that, away from the critical level, the divergence of the EP flux associated with steady, conservative and linear Rossby waves is zero; thus there is no forcing of the zonal mean flow by such waves. This rather surprising result is called the *Charney–Drazin nonacceleration theorem.*

Neither Eliassen and Palm nor Charney and Drazin discussed what happens when the waves encounter critical levels. The major difficulty arises from the fact that there is a singularity in the wave structure equation at the critical level. The problem was first investigated by Bretherton (1966a, 1969). Using the concept of wave packets in a slowly varying mean flow, Bretherton pointed out that a wave packet travelling with the appropriate local group velocity will approach the critical level but never reach it; thus the packet is neither transmitted nor reflected — it simply slows down until either dissipation or nonlinearity destroys it as a coherent entity. Booker and Bretherton (1967) gave an alternative description of the problem by introducing a complex phase velocity for internal gravity waves with a small imaginary part in order to remove the singularity associated with the critical level, and by looking at the asymptotics of the time-dependent initial value problem. Their result indicates that a wave travelling through the critical level has its amplitude attenuated by a constant factor dependent on the local Richardson number, thus the proof of the constancy of the eddy flux of momentum by Eliassen and Palm (1961) fails at the critical level. In other-words, a wave-induced force is exerted on the mean flow in the vicinity of the critical level (i.e., in the critical layer).

Hazel (1967) pointed out that a small, positive imaginary part of the phase velocity introduced by Booker and Bretherton (*op. cit.*) corresponds physically to a wave damped by mechanical and thermal dissipation. On the other hand, Benney and Bergeron (1969) and Davis (1969) argued that, if allowance is made for disturbances of finite amplitude, the critical layer need not be associated with dissipation effects, and there are some important cases where the nonlinear theory is very appealing. In fact, the importance of nonlinear processes in the problem had already been noted by Booker and Bretherton themselves. Incidentally, since the pioneering study of the linear, time-dependent Rossby-wave critical layer by Dickinson (1970), whose theory was further elucidated by Warn and Warn (1976), there has been continuing interest in the corresponding nonlinear problem for Rossby-wave critical layers (e.g., Geisler and Dickinson, 1974; Béland, 1976, 1978; Stewartson, 1978; Warn and Warn, 1978; Killworth and McIntyre, 1985; Haynes, 1985, 1989; McIntyre and Norton, 1990). The importance of the problem in the geophysical context is that it provides a paradigm for the phenomenon of Rossby-wave breaking processes that seems typical of large-scale atmospheric motions (McIntyre and Palmer, 1983, 1984, 1985).

Booker and Bretherton's theory for the critical layer problem is based on a non-rotating model for internal gravity waves. Applications of this theory to the rotating atmosphere were immediately made by Lindzen and Holton (1968) to show how upward-propagating, planetary-scale equatorial waves may drive the QBO in the zonal wind observed in the tropical stratosphere, and by Dickinson (1969a) to show how vertically propagating, mid-latitude Rossby waves may account for the remarkable zonal wind changes which occur during the sudden warming in the winter stratosphere. Dickinson's study also dealt with the effects of thermal dissipation without critical levels. Later on, Holton and Lindzen (1972) replaced their earlier QBO theory, which depended on the presence of critical levels, with a new theory wherein the damping of short-period waves by infrared cooling, rather than the critical levels, is directly relevant to the acceleration of the mean flow. Andrews and McIntyre (1976a,b, 1978a) and Boyd (1976) took a key step by deriving a general relationship between the acceleration of zonal mean flow and wave transience and nonconservative forcing, which may be expressed as

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla\cdot\mathbf{S} = \mathcal{D} + O(a^3), \tag{1.2.1}$$

where a is a dimensionless amplitude parameter, $\mathbf{S}$ is the EP flux vector whose divergence is a direct measure of the total forcing of the zonal-mean state by the eddies, $\mathcal{A}$ and $\mathcal{D}$, being respectively referred to as the density of wave activity and the non-conservative source (or sink) of wave activity, are $O(a^2)$ mean quadratic functions of disturbance quantities that generally involve parcel displacements associated with the waves. The nonlinear wave effects are implicitly represented by the $O(a^3)$ term. Eq.(1.2.1) is called the *generalized Eliassen–Palm theorem* (Edmon *et al.*, 1980). It makes explicit the dependence of the EP flux divergence on the wave transience ($-\partial\mathcal{A}/\partial t$) and non-conservative wave effects ($\mathcal{D}$). The effects of these physical mechanisms may be amplified by, but not crucially depend on, the presence of critical levels.

The generalized Eliassen–Palm theorem given in Eq.(1.2.1) does not explicitly include nonlinear wave effects. For linear, conservative waves, Andrews and McIntyre (1976a, 1978b) emphasized that wave transience by itself can cause only a temporary, reversible mean-flow change; in contrast, irreversible momentum transport can be induced by wave dissipation processes, which lead to permanent mean-flow changes. It is worth noting that the linear theory must be applied to the atmosphere with extreme caution. McIntyre and Palmer (1983, 1984, 1985) obtained some clear signs, which show that the dynamics of the stratosphere is highly nonlinear in some regions, where Rossby wave breaking appears to be a very common process. In order to recover a full description of wave generation, propagation, breaking, and their interaction with the mean-flow, a number of finite-amplitude generalisations of the generalized Eliassen–Palm theorem have been constructed (e.g., Andrews and McIntyre, 1978b,c; Killworth and McIntyre, 1985; McIntyre and Shepherd, 1987; Haynes, 1988). Having noted the facts that both dissipating and breaking waves tend to rearrange PVS along the isentropic surfaces, and that the PV distribution contains nearly all the dynamical information of the large-scale motion, McIntyre and Norton (1990) suggested that it might be convenient and useful to widen the sense of wave dissipation to include all wave-breaking processes, and to replace the notion of wave-induced mean force by the notion of wave-induced PVS transport.

Here we should take note of a remarkable similarity. In 1915, G. I. Taylor suggested that a disturbance-induced mean force in an incompressible two-dimensional fluid can be thought of as equivalent to an averaged transport of vorticity. Taylor realised well that the fundamental mean effect of eddies (waves or turbulence) is to transfer momentum from the layer in which they are generated to the layer in which they are dissipated (mixed). However, he felt that it would be difficult to analyze directly the eddy transport of momentum, because in the case of transfer of momentum the eddy can lose or gain velocity owing to the existence of local variations in pressure over a horizontal plane. Noticing that vorticity is a materially conserved quantity in a two-dimensional, homogeneous, and incompressible fluid and its evolution is not affected by the existence of local pressure gradients, Taylor suggested that the effect of eddies on the mean flow could be examined conveniently in terms of the diffusion of vorticity. To justify his ideas, Taylor derived an important identity to relate the eddy momentum flux divergence to the eddy vorticity transport, and further proved that the momentum transport theory turns out to be identical to the vorticity transport theory when the eddies are confined to a two-dimensional plane perpendicular to the direction of the mean flow.

Taylor's idea has been applied and generalized to geophysical fluid dynamics within the framework of quasi-geostrophic theory for a stably stratified, rapidly rotating fluid since the 1960s. Bretherton (1966a) first derived an important quasi-geostrophic relationship between the northward eddy flux of potential vorticity and the divergence of the EP flux. Such a relationship was re-derived by Green (1970) and applied to atmospheric data by Edmon *et al.* (1980). Dickinson (1969a) further pointed out that it is necessary to transport potential vorticity by the eddies in order to force a zonal flow. It is understood that the potential vorticity used by Bretherton (*op. cit.*) and Dickinson (*op. cit.*) is the quasi-geostrophic potential vorticity (henceforth denoted QGPV). Charney and Stern (1962) derived an elegant relationship between the derivatives of the QGPV and the PV, which shows that, when quasi-geostrophic scaling holds, the variation of the QGPV on a constant altitude or pressure surface is approximately proportional to the variation of the PV on an isentropic surface. Generally speaking, however, the QGPV differs from the PV in significant respects: for instance, it is not even a quasi-geostrophic approximation to the PV, and, in the absence of diabatic and frictional effects, the QGPV is conserved only in an approximate sense, i.e., it is conserved following the horizontal geostrophic flow, rather than following an air parcel (Hoskins *et al.*, 1985). Furthermore, there are difficulties in applying quasi-geostrophic theory to domains wide enough to include both the mid-latitudes and the tropics. In addition, the quasi-geostrophic theory is inaccurate near sloping tropopause transitions in the static stability and near jet streams with relative vorticity comparable to the planetary vorticity due to the Earth's rotation (Hoskins *et al.*, 1985; Hoskins, 1991). These facts, together with the PV fundamentals outlined in Section 1.1, directly motivate the renewed interest in recent years in using the PV to illuminate the geophysical fluid dynamics, including the wave–mean interaction processes in the atmosphere.

1.3 Zonal mean temperature and wind distributions

This thesis is about the interaction between the waves and the mean flow in the atmosphere. To begin with, we need some overview of the characteristics of the mean flow and the superposed waves in the atmosphere. In this section, the zonal mean temperature and wind distributions in the winter and summer atmosphere are briefly described. Important wave motions in the middle atmosphere will be outlined in the next section.

Fig.1.2 shows the observed zonal mean temperature cross sections for January and July in the troposphere and middle atmosphere. Data for the diagrams are obtained from the 1986 COSPAR International Reference Atmosphere compilation (see Fleming *et al.*, 1990). In the troposphere, the temperature structure is determined mainly by the balance between infrared radiative cooling and heat transport due to small-scale and synoptic-scale eddies. As a result, the temperature in the troposphere has its maximum on the Earth's surface in the equatorial region, decreasing horizontally toward both the winter and summer poles, and vertically toward the tropopause. Above the tropopause, on the other hand, the main features of the temperature structure are determined primarily by radiative processes. In the stratosphere, for instance, the increase of temperature with

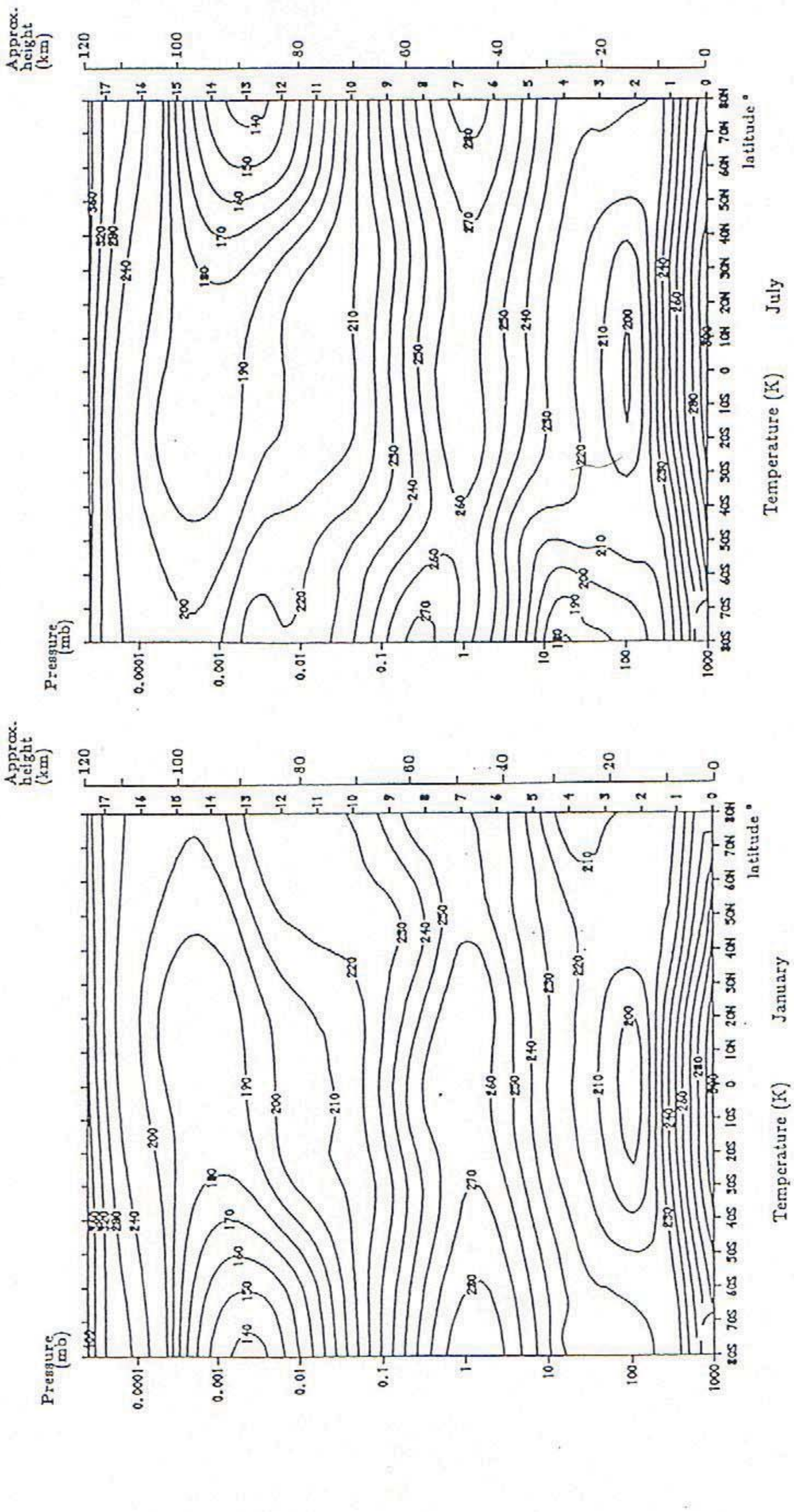


Figure 1.2: Observed monthly and zonally averaged temperature (K) in the lower and middle atmosphere for (a) January and (b) July. (After Fleming *et al.*, 1990.)

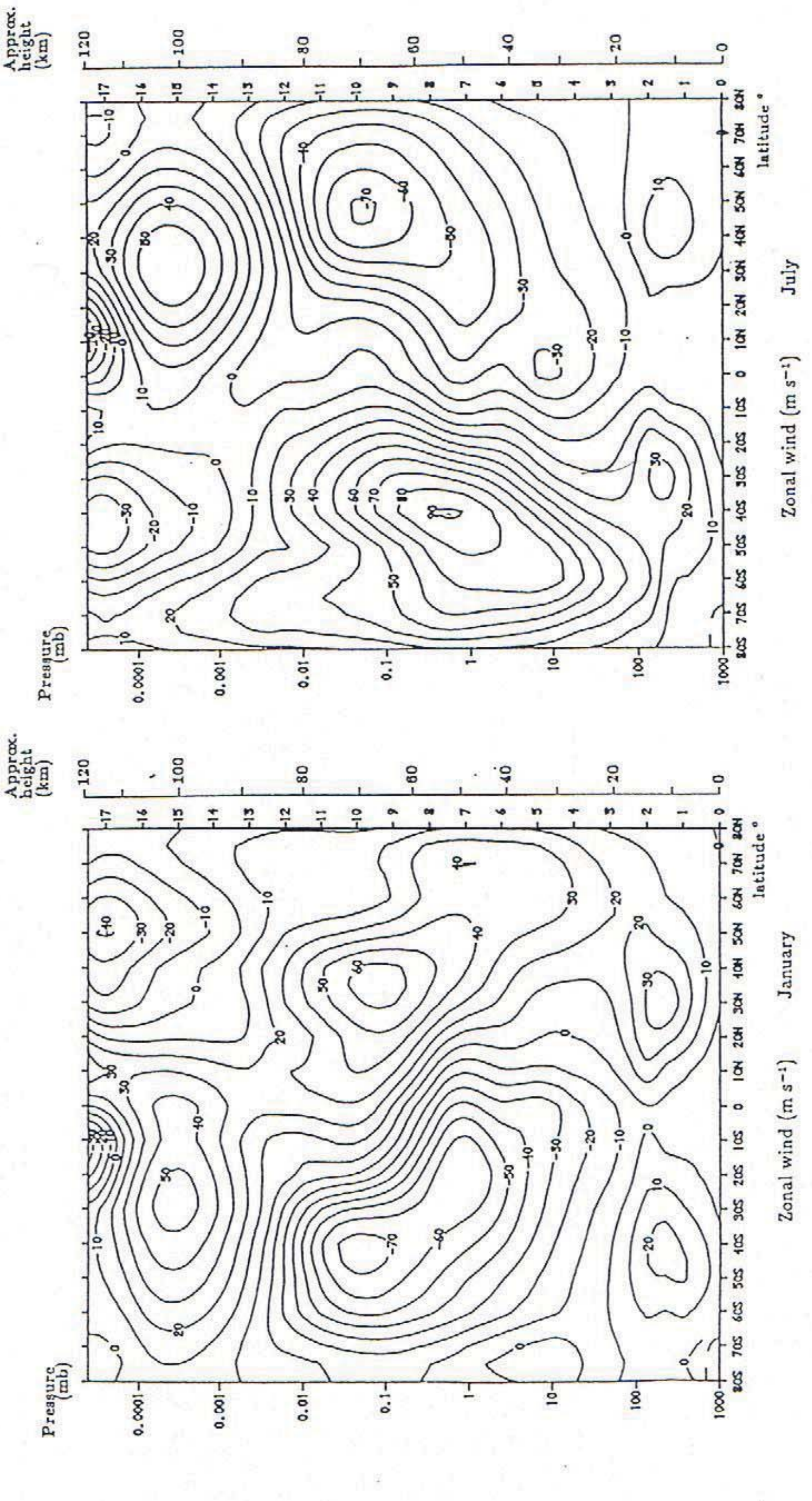


Figure 1.3: Zonal mean geostrophic winds (ms^{-1}) based on the temperatures shown in Fig.1.2 for (a) January and (b) July. (After Fleming *et al.*, 1990.)

height to a maximum at the stratopause near 50 km is due to absorption of solar radiation by ozone, and the uniform decrease from summer pole to winter pole above about 25 km is also in qualitative accord with radiative equilibrium conditions. In the mesosphere, temperature decreases with height owing to the reduced solar heating of ozone. Above the mesopause, solar ultraviolet radiation is strongly absorbed, particularly by molecular and atomic oxygen, resulting in the rapid increase of temperature in the thermosphere (Holton, 1992).

A most striking feature of the data of Fig.1.2 is that the lowest temperature occurs at the polar mesopause in the summer hemisphere rather than in the winter hemisphere where there is no solar radiation. Clearly, radiative considerations cannot provide an explanation of this phenomenon. Its existence, therefore, demonstrates that it is necessary to include some dynamical processes to model the climatological thermal structure of the middle atmosphere. Reference to Fig.1.1 shows that there is a global-scale mean mass circulation in the middle atmosphere. Associated with this mass circulation, air is being compressed and trying to warm adiabatically when it is descending, and *vice versa* (McIntyre, 1992). Therefore, it is the upper-mesospheric branch of the circulation, as sketched schematically by the heavy dashed streamline in Fig.1.1, that cause the remarkably cold temperatures at the summer polar mesopause. The stratospheric mass circulation shown in Fig.1.1 is also believed to be responsible for the departures from radiative equilibrium occurring in the lower stratosphere, where, as shown in Fig.1.2, the observed temperatures are substantially colder over the equator than over the poles (Holton, 1983a). As mentioned in Section 1.1, the mass circulation shown in Fig.1.1 should be regarded as primarily a wave-induced circulation. Its existence depends crucially on the irreversible transfer of angular momentum by certain atmospheric wave motions propagating or diffracting upwards from the troposphere and dissipating in the middle atmosphere (Andrews *et al.*, 1987; Holton, 1992; McIntyre, 1992).

The mean zonal winds derived from the associated temperature field are shown in Fig.1.3. The main features are an easterly jet in the summer hemisphere and a westerly jet in the winter hemisphere, with maxima in the wind speeds occurring near 60 km. This mean-flow configuration exerts a strong control on wave propagation in the atmosphere, as will be discussed in Section 1.4. In particular, the high-latitude westerly jets in the winter hemisphere provide wave guides for the vertical propagation of quasi-stationary Rossby waves, which are believed to be the essential dynamical mechanism responsible for the stratospheric sudden warming phenomena (Matsuno, 1971).

1.4 Important wave motions in the middle atmosphere

For middle atmosphere the two most important types of waves are those known as Rossby waves (or planetary waves) and internal gravity waves. Rossby waves are important in their own right as major components of the total circulation. In an inviscid barotropic atmosphere of constant depth, the Rossby wave is an absolute vorticity-conserving motion that owes its existence to the variation of the Coriolis force with latitude (Rossby, 1939, 1940, 1945). More generally, the restoring mechanism to which the Rossby wave owes its existence depends on the presence of gradients of potential vorticity along some or all of the isentropic surfaces of the stable stratification (Hoskins *et al.*, 1985). In the troposphere, the Rossby wave may arise as the result of baroclinic or barotropic instabilities (Charney, 1947; Eady, 1949; Kuo, 1949, 1953), or be excited by topographic forcing (Charney and Eliassen, 1949; Bolin, 1950) and by large-scale thermal asymmetries associated with ocean-continent surface heating contrasts (Smagorinsky, 1953). The possibility that the energy of Rossby waves excited in the troposphere may propagate into the stratosphere and mesosphere was firstly investigated by Charney and Drazin (1961). Using a β-plane geometry with constant basic zonal flow $\overline{U}$ and buoyancy frequency N, Charney and Drazin showed that vertical propagation of conservative, linear Rossby waves could occur only when

$$0 < (\overline{U} - c) < U_c \equiv \frac{\beta}{\left[K^2 + f^2 / (2NH)^2\right]}, \qquad (1.4.1)$$

where f is the Coriolis parameter, β the northward gradient of the planetary vorticity, K the horizontal wave number, H the standard scale height, c the zonal phase speed. Charney and Drazin's theory provides the conceptual cornerstone for much of our understanding of properties of the vertically propagating Rossby waves in the atmosphere. It suggests that the "window" for vertical propagation of Rossby waves, defined by Eq.(1.4.1), becomes smaller as the zonal wavenumber increases. This is in qualitative agreement with observations, which show that all Rossby-type disturbances tend to be absent in the summer stratosphere where the mean zonal winds are easterly (see Fig.1.3), and only the planetary-scale stationary disturbances can penetrate from the troposphere into the winter stratosphere where the mean zonal winds are strong and westerly (Holton, 1975; Andrews *et al.*, 1987).

Eq.(1.4.1) is derived under the assumption of constant mean zonal flow $\overline{U}$ (and constant buoyancy frequency N as well). However, $\overline{U}$ is not uniformly distributed in the atmosphere. Matsuno (1970) first tried with a linear quasi-geostrophic numerical model to study the propagation of Rossby waves in the stratosphere in the presence of realistic mean zonal wind profiles. His results show that the wave energy, or rather the wave activity, is guided along the axis of maximum westerly mean zonal winds. This property agrees well with the observations and is confirmed by the more sophisticated models subsequently developed (Matsuno, 1971; van Loon *et al.*, 1973; Holton, 1975; Schoeberl and Geller, 1977; Lin, 1982; Andrews *et al.* 1987; Chen and Robinson, 1992).

Pure internal gravity waves owe their existence to buoyancy restoring forces. The detailed theory of these waves can be found in standard texts, such as Lighthill (1978). In the atmosphere, gravity waves are generally affected to some extent by the rotation of the earth. This is particularly true for the gravity waves with horizontal scales greater than a few hundred kilometers and periods greater than a few hours, which are usually called *inertia-gravity waves* (see, e.g., Gill, 1982; Andrews *et al.*, 1987; Holton, 1992). The excitations of internal gravity waves in the atmosphere can arise from fluid instabilities, collapse of fronts, forcing of topography, and so forth. Although they are not a dominant part of the tropospheric large-scale circulation in mid-latitudes, the internal gravity waves become dynamically important when they propagate into the upper mesosphere where the growth of wave amplitudes with height due to the evanescence of the density may lead to gravity-wave breaking (Lindzen, 1981; Holton, 1982). For conditions of constant mean wind $\overline{U}$ and buoyancy frequency N, it can be shown (e.g., Gill, 1982; Holton, 1992) that vertical propagation of internal gravity waves requires that

$$-\frac{N}{K} < (\overline{U} - c) < \frac{N}{K}, \tag{1.4.2}$$

where K and c are, again, the horizontal wavenumber and phase speed, respectively. Eq.(1.4.2) indicates that only for magnitudes of mean wind less than the critical value N/K will vertical propagation of stationary internal gravity waves occur.

The properties of waves mentioned above are all derived from linear theory, which restricts attention to sufficiently small, reversible displacements of material contours from some known state of equilibrium. However, owing to the decay with altitude of the density of air, linear, nondissipative wave theory predicts that the velocities and geopotential associated with vertically propagating disturbances grow with altitude. Thus, at some height the nonlinear terms that have been neglected will become important, and wave breaking followed by turbulence, small-scale mixing, and dissipation turns out to be a common process under such circumstances. As generally believed, it is mainly the irreversible angular-momentum transport due to gravity-wave breaking that drives the upper-mesospheric mass circulation (see Fig.1.1), which is responsible for the remarkable coldness of the summer polar mesopause (Houghton, 1978; Lindzen, 1981; Holton, 1982, 1983b; Fritts, 1984, 1989; McIntyre, 1992). On the other hand, Rossby-wave breaking is believed to dominate the wintertime stratospheric dynamics, leading to strong quasi-horizontal mixing and irreversible tracer transport in the midlatitude "surf zone" and polar-vortex erosion (McIntyre, 1982, 1992; McIntyre and Palmer, 1983, 1984, 1985; Juckes and McIntyre, 1987; Juckes, 1989; Norton, 1994).

Breaking Rossby waves may be compared with and contrasted with breaking gravity waves: whereas the latter tend (i) to generate three-dimensional turbulence, (ii) to deform isentropic surfaces irreversibly, and (iii) to rearrange entropy downgradient in the vertical, the former tend (i) to generate layerwise-two-dimensional geostrophic turbulence, (ii) to deform PV contours irreversibly, and (iii) to rearrange PVS and chemical tracers downgradient along the same isentropic surfaces (McIntyre and Palmer, 1985; McIntyre, 1987, 1992; McIntyre and Norton, 1990). Concerning the qualitative behaviour of a certain set of material contours, McIntyre and Palmer (1985) took the defining property of wave breaking to be the rapid and irreversible deformation of those material contours that would otherwise be undulated reversibly by the waves' restoring mechanism under the conditions assumed by the linear, nondissipative wave theory. Noticing the conservation and transport properties of PVS outlined in Section 1.1, McIntyre and Norton (1990) further pointed out that the effects of breaking Rossby and gravity waves upon the global mean circulation can be understood in a unified way by viewing all the phenomena in terms of wave-induced PVS transport along the isentropic surfaces. The main approach taken in the first part (i.e., Chapters 2 and 3) of this thesis is in this direction.

1.5 Fundamental properties of PV and PV-substance

Emphasis in this study is placed on the description of wave–mean interaction processes in terms of PVS transport. It is perhaps informative at this stage to recall more extensively the PV fundamentals which have been briefly mentioned in Section 1.1.

Consider an atmosphere surrounding the rotating Earth. The basic equations of motion may be written as follows,

$$\frac{D\mathbf{u}}{Dt} + 2\boldsymbol{\Omega} \times \mathbf{u} + \frac{1}{\rho}\nabla p + \nabla\Phi = \mathbf{F}, \tag{1.5.1}$$

$$\frac{D\theta}{Dt} = Q, \tag{1.5.2}$$

$$\frac{D\rho}{Dt} + \rho\nabla\cdot\mathbf{u} = 0, \tag{1.5.3}$$

$$p = \rho RT, \tag{1.5.4}$$

$$\theta = T\left(\frac{p_s}{p}\right)^{\kappa} = \frac{p}{\rho R}\left(\frac{p_s}{p}\right)^{\kappa}, \tag{1.5.5}$$

where Ω is the Earth's angular velocity vector, $\mathbf{u}$ is the velocity vector relative to the Earth, ρ is the density of air, p is the pressure, p_s is a constant reference pressure, T is the temperature, θ is the potential temperature, Φ is a potential for the sum of the gravitational and centrifugal forces, $\mathbf{F}$ is the viscous force vector, Q is proportional to the rate of diabatic heating, R is the gas constant for dry air, $\kappa \equiv R/c_p \approx 2/7$, with c_p being the specific heat at constant pressure, and D/Dt is the material derivative defined by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla. \tag{1.5.6}$$

Taking the curl of Eq.(1.5.1) and then forming the scalar product of it with $\nabla\theta$, with the aid of Eq.(1.5.2) we obtain the following equation for the PV:

$$\frac{\partial}{\partial t}(\rho P) + \nabla\cdot\mathbf{J} = 0, \tag{1.5.7}$$

where

$$\mathbf{J} = \rho P\mathbf{u} - Q\boldsymbol{\zeta}_a - \mathbf{F}\times\nabla\theta, \tag{1.5.8}$$

and P, $\boldsymbol{\zeta}_a$ are defined by Eqs.(1.1.1) and (1.1.2), respectively. Eq.(1.5.7) is in general conservation form (Haynes and McIntyre, 1987, 1990; McIntyre, 1992). It should be emphasized that the notion of conservation used here is not in the material or Lagrangian sense, but in the traditional, general sense used in theoretical chemistry and physics. In other words, it is used to highlight the fact that the PV behaves like the mixing ratio of a peculiar chemical 'substance', namely the PVS, which has zero source away from boundaries. A chemical substance with zero source means a chemical substance whose molecules are neither created nor destroyed, and then its mixing ratio can change only by transport processes. Eq.(1.5.7) shows that sources and sinks for the PVS can occur only at the ground, and not within the atmosphere itself. Frictional and diabatic terms do not act to create or destroy the PVS, but only help redistribute it.

Another important property of PVS hidden behind Eq.(1.5.7) can be revealed by rewriting the flux $\mathbf{J}$ as

$$\mathbf{J} = \rho P \mathbf{u}_{\theta\perp} + \rho P \mathbf{u}_{\|} - Q\boldsymbol{\zeta}_{\mathrm{a}\|} - \mathbf{F} \times \nabla\theta, \tag{1.5.9}$$

where

$$\mathbf{u}_{\theta\perp} = -\frac{\partial\theta/\partial t}{|\nabla\theta|^2}\nabla\theta\,, \quad \mathbf{u}_{\|} = \mathbf{u} - \frac{\mathbf{u}\cdot\nabla\theta}{|\nabla\theta|^2}\nabla\theta\,, \quad \boldsymbol{\zeta}_{\mathrm{a}\|} = \boldsymbol{\zeta}_{\mathrm{a}} - \frac{\boldsymbol{\zeta}_{\mathrm{a}}\cdot\nabla\theta}{|\nabla\theta|^2}\nabla\theta, \tag{1.5.10}$$

with "$\|$" denoting projection parallel to the local θ-surface, and $\mathbf{u}_{\theta\perp}$ being just the velocity of the θ-surface normal to itself. Thus, the last three terms in Eq.(1.5.9) all represent vectors lying parallel to the local isentropic surface, while the first is just ρP times the normal velocity $\mathbf{u}_{\theta\perp}$ of that surface. Therefore it follows that a point moving with velocity $\mathbf{J}/\rho P$ must always remain on exactly the same isentropic surface. In other words, PVS 'molecules' behave as if they never cross isentropic surfaces, even though air can move across these surfaces in the presence of the diabatic heating. This result is called the impermeability theorem (Haynes and McIntyre, 1987, 1990). An alternative interpretation of this theorem can be provided by noting that the advective PVS transport across the isentropic surface due to the diabatic velocity $\mathbf{u}_{\mathrm{d}\perp} = \mathbf{u} - \mathbf{u}_{\theta\perp} - \mathbf{u}_{\|}$ is cancelled exactly by the term $(-\boldsymbol{\zeta}_{d\perp}Q)$, where $\boldsymbol{\zeta}_{\mathrm{a}\perp} = \boldsymbol{\zeta}_{\mathrm{a}} - \boldsymbol{\zeta}_{\mathrm{a}\|}$ is normal to the θ-surface. This cancellation is manifest, because

$$\mathbf{u}_{\mathrm{d}\perp} \equiv \mathbf{u} - \mathbf{u}_{\theta\perp} - \mathbf{u}_{\|} = \frac{Q}{|\nabla\theta|^2}\nabla\theta, \tag{1.5.11}$$

$$\boldsymbol{\zeta}_{\mathrm{a}\perp} \equiv \boldsymbol{\zeta}_{\mathrm{a}} - \boldsymbol{\zeta}_{\mathrm{a}\|} = \frac{\rho P \nabla\theta}{|\nabla\theta|^2}\,, \tag{1.5.12}$$

so that

$$\rho P \mathbf{u}_{\mathrm{d}\perp} - Q\boldsymbol{\zeta}_{\mathrm{a}\perp} \equiv 0. \tag{1.5.13}$$

A useful alternative form of the conservation equation (1.5.7) is the Lagrangian form

$$\frac{DP}{Dt} = \rho^{-1}\nabla\cdot(\,Q\boldsymbol{\zeta}_{\mathrm{a}} + \mathbf{F}\times\nabla\theta\,) = \rho^{-1}\boldsymbol{\zeta}_{\mathrm{a}}\cdot\nabla Q + \rho^{-1}(\nabla\times\mathbf{F})\cdot\nabla\theta, \tag{1.5.14}$$

which follows from substituting the mass-conservation equation (1.5.3) into Eq.(1.5.7) and performing a little manipulation (see Ertel, 1942; Truesdell, 1951; Haynes and McIntyre, 1987, 1990). Eq.(1.5.14) shows that PV is materially conserved if diabatic heating and nonconservative forces are negligible. This result is known as Ertel's theorem.

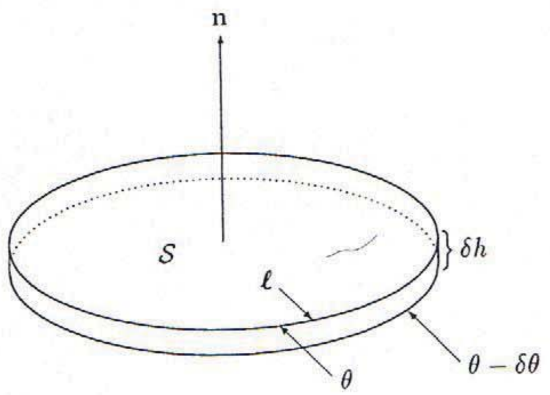


Figure 1.4: A cylindrical fluid parcel bounded by two nearby neighbouring isentropic surfaces, θ and $\theta - \delta\theta$ with a small distance δh apart, and a lateral material surface. ℓ is the intersection of the lateral material surface with the isentropic surface θ. The area enclosed by ℓ is denoted by $\mathcal{S}$.

Ertel's theorem has a very simple physical meaning. To see this, consider a cylindrical fluid parcel bounded by two nearby neighbouring isentropic surfaces, say θ and $\theta - \delta\theta$, and a lateral material surface, as shown in Fig.1.4. Assume the fluid motion is inviscid and adiabatic. Now let ℓ be the intersection of the lateral material surface with the isentropic surface θ. A kinematic relation between the absolute circulation along ℓ, say C_{a}, and the absolute vorticity $\boldsymbol{\zeta}_{\mathrm{a}}$ is given by Stokes' theorem

$$C_{\mathrm{a}} = \oint_{\ell} (2\mathbf{\Omega} \times \mathbf{x} + \mathbf{u}) \cdot d\boldsymbol{\ell} = \int\!\!\int_{\mathcal{S}} (2\mathbf{\Omega} + \nabla \times \mathbf{u}) \cdot \mathbf{n} dS = \int\!\!\int_{\mathcal{S}} \boldsymbol{\zeta}_{\mathrm{a}} \cdot \mathbf{n} dS, \qquad (1.5.15)$$

where $\mathbf{x}$ is the position vector, n is the unit vector normal to the isentropic surface θ, $\mathcal{S}$ denotes the area enclosed by ℓ, and dS denotes an element of area. Obviously, ℓ in this case is a material contour on the isentropic surface. Thus, C_{a} is materially conserved by Kelvin's circulation theorem. For the case of an infinitesimal cylinder (both δh and $\mathcal{S}$ arbitrarily small), $C_{\mathrm{a}} = \boldsymbol{\zeta}_{\mathrm{a}} \cdot \mathbf{n}\mathcal{S}$. In addition, the mass in the cylinder is given by $m = \rho \mathcal{S} \delta h$ (δh is the height of the cylinder). Thus, it follows that

$$\lim_{\delta h \to 0} \frac{C_{\mathrm{a}} \delta\theta}{m} = \lim_{\delta h \to 0} \frac{1}{\rho} \boldsymbol{\zeta}_{\mathrm{a}} \cdot \mathbf{n} \frac{\delta\theta}{\delta h} = \frac{1}{\rho} \boldsymbol{\zeta}_{\mathrm{a}} \cdot \nabla\theta = P. \qquad (1.5.16)$$

Since C_{a}, m, and $\delta\theta$ are all conserved following the motion in a frictionless, adiabatic fluid, P will be conserved also, and then Ertel's theorem follows (Gill, 1982; Hoskins, 1991).

The above interpretation shows how the PV combines the dynamics and thermodynamics into a single equation of great usefulness in a large class of problems having meteorological and oceanographical interest. The reasons are the conservation properties just discussed, and the invertibility principle mentioned in Section 1.1. According to the invertibility principle, the PV field can be inverted to yield the velocity field and any other relevant dynamical information (Hoskins *et al.*, 1985). Note that motions such obtained are called *balanced motions*, referring to the fact, as was said in Section 1.1, that the invertibility principle depends on a 'suitable balance condition'. Mathematical subtleties are involved here, but one way to described 'balance' is the approximate absence of acoustic and inertia-gravity waves. Thus, balanced motions are vortical motions, with the implication that some approximation is unavoidable (McIntyre and Norton, 1995; Ford and McIntyre, 1995 and references therein). McIntyre and Norton (1990,1995) pointed out, however, that in practice the approximations involved are much smaller than might be imagined from the standard approximate inversion theories which restrict attention to Mach, Froude and/or Rossby numbers much less than unity.

The invertibility principle with its property of accuracy greatly simplifies the prognostic aspects of the atmospheric motion. The problem now reduces to timestepping the distribution of PV on each isentropic surface, the distribution of potential temperature at the lower boundary, and the mass under each isentropic surface; otherwise it would involve all three velocity components and a thermodynamic variable (McIntyre, 1992).

1.6 Scope of the thesis

As already mentioned, this study focuses on some fundamentals of wave–mean interaction necessary for understanding the middle atmospheric circulation. An outline of the organisation of the rest of the thesis is the following: Chapter 2 describes a simple thought experiment to highlight the usefulness of the description of wave–mean interaction in terms of wave-induced PVS transport. The same thought experiment, which focuses attention on the mean motion induced by inertia-gravity waves, was examined by McIntyre (1980a) and Andrews (1980) from a perspective of the wave-induced momentum transport. Our major objective is to show how, and at what order of accuracy, the PV description of the problem is consistent with the momentum description. Many ideas and methods used in the subsequent chapters are presented in this chapter.

Chapter 3 examines in further detail the general relationship between the descriptions of wave–mean interaction in terms of momentum transport and in terms of PVS transport. Both integral and differential properties are discussed. Some general identities linking the wave-induced PVS transport to the rate of dissipation of quasimomentum are derived. These identities generalize the well-known identity obtained by Taylor (1915) to the general geophysical fluid system.

In Chapter 4 we turn to the study of Rossby wave dynamics within the framework of quasi-geostrophic theory. Particular attention is paid to the EP flux divergences induced by steady, nonbreaking, dissipating Rossby waves. Also discussed is the difference between the the *transformed Eulerian-mean* and the *generalized Lagrangian-mean* meridional circulations. It will be shown that the difference is not always negligible in the middle atmosphere.

Chapter 5 continues the detailed examination of the ideas introduced in Chapter 4, looking at a highly idealized model for the polar vortex in the winter stratosphere. The aim of this chapter is to highlight the dynamical effects of Rossby waves associated with the polar vortex. Discussion presented in this chapter has direct relevance to current question about the effectiveness of the Rossby-wave restoring mechanism in inhibiting chemical transport across the edge of the polar vortex marked by steep PV gradients on isentropic surfaces, which is believed to be crucial to the formation of the Antarctic ozone hole (McIntyre, 1989).

Chapter 6 summarises the main conclusions obtained in this thesis.

CHAPTER 2

DESCRIPTION OF WAVE-INDUCED MEAN MOTIONS IN TERMS OF WAVE-INDUCED TRANSPORT OF PV-SUBSTANCE: A THOUGHT EXPERIMENT

2.1 Introduction

It is possible, according to the invertibility principle, to describe the global circulation entirely in terms of the source, sink, and transport of PV-substance (PVS). The main point is the idea that, as well as being easy to visualise, distributions of PV on isentropic surfaces and of potential temperature on boundaries contain nearly all the information about the dynamics of fluid motion, apart from any inertia-gravity oscillations that may be present.

This chapter describes a simple thought experiment, in which perhaps the simplest methodology for applying the invertibility principle is demonstrated. Our main attempt is to show how, and at what order of accuracy, the wave-induced mean motion can be described in terms of the wave-induced PVS transport. Although some detailed calculations involved are somewhat cumbersome, the nature of our thought experiment is, in fact, very simple. It concerns only the mean flows induced by linear, dissipating inertia-gravity waves, without taking into account the modification by the mean-flow configuration of the wave motions. It is, therefore, a self-consistent problem only when the wave-induced mean flows remain small in the sense that their effect on the wave fields is always negligible.

A general description of the thought experiment is given in Section 2.2. In Section 2.3 some useful approximations are introduced, and the set of equations to be used in later sections is written in its nondimensional form. Section 2.4 is devoted to the study of inertia-gravity waves. The waves are assumed to be generated by a corrugated bottom boundary and to be weakly dissipated in some layer away from the bottom boundary. Wave solutions are calculated from the linearised equations by utilising the asymptotic expansion method. It is the mean effect of these dissipating inertia-gravity waves that form the main concern of the thought experiment.

Section 2.5 returns to the discussion of wave-induced PVS transport. Detailed calculations of the PV anomalies induced by the dissipating inertia-gravity waves are carried out, and gradient-independent PVS transport is demonstrated in this section. The wave-induced balanced mean motions associated with the PVS anomalies are calculated in Section 2.6, where the application of the invertibility principle is taken up from both the coarse-grain and fine-grain perspectives. The wave-induced mean motions in the Eulerian-mean and the generalised Lagrangian-mean frameworks are considered in Section 2.7. Some conclusions are presented in Section 2.8.

2.2 Description of the thought experiment

Consider a stably-stratified fluid in a Cartesian coordinate system (x, y, z), where x, y, and z are respectively the eastward, northward, and vertical distances to the origin of coordinates, which is chosen on the Earth's surface with some reference latitude and longitude. The velocity components in the x-, y-, and z-directions are denoted respectively by u, v, and w, i.e., $\mathbf{u} = (u, v, w)$. For simplicity without loss of generality, we assume that the fluid is contained in a channel between two rigid vertical walls at $y = 0$, b and a moving lower boundary $z = h(x, y, t)$; it is unbounded above (as $z \to \infty$). The fluid is initially in an undisturbed, zonally symmetric, steady, hydrostatic equilibrium state described by

$$\mathbf{u} = 0, \quad \rho = \rho_{\rm B}(z), \quad p = p_{\rm B}(z), \quad T = T_{\rm B}(z), \quad \theta = \theta_{\rm B}(z), \tag{2.2.1}$$

where

$$\frac{dp_{\rm B}}{dz} = -\rho_{\rm B} g, \quad T_{\rm B} = \frac{p_{\rm B}}{R\rho_{\rm B}}, \quad \theta_{\rm B} = \frac{p_{\rm B}}{R\rho_{\rm B}} \left(\frac{p_{\rm s}}{p_{\rm B}}\right)^{\kappa}, \tag{2.2.2}$$

with g being the magnitude of the gravity acceleration. $\rho_{\rm B}$, $p_{\rm B}$, $T_{\rm B}$, and $\theta_{\rm B}$ can be thought of as respectively the density, pressure, temperature and potential temperature of the "background".

At $t = 0$ some waves are generated by a corrugated bottom boundary moving parallel to the x axis with constant velocity c. The waves are imagined to be dissipated in a layer $\mathcal{D}$ above the bottom boundary, as suggested schematically in Fig.2.1. Thus, according to the generalised Charney–Drazin theorem discussed in Section 1.2, a wave-drag force exerted by the boundary on the fluid is felt in the layer $\mathcal{D}$. This force acts to

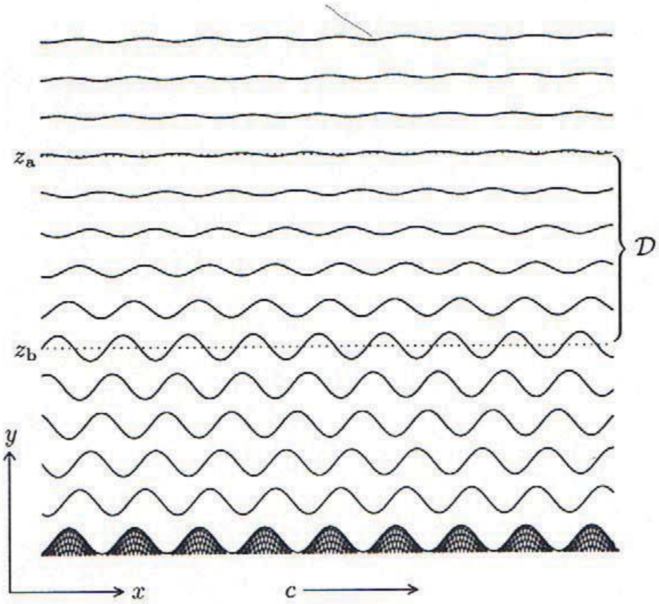


Figure 2.1: Schematic illustration of waves generated in a resting, stably-stratified fluid by a corrugated bottom boundary moving parallel to the x axis with constant velocity c, propagating upward and dissipated in the layer $\mathcal{D}$ (i.e. $z_b < z < z_a$) above the bottom boundary.

accelerate the fluid there in the direction of the phase velocity c. As a result, a mean flow, with some appropriate definition of "mean", is induced. It should be noted that in a stratified, rotating fluid, such as the middle atmosphere, changes in velocity field are always companied by changes in temperature field.

There are, at least, two routes by which the wave-induced motions in the above problem can be approached. Traditionally, the problem is analyzed by evaluating directly the wave-induced momentum and heat transport. More precisely, the mean velocity and temperature fields can be obtained by substituting the wave-induced mean force and diabatic heating into the mean momentum and thermodynamic equations respectively, and then integrating the resulting equations to obtain the mean velocity and temperature fields. This approach has been discussed intensively by several authors under various circumstances (e.g., Matsuno and Nakamura, 1979; McIntyre, 1980a; Andrews, 1980; Uryu, 1980; Takahashi and Uryu, 1981). These works provide some simple yet beautiful illustrations of the well-known ability of waves to transport momentum between regions of wave generation and regions of wave dissipation.

Wave-induced mean motions can also be approached from the viewpoint of wave-induced PVS transport, as suggested by McIntyre and Norton (1990). Such an approach carries with it an implication that the wave-induced mean motions are approximately balanced motions that can be obtained by applying the invertibility principle. McIntyre and Norton (*op. cit.*) argued that the robust parts of the dissipative type of wave-induced mean motions are controlled by PV evolution, insensitive to what kind of mean is taken, and cumulatively much larger than the reversible, dissipation-independent parts of the mean motions that are uniformly bounded by $O(a^2)$ and depend on the continued presence of the waves and on the choice of averaging operator. Thus the general nature of the dissipative type of wave-induced mean motions can be understood by viewing the phenomena in terms of wave-induced PVS transport along the isentropic surfaces. More precisely, wave-induced mean motions can be examined by evaluating the PV distribution from the PV evolution equation, and then applying the invertibility principle to deduce the mean velocity field and other relevant dynamical information.

To make the foregoing ideas more concrete and explicit, in what follows we shall confine attention to one of the simplest possible models, in which dissipating inertia-gravity waves are the primary agents for the forcing of mean flow. The wave-induced mean motions will be derived from both the viewpoints of wave-induced momentum transport and PVS transport. Before proceeding with the detailed formulation, we shall introduce some commonly used approximations to simplify the discussion.

2.3 Approximations to the governing equations and nondimensionalization

The fundamental equations governing the atmospheric motion have been outlined in Section 1.5. They are the momentum equation (1.5.1), the thermodynamic energy equation (1.5.2), the mass conservation equation (1.5.3), and the equation of state (1.5.4). These equations, however, are far more complicated than necessary for description of the thought experiment considered in this chapter. Therefore some appropriate approximation becomes essential. This section introduces three commonly used approximations, the most important being the Boussinesq approximation.

2.3.1 *The Boussinesq approximation*

The Boussinesq approximation was introduced independently by Oberbeck (1879) and Boussinesq (1903). It consists essentially of neglecting the departures of density from the background density except insofar as they are coupled with gravity to give rise to a buoyancy force. Originally, this approximation was developed for a stratified, incompressible fluid. It was extended by Spiegel and Veronis (1960) to apply to certain flows of compressible fluids, in which the vertical length scale, D say, of the motion is very small compared to the scale height, H say, of the reference state. Mathematically, the Boussinesq approximation is satisfactorily applicable in an compressible fluid only when the following condition

$$\frac{D}{H} \ll 1 \tag{2.3.1}$$

is satisfied.

This chapter is concerned with the nature of mean motion induced by dissipating inertia-gravity waves. In the middle atmosphere, H is about 7 km. On the other hand, observational studies have shown that the vertical wavelengths of inertia-gravity waves detected in the middle atmosphere range from 1 km to 15 km (Sawyer, 1961; Hearth *et al.*, 1974; Balsley and Garello, 1985; Hirota and Niki, 1985, 1986; Ushimaru and Tanaka, 1990; Wilson *et al.*, 1991; Nakamura *et al.*, 1993). This implies that the vertical length scale of these waves, which equals to the wavelength divided by 2π, is small enough to satisfy Eq.(2.3.1). Under such circumstances, Eqs.(1.5.1)–(1.5.5) reduce to the following Boussinesq equations (for the detailed derivation, see Appendix 2A):

$$\frac{D\mathbf{u}}{Dt} + f\hat{\mathbf{z}} \times \mathbf{u} - \Theta\hat{\mathbf{z}} + \nabla\Pi = \mathbf{F}, \tag{2.3.2}$$

$$\frac{D\Theta}{Dt} + N^2 w = \mathcal{J}, \tag{2.3.3}$$

$$\nabla\cdot\mathbf{u} = 0, \tag{2.3.4}$$

$$\frac{\delta\rho}{\rho_B} = -\frac{\delta\theta}{\theta_B}, \tag{2.3.5}$$

where

$$f = 2\Omega\sin\phi, \tag{2.3.6}$$

$$N^2 = \frac{g}{\theta_B}\frac{d\theta_B}{dz} = \frac{\kappa g}{H}, \tag{2.3.7}$$

and

$$\Pi = \frac{\delta p}{\rho_B}, \quad \Theta = \frac{g\delta\theta}{\theta_B}, \quad \mathcal{J} = \frac{gQ}{\theta_B}, \tag{2.3.8}$$

with ϕ being the latitude on the Earth, $\delta\rho$, δp, and $\delta\theta$ representing departures from the background density ρ_{B}, background pressure p_{B}, and background potential temperature θ_{B}, respectively. Note that ρ_{B}, p_{B}, and θ_{B} represent the hydrostatic equilibrium state described by Eqs.(2.2.1) and (2.2.2), and it is assumed that

$$\delta\rho \ll \rho_{\mathrm{B}}, \quad \delta p \ll p_{\mathrm{B}}, \quad \delta\theta \ll \theta_{\mathrm{B}}. \tag{2.3.9}$$

Incidentally, f and Θ are referred to as the Coriolis parameter and the buoyancy acceleration, respectively.

In the spirit of the Boussinesq approximation, the PV defined by Eq.(1.1.1) can be approximated as

$$P = \frac{1}{\rho}\boldsymbol{\zeta}_{\mathrm{a}} \cdot \nabla\theta \approx \frac{\theta_{\mathrm{B}}}{g\rho_{\mathrm{B}}}\left(f\hat{\mathbf{z}} + \nabla\times\mathbf{u}\right)\cdot\left(N^2\hat{\mathbf{z}} + \nabla\Theta\right), \tag{2.3.10}$$

and the PV equation (1.5.14) can be approximated as

$$\frac{DP}{Dt} = \nabla\cdot\left\{\frac{\theta_{\mathrm{B}}}{g}\left[\mathcal{J}\boldsymbol{\zeta}_{\mathrm{a}} + \mathbf{F}\times\left(\mathcal{N}^2\hat{\mathbf{z}} + \nabla\Theta\right)\right]\right\}. \tag{2.3.11}$$

Eqs.(2.3.4), (2.3.2), (2.3.3), (2.3.11), whose nondimensional counterparts will be given in §2.3.3, form the basic equations for the motions considered in this chapter.

2.3.2 *f-plane approximation*

As mentioned already, the vertical Coriolis force can be consistently neglected in the relevant equations. The horizontal Coriolis force is represented by the the Coriolis parameter f defined by Eq.(2.3.4). In general, f is a function of latitude ϕ. In this chapter, we shall concentrate exclusively on extratropical flows for which the meridional extent of the motion is so small that the effect of the variation of the Coriolis parameter with latitude is not dynamically important. Under such circumstances, the horizontal can be taken to be a plane surface that has a fixed inclination to the rotation vector, and f can be regarded as constant. These assumptions are called the f-plane approximation.

2.3.3 *Non-dimensional equations and the small turbulent dissipation assumption*

In Eqs.(2.3.2) and (2.3.3), the effects of dissipation are represented by the viscous force $\mathbf{F}$ and the diabatic heating $\mathcal{J}$, respectively. From the molecular viewpoint, the viscous force results from momentum exchanges in collisions between the molecules. It should also include the intermolecular forces. A detailed discussion of the this force was given, e.g., by Batchelor (1967). It turns out that if changes of the *dynamic viscosity coefficient* and compressibility effects can be ignored, the viscous force can be expressed as

$$\mathbf{F} = \nu \nabla^2 \mathbf{u}, \tag{2.3.12}$$

where ν is the kinematic viscosity coefficient. $\mathbf{F}$ given by Eq.(2.3.12) obviously results from a net downgradient transport of momentum by random motion of the molecules.

It can be shown that, for the atmosphere below 100 km, ν is so small that the molecular viscous force is invariably negligible on atmospheric scales (Pedlosky, 1987; Holton, 1992). From the practical viewpoint, however, the viscous force in the atmosphere may arise from momentum transport due to turbulent flows.[1] Observations in the atmosphere suggest that the atmospheric turbulence exists on almost every scale. Therefore, even when observations are taken with very short temporal and spatial separations, some smaller-scale turbulent flows will still be unobservable. One common notion, being first put forward by Boussinesq (1877), is that the smaller-scale motions act to smooth and mix fluid properties (e.g., momentum, heat, etc.) on the motions with larger scales, which are the focus of our interest, by processes analogous to molecular, diffusive transport. Therefore one way to estimate the dissipative influence of turbulent motions is to retain Eq.(2.3.12), but replace ν by a *turbulent viscosity* K_{m} of much larger magnitude than the molecular value, namely

$$\mathbf{F} = K_{\mathrm{m}} \nabla^2 \mathbf{u}. \tag{2.3.13}$$

In this chapter, this concept of turbulent viscosity will be adopted. Further discussion of its validity and limitations can be seen, e.g., in Pedlosky (1987) and Brown (1991).

In a similar fashion, the thermal dissipation $\mathcal{J}$ due to turbulent heat transport is modelled by

$$\mathcal{J} = K_{\mathrm{h}} \nabla^2 \Theta, \tag{2.3.14}$$

[1]In general, the main characteristic of turbulent motion is the random nature of the velocity field in time and space, so that the classical hydrodynamic problem of finding positions of all fluid particles at any instant t from the given positions and velocities of the particles at an initial instant t_0 and boundary conditions becomes meaningless in the case of turbulent motion. In atmospheric applications, however, turbulent flows are usually referred to as the motions that have spatial and temporal scales much smaller than the scales of separation of the observations. These flows may include both the wavy motions and the apparently disorganised motions of turbulence. But since it is impossible to identify these small-scale motions, we have to call them turbulence. Although this is not a very satisfactory definition, it is adequate for practical purposes.

where $K_{\rm h}$ represents the *turbulent diffusivity of heat.* For the present purposes, both $K_{\rm m}$ and $K_{\rm h}$ are assumed to be smoothly varying function of altitude alone, and to be zero outside the dissipative layer $\mathcal{D}$ defined in Fig.2.1. The dissipation due to molecular transport will be neglected.

It is now convenient to introduce non-dimensional coordinates and variables by the relations

$$\left.\begin{array}{lll}(\check{x},\check{y}) = (x,y)/L, & \check{z} = z/D, & \check{t} = t/(L/V),\\ (\check{u},\check{v}) = (u,v)/V, & \check{w} = w/(VD/L), & \check{\zeta}_{\rm a} = \zeta_{\rm a}/(2\Omega),\\ \check{\Pi} = \Pi/(2\Omega VL), & \check{\Theta} = \Theta/(2\Omega VL/D), & \check{P} = P/(8\Omega^3 T_{\rm c} L^2/g\rho_{\rm s} D^2),\\ \check{f} = f/(2\Omega), & \check{\rho}_{\rm B} = \rho_{\rm B}/\rho_{\rm s}, & \check{\theta}_{\rm B} = \theta_{\rm B}/T_{\rm c},\\ \check{\mathbf{F}} = \mathbf{F}/(V^2/L), & \check{\mathcal{J}} = \mathcal{J}/(2\Omega V^2/D), & (\check{K}_{\rm m},\check{K}_{\rm h}) = (K_{\rm m},K_{\rm h})/K_0,\end{array}\right\} \quad (2.3.15)$$

where L, D, V, and K_0 are characteristic scales of horizontal length, vertical length, horizontal velocity, and turbulent viscosity (or turbulent diffusivity of heat), respectively. Substituting Eqs.(2.3.13)–(2.3.15) into Eqs.(2.3.2), (2.3.3), and (2.3.4), and, for convenience without risk of confusion, dropping the checks (ˇ) from the non-dimensional coordinates and variables, we readily obtain the following non-dimensional equations of motion

$$\mathcal{R}\frac{Du}{Dt} - fv + \frac{\partial \Pi}{\partial x} = \mu\mathcal{R}K_{\rm m}\check{\nabla}^2 u, \quad (2.3.16)$$

$$\mathcal{R}\frac{Dv}{Dt} + fu + \frac{\partial \Pi}{\partial y} = \mu\mathcal{R}K_{\rm m}\check{\nabla}^2 v, \quad (2.3.17)$$

$$\mathcal{R}\iota^2\frac{Dw}{Dt} - \Theta + \frac{\partial \Pi}{\partial z} = \mu\,\iota^2\mathcal{R}K_{\rm m}\check{\nabla}^2 w, \quad (2.3.18)$$

$$\mathcal{R}\frac{D\Theta}{Dt} + \mathcal{N}^2 w = \mu\mathcal{R}K_{\rm h}\check{\nabla}^2\Theta, \quad (2.3.19)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \quad (2.3.20)$$

where $\mathcal{R}$, ι, $\mathcal{N}$, and μ are dimensionless parameters defined by

$$\mathcal{R} = \frac{V}{2\Omega L}, \qquad \iota = \frac{D}{L}, \qquad \mathcal{N} = \frac{ND}{2\Omega L}, \qquad \mu = \frac{K_0 L}{VD^2}; \quad (2.3.21)$$

and $\check{\nabla}^2$ is the dimensionless Laplacian operator defined by

$$\check{\nabla}^2 = \iota^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \frac{\partial^2}{\partial z^2}. \quad (2.3.22)$$

The non-dimensional version of Eq.(2.3.11) is

$$\rho_{\rm B}\frac{DP}{Dt} = \mathcal{R}\check{\nabla}\cdot\left\{\theta_{\rm B}\left[\mathcal{J}\zeta_{\rm a} + \iota^{-1}\mathbf{F}\times\left(\mathcal{N}^2\hat{\mathbf{z}} + \mathcal{R}\check{\nabla}\Theta\right)\right]\right\}, \quad (2.3.23)$$

where P, ζ_a, $\mathbf{F}$, and $\mathcal{J}$ are all non-dimensional variables, defined respectively by

$$P = \frac{\theta_B}{\rho_B}\left[\left(f\hat{\mathbf{z}} + \check{\nabla}\times\mathbf{u}\right)\cdot\left(\mathcal{N}^2\hat{\mathbf{z}} + \mathcal{R}\check{\nabla}\Theta\right)\right], \tag{2.3.24}$$

$$\zeta_a = f\hat{\mathbf{z}} + \check{\nabla}\times\mathbf{u}, \tag{2.3.25}$$

$$\mathbf{F} = \mu K_m\left(\check{\nabla}^2 u,\ \check{\nabla}^2 v,\ \iota\check{\nabla}^2 w\right), \tag{2.3.26}$$

$$\mathcal{J} = \mu K_h\check{\nabla}^2\Theta. \tag{2.3.27}$$

and the non-dimensional operator $\check{\nabla}$ is defined by

$$\check{\nabla}\cdot\mathbf{A} = \iota\left(\frac{\partial A_{(x)}}{\partial x} + \frac{\partial A_{(y)}}{\partial y}\right) + \frac{\partial A_{(z)}}{\partial z}, \tag{2.3.28}$$

$$\check{\nabla}\Theta = \iota\left(\frac{\partial\Theta}{\partial x}\hat{\mathbf{x}} + \frac{\partial\Theta}{\partial y}\hat{\mathbf{y}}\right) + \frac{\partial\Theta}{\partial z}\hat{\mathbf{z}}, \tag{2.3.29}$$

$$\check{\nabla}\times\mathbf{A} = \left(\iota\frac{\partial A_{(z)}}{\partial y} - \frac{1}{\iota}\frac{\partial A_{(y)}}{\partial z}\right)\hat{\mathbf{x}} + \left(\frac{1}{\iota}\frac{\partial A_{(x)}}{\partial z} - \iota\frac{\partial A_{(z)}}{\partial x}\right)\hat{\mathbf{y}} + \left(\frac{\partial A_{(y)}}{\partial x} - \frac{\partial A_{(x)}}{\partial y}\right)\hat{\mathbf{z}}. \tag{2.3.30}$$

Note that ι is the ratio of the vertical and horizontal length scales. If $\iota^2 \ll 1$, then Eq.(2.3.18) reduces to the hydrostatic equation. For gravity wave motions, the vertical length scale is not necessarily much less than the horizontal length scale, so that the hydrostatic relation may not always hold. μ is a parameter measuring the magnitude of turbulent dissipation. The square of $\mathcal{N}$ is sometimes called the Burger number. $\mathcal{R}$ is the Rossby number. For large-scale atmospheric motions $\mathcal{R}$ is usually a small number and then the motions are characterised by geostrophic balance between the Coriolis force and the pressure gradient force in the horizontal momentum equations. For inertia-gravity wave motions, on the other hand, $\mathcal{R}$ is $O(1)$. In this chapter we are concerned with the mean motions induced by dissipating inertia-gravity waves. Therefore for simplicity without loss of generality, we shall let

$$\mathcal{R} = 1. \tag{2.3.31}$$

It should be further noted that, although the turbulent viscosity and diffusivity of heat are much larger than their molecular counterparts, the direct effects of mechanical and thermal dissipation associated with much smaller-scale turbulence on the large-scale atmospheric flow are, more often than not, completely negligible away from the planetary boundary layer. However, in the theory of wave–mean interaction turbulent dissipation nevertheless always plays an important role in damping the gravity waves and then leading to violation of the Charney–Drazin nonacceleration theorem. Therefore it is necessary to include the turbulent dissipation terms in the equations for gravity wave motions, though these terms may still be small in comparison with inertial and pressure-gradient forces. In Eqs.(2.3.16)–(2.3.19), the smallness of the turbulent dissipation is measured by the dimensionless parameter μ. In our thought experiment, inertia-gravity waves are assumed to be weakly dissipated by turbulent processes, so that

$$\mu \ll 1. \tag{2.3.32}$$

In the following analysis, μ will serve as the small parameter, in powers of which solutions of the problem considered can be solved asymptotically.

2.4 The linearised inertia-gravity waves

In this section linear, dissipating inertia-gravity waves in the f-plane channel model are examined. We shall separate each flow quantity into an x-averaged part (with some appropriate definition of "average") and a disturbance part. All disturbance quantities are taken to be $O(a)$, where $a \ll 1$. Furthermore, all wave-induced mean quantities must be $O(a^2)$ or smaller. A demonstration of the self-consistency of these conditions on the disturbances and the mean fields can be found in Andrews *et al.* (1987).

In our thought experiment, the fluid is initially in an undisturbed, zonally symmetric, steady, hydrostatic equilibrium state that is described by Eqs.(2.2.1) and (2.2.2). Thus, it follows that

$$(u, v, w, \Theta, \Pi) = (u', v', w', \Theta', \Pi') + O(a^2). \tag{2.4.1}$$

Substitution of Eq.(2.4.1) into Eqs.(2.3.16)–(2.3.20) leads to the following set of linearised equations for inertia-gravity waves:

$$\frac{\partial u'}{\partial t} - fv' + \frac{\partial \Pi'}{\partial x} = \mu K_{\mathrm{m}} \check{\nabla}^2 u', \tag{2.4.2}$$
$$\frac{\partial v'}{\partial t} + fu' + \frac{\partial \Pi'}{\partial y} = \mu K_{\mathrm{m}} \check{\nabla}^2 v', \tag{2.4.3}$$
$$\iota^2 \frac{\partial w'}{\partial t} - \Theta' + \frac{\partial \Pi'}{\partial z} = \mu\, \iota^2 K_{\mathrm{m}} \check{\nabla}^2 w', \tag{2.4.4}$$
$$\frac{\partial \Theta'}{\partial t} + \mathcal{N}^2 w' = \mu K_{\mathrm{h}} \check{\nabla}^2 \Theta', \tag{2.4.5}$$
$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} + \frac{\partial w'}{\partial z} = 0. \tag{2.4.6}$$

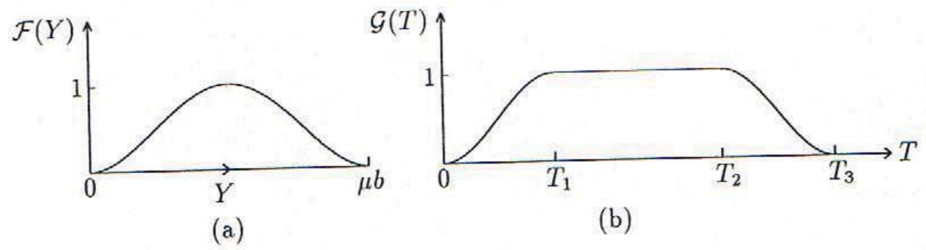


Figure 2.2: $\mathcal{F}(Y)$ and $\mathcal{G}(T)$.

The initial condition is that all disturbances are zero, namely

$$(\ )' = 0 \quad \text{for} \quad t < 0. \tag{2.4.7}$$

At $t = 0$ a moving topographic disturbance, $h' = O(a)$, is slowly switched on. The specific form of h' is assumed to be

$$h' = a\mathcal{F}(Y)\mathcal{G}(T)\sin(kx - \sigma t), \tag{2.4.8}$$

where T and Y represent respectively the 'slow time' and 'slow length' variables defined by

$$Y = \mu y\,, \qquad T = \mu t\,, \qquad (\mu \ll 1). \tag{2.4.9}$$

Similarly, a 'slow height' variable Z is also defined by

$$Z = \mu z, \tag{2.4.10}$$

and, because of the condition (2.3.1), the dimensionless ρ_{B} and θ_{B} may be expressed as

$$\rho_{\text{B}} = \exp(-2n_H Z), \qquad \theta_{\text{B}} = \exp(2\kappa n_H Z), \tag{2.4.11}$$

where n_H is an $O(1)$ dimensionless parameter defined by

$$\frac{D}{2H} \equiv \mu n_H \ll 1. \tag{2.4.12}$$

$\mathcal{F}$ and $\mathcal{G}$, depicted schematically in Fig.2.2, are smoothly varying functions of Y and T, respectively. k is the zonal wavenumber, and σ the frequency of the sinusoidal topographic corrugations which move with phase speed $c = \sigma/k$. For later reference, $\mathcal{F}^2(Y)$ is expanded in the Fourier series as

$$\mathcal{F}^2(Y) = \sum_{n=1}^{\infty} M_n \sin(\ell_n Y), \tag{2.4.13}$$

where

$$\ell_n = \frac{n\pi}{\mu b}, \qquad M_n = \frac{2}{\mu b}\int_0^{\mu b} \mathcal{F}^2(Y)\sin(\ell_n Y)dY. \tag{2.4.14}$$

The boundary conditions are

$$u' = v' = w' = 0 \qquad \text{on } y = 0, b \text{ within the layer } \mathcal{D}; \tag{2.4.15}$$

$$v' = 0 \qquad \text{on } y = 0, b \text{ outside the layer } \mathcal{D}; \tag{2.4.16}$$

$$w' = \frac{\partial h'}{\partial t} \qquad \text{on } z = 0; \tag{2.4.17}$$

$$(\,)' \to 0 \qquad \text{as } z \to \infty \ (t < \infty). \tag{2.4.18}$$

Substitution of Eq.(2.4.8) into Eq.(2.4.17) gives

$$\begin{aligned} w' = &-a\sigma\mathcal{F}(Y)\mathcal{G}(T)\cos(kx - \sigma t) \\ &+\mu a\mathcal{F}(Y)\frac{\partial}{\partial T}[\mathcal{G}(T)]\sin(kx - \sigma t), \quad \text{on } z = 0. \end{aligned} \tag{2.4.19}$$

Now let us seek solutions to the linearised equations Eqs.(2.4.2)–(2.4.6) in the form of asymptotic expansions with respect to the small parameter μ, namely

$$(u', v', w', \Theta', \Pi') = \sum_{n=0}^{\infty} (u'_n, v'_n, w'_n, \Theta'_n, \Pi'_n)\, \mu^n, \tag{2.4.20}$$

where $\{u'_n, v'_n, w'_n, \cdots\}$ are all $O(a)$ and independent of the small parameter μ. Furthermore, we assume that

$$(u'_n, v'_n, w'_n, \Theta'_n, \Pi'_n) = \mathrm{Re}\left\{\left(\widehat{u}_n, \widehat{v}_n, \widehat{w}_n, \widehat{\Theta}_n, \widehat{\Pi}_n\right)\exp\left[i(kx + mz - \sigma t)\right]\right\}, \tag{2.4.21}$$

where m is a real vertical wavenumber and the complex amplitudes $\{\widehat{u}_n, \widehat{v}_n, \widehat{w}_n, \cdots\}$ are functions only of Y, T, and Z.

Substituting Eq.(2.4.20) with (2.4.21) into Eqs.(2.4.2)–(2.4.6) and equating the coefficients of μ^0 yields

$$-i\sigma\widehat{u}_0 - f\widehat{v}_0 + ik\widehat{\Pi}_0 = 0, \tag{2.4.22}$$

$$-i\sigma\widehat{v}_0 + f\widehat{u}_0 = 0, \tag{2.4.23}$$

$$-i\iota^2\sigma\widehat{w}_0 - \widehat{\Theta}_0 + im\widehat{\Pi}_0 = 0, \tag{2.4.24}$$

$$-i\sigma\widehat{\Theta}_0 + \mathcal{N}^2\widehat{w}_0 = 0, \tag{2.4.25}$$

$$ik\widehat{u}_0 + im\widehat{w}_0 = 0, \tag{2.4.26}$$

which lead to the following identities

$$\widehat{u}_0 = -\frac{m}{k}\widehat{w}_0, \quad \widehat{v}_0 = \frac{ifm}{\sigma k}\widehat{w}_0, \quad \widehat{\Theta}_0 = -\frac{i\mathcal{N}^2}{\sigma}\widehat{w}_0, \quad \widehat{\Pi}_0 = -\frac{m(\sigma^2 - f^2)}{\sigma k^2}\widehat{w}_0, \tag{2.4.27}$$

and the dispersion relation

$$\sigma^2 = \frac{\mathcal{N}^2 k^2 + f^2 m^2}{\iota^2 k^2 + m^2}. \tag{2.4.28}$$

In the atmosphere, the dimensional buoyancy frequency usually exceeds the dimensional Coriolis parameter by a large factor. This corresponds to the condition $\iota^{-1}\mathcal{N} \gg 1$. In addition, $f = \sin\phi \leq 1$. Therefore Eq.(2.4.28) indicates that the existence of propagating waves, with both k and m real, requires that

$$|f| \leq |\sigma| \leq \iota^{-1}\mathcal{N}. \tag{2.4.29}$$

The vertical group velocity can be obtained from Eq.(2.4.28) as,

$$c_{gz} = \frac{\partial \sigma}{\partial m} = -\frac{m(\sigma^2 - f^2)}{\sigma(\iota^2 k^2 + m^2)}. \tag{2.4.30}$$

Noticing the facts that σ/m represents vertical phase velocity and σ satisfies Eq.(2.4.29), we see from Eq.(2.4.30) that wave phase and wave energy propagate in opposite vertical directions.

To determine $\widehat{w}_0$ we must proceed to the $O(\mu^1)$ equations. Substituting Eq.(2.4.20) with (2.4.21) into Eqs.(2.4.2)–(2.4.6) and equating the coefficients of μ^1 yields

$$-i\sigma\widehat{u}_1 - f\widehat{v}_1 + ik\widehat{\Pi}_1 = -\left[K_{\mathrm{m}}\left(\iota^2 k^2 + m^2\right)\widehat{u}_0 + \frac{\partial \widehat{u}_0}{\partial T}\right], \tag{2.4.31}$$

$$-i\sigma\widehat{v}_1 + f\widehat{u}_1 = -\left[K_{\mathrm{m}}\left(\iota^2 k^2 + m^2\right)\widehat{v}_0 + \frac{\partial \widehat{v}_0}{\partial T} + \frac{\partial \widehat{\Pi}_0}{\partial Y}\right], \tag{2.4.32}$$

$$-i\sigma\iota^2\widehat{w}_1 - \widehat{\Theta}_1 + im\widehat{\Pi}_1 = -\iota^2\left[K_{\mathrm{m}}\left(\iota^2 k^2 + m^2\right)\widehat{w}_0 + \frac{\partial \widehat{w}_0}{\partial T}\right] - \frac{\partial \widehat{\Pi}_0}{\partial Z}, \tag{2.4.33}$$

$$-i\sigma\widehat{\Theta}_1 + \mathcal{N}^2\widehat{w}_1 = -\left[K_{\mathrm{h}}\left[\iota^2 k^2 + m^2\right)\widehat{\Theta}_0 + \frac{\partial \widehat{\Theta}_0}{\partial T}\right], \tag{2.4.34}$$

$$ik\widehat{u}_1 + im\widehat{w}_1 = -\frac{\partial \widehat{v}_0}{\partial Y} - \frac{\partial \widehat{w}_0}{\partial Z}. \tag{2.4.35}$$

From Eqs.(2.4.31)–(2.4.35) we obtain the following equation for $\widehat{w}_0$:

$$\frac{\partial \widehat{w}_0}{\partial T} + c_{gz}\left[\frac{\partial \widehat{w}_0}{\partial Z} + \Lambda(Z)\widehat{w}_0\right] = 0, \tag{2.4.36}$$

where $\Lambda(Z)$ is defined by

$$\Lambda(Z) = K_{\mathrm{m}}(Z)\frac{(\mathcal{N}^2k^2 + 2f^2m^2)}{2\sigma^2 c_{gz}} + K_{\mathrm{h}}(Z)\frac{\mathcal{N}^2k^2}{2\sigma^2 c_{gz}}. \tag{2.4.37}$$

Using the lower boundary condition (2.4.19), we obtain the solution of Eq.(2.4.36) as

$$\widehat{w}_0 = -a\sigma\mathcal{F}(Y)\mathcal{H}(Z)\mathcal{G}(\widehat{T}), \tag{2.4.38}$$

where

$$\mathcal{H}(Z) \equiv \exp\left[-\int_0^Z \Lambda(Z)dZ\right], \tag{2.4.39}$$

$$\widehat{T} \equiv T - Z/c_{gz}. \tag{2.4.40}$$

Applying the upper boundary condition (2.4.18) to Eq.(2.4.38) shows that

$$\Lambda(Z) \geq 0. \tag{2.4.41}$$

Thus, from Eqs.(2.4.37) and (2.4.30) it follows that

$$c_{gz} \geq 0, \qquad \text{namely } \operatorname{sgn}(m) = -\operatorname{sgn}(\sigma). \tag{2.4.42}$$

Eq.(2.4.42) indicates that the wave phase has a downward component of propagation ($\sigma/m < 0$), in contrast to the upward energy propagation ($c_{gz} \geq 0$).

Now a particle displacement function $\boldsymbol{\xi}' = (\xi', \eta', \zeta')$ can be defined by (Andrews and McIntyre, 1978b; Andrews, 1980)

$$\frac{\partial \boldsymbol{\xi}'}{\partial t} = \mathbf{u}' + O(a^2). \tag{2.4.43}$$

From Eqs.(2.4.6) and (2.4.7) it follows that

$$\nabla\cdot\boldsymbol{\xi}' = 0. \tag{2.4.44}$$

By analogy with Eqs.(2.4.20) and (2.4.21) we let

$$\boldsymbol{\xi}' = \sum_{n=0}^{\infty}(\xi'_n, \eta'_n, \zeta'_n)\mu^n = \mathrm{Re}\left\{(\widehat{\xi}_0, \widehat{\eta}_0, \widehat{\zeta}_0)\exp\left[i(kx + mz - \sigma t)\right]\right\} + O(\mu a), \tag{2.4.45}$$

where $\left\{\widehat{\xi}_0, \widehat{\eta}_0, \widehat{\zeta}_0\right\}$ depend on (Y, Z, T) only. From Eqs.(2.4.43) and (2.4.27) it follows that

$$\widehat{\xi}_0 = -\frac{im}{\sigma k}\widehat{w}_0, \qquad \widehat{\eta}_0 = -\frac{fm}{\sigma^2 k}\widehat{w}_0, \qquad \widehat{\zeta}_0 = \frac{i}{\sigma}\widehat{w}_0. \tag{2.4.46}$$

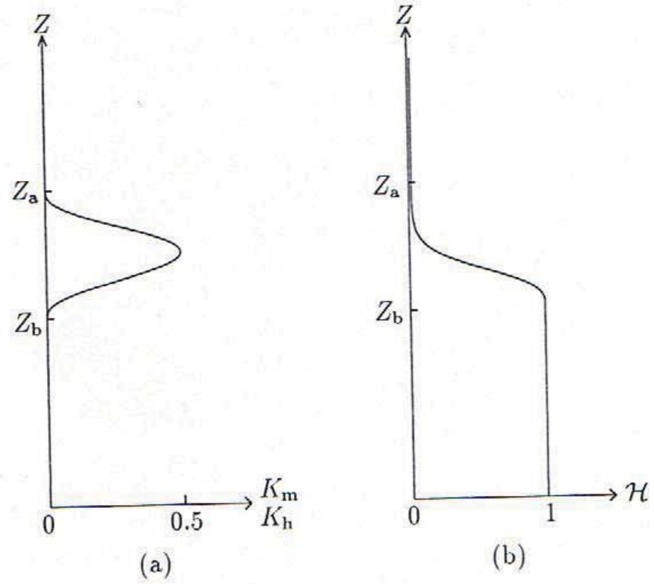


Figure 2.3: (a) The dimensionless (turbulent) viscosity $K_m(Z)$ and diffusivity of heat $K_h(Z)$ specified in the thought experiment. For simplicity we assume that $K_m(Z) = K_h(Z)$. (b) The amplitude function $\mathcal{H}(Z)$ of an inertia-gravity wave, whose horizontal and vertical wavelengths are, respectively, 200 km and 1.5 km. It is assumed in this case that $\phi = 45°$N and $N = 0.02\,s^{-1}$.

As an example, a mid-latitude ($\phi = 45°$N) inertia-gravity wave is considered in our thought experiment. The dimensional horizontal and vertical wavelengths of this wave are 200 km and 1.5 km, respectively. For simplicity we shall let $K_m(Z) = K_h(Z)$ and specify their vertical structure as that shown in Fig.2.3a. Then the amplitude function $\mathcal{H}(Z)$ for the specified wave is shown in Fig.2.3b. As expected, the wave is damped in the dissipation layer (i.e. $Z_b < Z < Z_a$). Its dynamical effect on the mean flow will be examined in later sections.

2.5 PVS redistribution due to the wave dissipation

Now let us turn to the detailed formulation of the thought experiment described in Section 2.2. As a first step towards the description of the wave-induced mean motions in terms of PVS transport, in this section we shall first consider how the PVS is transported by dissipating inertia-gravity waves.

2.5.1 The generalised Lagrangian-mean PV distribution

We start with a consideration of the PVS budget in a material tube $\mathcal{V}$ of fluid enclosed by a surface $\mathcal{S}$, where $\mathcal{S}$ is specified to consist of four material surfaces $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$, as shown in Fig.2.4. Initially, $\mathcal{V}$ lies along an arbitrary latitude, and $\mathcal{S}_2$ and $\mathcal{S}_4$ are specified to coincide with two isentropic surfaces θ_{AD} and θ_{BC}, respectively. We shall consider sufficiently small amplitudes to prevent wave breaking. It is understood that the wave dissipation is capable of inducing a Lagrangian mean meridional circulation, which, in principle, may lead to the permanent migration of the material tube $\mathcal{V}$ away from its original position. But it will turn out that such wave-induced mean meridional circulation is uniformly bounded by $O(\mu a^2)$ (Section 2.7). Thus we may assume that the wave dissipation is weak enough ($\mu \ll 1$) to ensure that the material tube remains approximately in its original position throughout the thought experiment.

The total amount of PVS in the tube at instant t is defined by

$$\mathcal{P}(t;Y,Z) = \iiint\limits_{\mathcal{V}} \rho_{\mathrm{B}} P d\tau, \tag{2.5.1}$$

where the coordinates Y and Z serve to indicate the average position of the tube, and $d\tau$ is an element volume. If both the viscous force and diabatic heating are absent, Ertel's theorem says that PV is materially conserved. Under such circumstances, the total amount of PVS within the material tube $\mathcal{V}$ must be also unchanged, i.e., $\mathcal{P}(t;Y,Z) = \mathcal{P}(0;Y,Z)$ =const., provided that waves do not break.

In our thought experiment, however, the inertia-gravity waves are assumed to be dissipated within the layer $\mathcal{D}$. Therefore Ertel's theorem can be expected to be violated and the total PVS within the material tube $\mathcal{V}$ can change by PVS transfer across the material surface $\mathcal{S}$ due to the wave dissipation. Differentiation of Eq.(2.5.1) with respect to time t and use of the mass-conservation equation (1.5.2) gives (see Batchelor, 1967)

$$\frac{d}{dt}\left[\mathcal{P}(t;Y,Z)\right] = \frac{d}{dt}\iiint\limits_{\mathcal{V}} \rho_{\mathrm{B}} P d\tau = \iiint\limits_{\mathcal{V}} \rho_{\mathrm{B}} \frac{DP}{Dt} d\tau. \tag{2.5.2}$$

Substitution of Eq.(2.3.23) into Eq.(2.5.2) and use of the divergence theorem (and $\mathcal{R} = 1$) leads to

$$\begin{aligned}\frac{d}{dt}\left[\mathcal{P}(t;Y,Z)\right] &= \iiint\limits_{\mathcal{V}} \check{\nabla}\cdot\left\{\theta_{\mathrm{B}}\left[\mathcal{J}\boldsymbol{\zeta}_{\mathrm{a}} + \iota^{-1}\mathbf{F}\times\left(\mathcal{N}^2\hat{\mathbf{z}} + \check{\nabla}\Theta\right)\right]\right\} d\tau \\ &= \iint\limits_{\mathcal{S}} \left\{\theta_{\mathrm{B}}\left[\mathcal{J}\boldsymbol{\zeta}_{\mathrm{a}} + \iota^{-1}\mathbf{F}\times\left(\mathcal{N}^2\hat{\mathbf{z}} + \check{\nabla}\Theta\right)\right]\right\}\cdot\hat{\mathbf{n}} dS, \end{aligned} \tag{2.5.3}$$

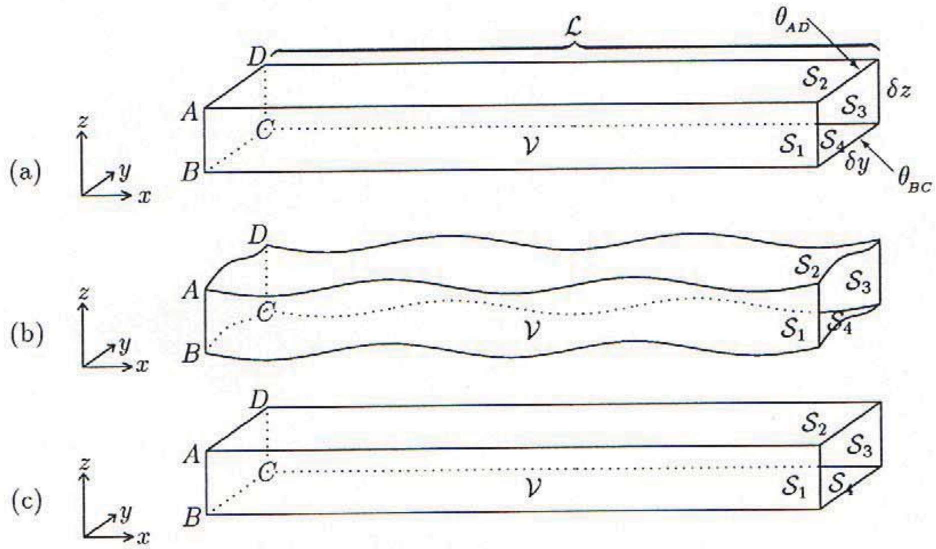


Figure 2.4: A closed material tube $\mathcal{V}$ of fluid specified to be bounded by four material surfaces $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$. The length, width, and height of $\mathcal{V}$ are denoted by $\mathcal{L}$, δy, and δz, respectively. The coordinates of four reference particles A, B, C, and D are denoted by (x_A, y_A, z_A), (x_B, y_B, z_B), (x_C, y_C, z_C), and (x_D, y_D, z_D), respectively. (a) Initially, $\mathcal{V}$ lies along an arbitrarily latitude, and $\mathcal{S}_2$ and $\mathcal{S}_4$ are specified to coincide with two isentropic surfaces θ_{AD} and θ_{BC}, respectively. In this state $(t < 0)$, there is no disturbance anywhere. (b) The undulation of the material tube when the waves arrive from below. (c) The waves fade down to nothing in the final state.

where $\hat{\mathbf{n}}$ denotes the outward unit vector normal to the surface of the material tube and dS is an element area.

Integrating Eq.(2.5.3) with respect to time t gives

$$\mathcal{P}(t; Y, Z) = \mathcal{P}(0; Y, Z) + \delta\mathcal{P}(t; Y, Z), \tag{2.5.4}$$

where $\delta\mathcal{P}$ represents the increment of PVS in the tube $\mathcal{V}$, given by

$$\begin{aligned}\delta\mathcal{P}(t; Y, Z) &= -\int_0^t \left\{ \iint\limits_{\mathcal{S}} \theta_{\mathrm{B}} \left[\mathcal{J}\boldsymbol{\zeta}_{\mathrm{a}} + \iota^{-1}\mathbf{F} \times \left(\mathcal{N}^2\hat{\mathbf{z}} + \check{\nabla}\Theta \right) \right] \cdot \hat{\mathbf{n}} dS \right\} dt \\ &= \int_0^t \left\{ \left(\iint\limits_{\mathcal{S}_1} + \iint\limits_{\mathcal{S}_2} + \iint\limits_{\mathcal{S}_3} + \iint\limits_{\mathcal{S}_4} \right) \theta_{\mathrm{B}} \left[\mathcal{J}\boldsymbol{\zeta}_{\mathrm{a}} + \iota^{-1}\mathbf{F} \times \left(\mathcal{N}^2\hat{\mathbf{z}} + \check{\nabla}\Theta \right) \right] \cdot \hat{\mathbf{n}} dS \right\} dt. \end{aligned} \tag{2.5.5}$$

In addition, the total mass in the tube $\mathcal{V}$ is

$$\mathcal{M} = \iiint\limits_{\mathcal{V}} \rho_{\mathrm{B}} d\tau. \tag{2.5.6}$$

Note that, since $\mathcal{V}$ is a material tube and the waves in question are assumed not to break, $\mathcal{M}$ given by Eq.(2.5.6) should be independent of time.

Now an average of P weighted by mass along the material tube $\mathcal{V}$ may be defined by

$$\overline{P}^{\dagger} = \mathcal{P}(t; Y, Z)/\mathcal{M}. \tag{2.5.7}$$

It should be noticed that the material tube $\mathcal{V}$ was initially straight and uniform (Fig.2.4a), and has been distorted by the waves into the shape shown in Fig.2.4b. Therefore in the limit $\delta y \equiv (y_{D} - y_{A}) \to 0$ and $\delta z \equiv (z_{A} - y_{B}) \to 0$, $\overline{P}^{\dagger}$ defined by Eq.(2.5.7) is equivalent to the *generalised Lagrangian mean* (GLM) of P, which is usually denoted by $\overline{P}^{\mathrm{L}}$ with $\overline{(\,)}^{\mathrm{L}}$ representing a GLM quantity. For the concept and detailed mathematical theory of the GLM formulation, see Andrews and McIntyre (1978b,c). The fact that

$$\overline{P}^{\mathrm{L}} \equiv \lim_{\substack{\delta y \to 0 \\ \delta z \to 0}} \overline{P}^{\dagger} \tag{2.5.8}$$

was pointed out by McIntyre (1980a,b). Since the tube $\mathcal{V}$ is arbitrarily chosen, the overall distribution of $\overline{P}^{\mathrm{L}}$ can be evaluated by using Eqs.(2.5.8) and (2.5.7), provided that $\mathcal{P}$ and $\mathcal{M}$ are obtainable from Eqs.(2.5.4)-(2.5.6).

To carry out the surface integral in Eq.(2.5.5), we need first to calculate the outward element area vector $\hat{\mathrm{n}} dS$ normal to the material surfaces $\mathcal{S}_1$, $\mathcal{S}_2$, $\mathcal{S}_3$, and $\mathcal{S}_4$. Suppose these surfaces are given parametrically by the equations

$$x = x(r_{\mathrm{p}}, s_{\mathrm{p}}), \qquad y = y(r_{\mathrm{p}}, s_{\mathrm{p}}), \qquad z = z(r_{\mathrm{p}}, s_{\mathrm{p}}), \tag{2.5.9}$$

where r_{p} and s_{p} are parameter variables, ranging over some domains of the $(r_{\mathrm{p}}, s_{\mathrm{p}})$-plane. Thus $\hat{\mathrm{n}} dS$ can be given by (e.g., see Bronshtein and Semendyayev, 1979)

$$\begin{aligned}\hat{\mathrm{n}} dS = \pm \Bigg[\hat{\mathrm{x}} \left(\frac{\partial y}{\partial r_{\mathrm{p}}} \frac{\partial z}{\partial s_{\mathrm{p}}} - \frac{\partial z}{\partial r_{\mathrm{p}}} \frac{\partial y}{\partial s_{\mathrm{p}}} \right) + \hat{\mathrm{y}} \left(\frac{\partial z}{\partial r_{\mathrm{p}}} \frac{\partial x}{\partial s_{\mathrm{p}}} - \frac{\partial x}{\partial r_{\mathrm{p}}} \frac{\partial z}{\partial s_{\mathrm{p}}} \right) \\ + \hat{\mathrm{z}} \left(\frac{\partial x}{\partial r_{\mathrm{p}}} \frac{\partial y}{\partial s_{\mathrm{p}}} - \frac{\partial y}{\partial r_{\mathrm{p}}} \frac{\partial x}{\partial s_{\mathrm{p}}} \right) \Bigg] dr_{\mathrm{p}} ds_{\mathrm{p}},\end{aligned} \tag{2.5.10}$$

where the sign must be specified to guarantee $\hat{\mathrm{n}}$ to be an outward vector normal to the surface.

Now let us fix attention to a material particle arbitrarily specified on $\mathcal{S}$. The position of this particle is denoted by $\tilde{\mathbf{x}} + \boldsymbol{\xi}'(\tilde{\mathbf{x}}, t)$, where $\boldsymbol{\xi}'$ is the particle displacement vector defined by Eq.(2.4.43), and $\tilde{\mathbf{x}}$ the position vector on a suitably defined line parallel to the x axis and moving with the GLM velocity $\overline{\mathbf{u}}^{\mathrm{L}}$ (see Andrews and McIntyre, 1978b). Thus, at the leading order in μ we have

$$x = \tilde{x} + \mathrm{Re}\left\{\hat{\xi}_0(\tilde{Y}, \tilde{Z}) \exp\left[i(k\tilde{x} + m\tilde{z} - \sigma t)\right]\right\} + O(\mu a), \tag{2.5.11}$$

$$y = \tilde{y} + \mathrm{Re}\left\{\hat{\eta}_0(\tilde{Y}, \tilde{Z}) \exp\left[i(k\tilde{x} + m\tilde{z} - \sigma t)\right]\right\} + O(\mu a), \tag{2.5.12}$$

$$z = \tilde{z} + \mathrm{Re}\left\{\hat{\zeta}_0(\tilde{Y}, \tilde{Z}) \exp\left[i(k\tilde{x} + m\tilde{z} - \sigma t)\right]\right\} + O(\mu a), \tag{2.5.13}$$

where $\tilde{Y} = \mu\tilde{y}$, $\tilde{Z} = \mu\tilde{z}$. Clearly Eqs.(2.5.11)–(2.5.13) can be regarded as the parametric equations representing the material surfaces $\mathcal{S}_1$ and $\mathcal{S}_3$ when $\tilde{y}$ is specified as a constant and $\tilde{x}$, $\tilde{z}$ as the parameter variables, and representing the material surfaces $\mathcal{S}_2$ and $\mathcal{S}_4$ when $\tilde{z}$ is specified as a constant and $\tilde{x}$, $\tilde{y}$ as the parameter variables (see Fig.2.4). Therefore, substitution of Eqs.(2.5.11)–(2.5.13) into Eq.(2.5.10) and use of Eqs.(2.4.20), (2.4.27), and (2.4.46) will lead to

$$\hat{\mathbf{n}} dS = -\left\{\iota\frac{kv_0'}{\sigma}\hat{\mathbf{x}} + \iota\left(1 - \frac{ku_0'}{\sigma} - \frac{mw_0'}{\sigma}\right)\hat{\mathbf{y}} + \frac{mv_0'}{\sigma}\hat{\mathbf{z}} + O(a^2, \mu a)\right\} d\tilde{x}d\tilde{z}, \quad \text{on } \mathcal{S}_1; \tag{2.5.14}$$

$$\hat{\mathbf{n}} dS = \left\{\frac{kw_0'}{\sigma}\hat{\mathbf{x}} + \left(1 - \frac{ku_0'}{\sigma}\right)\hat{\mathbf{z}} + O(a^2, \mu a)\right\} d\tilde{x}d\tilde{y}, \quad \text{on } \mathcal{S}_2; \tag{2.5.15}$$

$$\hat{\mathbf{n}} dS = \left\{\iota\frac{kv_0'}{\sigma}\hat{\mathbf{x}} + \iota\left(1 - \frac{ku_0'}{\sigma} - \frac{mw_0'}{\sigma}\right)\hat{\mathbf{y}} + \frac{mv_0'}{\sigma}\hat{\mathbf{z}} + O(a^2, \mu a)\right\} d\tilde{x}d\tilde{z}, \quad \text{on } \mathcal{S}_3; \tag{2.5.16}$$

$$\hat{\mathbf{n}} dS = -\left\{\frac{kw_0'}{\sigma}\hat{\mathbf{x}} + \left(1 - \frac{ku_0'}{\sigma}\right)\hat{\mathbf{z}} + O(a^2, \mu a)\right\} d\tilde{x}d\tilde{y}, \quad \text{on } \mathcal{S}_4. \tag{2.5.17}$$

Assume that all wave-induced mean quantities depend on (Y, Z, T) only. In addition, all these quantities are $O(a^2)$, as mentioned in Section 2.4. Thus it follows that

$$\boldsymbol{\zeta}_{\mathrm{a}} = f\hat{\mathbf{z}} + \check{\nabla} \times \mathbf{u}_0' + O(\mu a^2), \tag{2.5.18}$$

$$\check{\nabla}\Theta = \check{\nabla}\Theta_0' + O(\mu a^2), \tag{2.5.19}$$

$$\mathcal{J} = \mu K_{\mathrm{h}} \check{\nabla}^2\Theta_0' + O(\mu^2 a), \tag{2.5.20}$$

$$\mathbf{F} = \mu K_{\mathrm{m}} \left(\check{\nabla}^2 u_0', \, \check{\nabla}^2 v_0', \, \iota\check{\nabla}^2 w_0'\right) + O(\mu^2 a). \tag{2.5.21}$$

Now we are able to calculate the surface integral in Eq.(2.5.5) by using Eqs.(2.5.14)–(2.5.21). For example, in the limit of $\delta z \to 0$,

$$\iint_{\mathcal{S}_1} \theta_{\mathrm{B}} \left[\mathcal{J}\boldsymbol{\zeta}_{\mathrm{a}} + \mathbf{F} \times \left(\mathcal{N}^2\hat{\mathbf{z}} + \check{\nabla}\Theta\right)\right] \cdot \hat{\mathbf{n}} dS$$

$$
\begin{aligned}
&= -\mu \int_{z_B}^{z_A} \int_0^{\mathcal{L}} \theta_{\mathrm{B}} \left\{ K_{\mathrm{m}} \left[\frac{k\mathcal{N}^2}{\sigma} v_0' \check{\nabla}^2 v_0' + \left(\frac{\mathcal{N}^2}{\sigma}(ku_0' + mw_0') - \mathcal{N}^2 - \frac{\partial \Theta_0'}{\partial z} \right) \iota^2 \check{\nabla}^2 u_0' \right.\right. \\
&\qquad \left.\left. + \frac{\partial \Theta_0'}{\partial x} \check{\nabla}^2 w_0' \right] + K_{\mathrm{h}} \left(\frac{fm}{\sigma} v_0' + \frac{\partial u_0'}{\partial z} - \iota^2 \frac{\partial w_0'}{\partial x} \right) \check{\nabla}^2 \Theta_0' + O(a^3, \mu a^2) \right\} d\tilde{x} d\tilde{z} \\
&= -\frac{\mu m \mathcal{N}^2 (\sigma^2 - f^2)}{k\sigma^2} \int_{z_B}^{z_A} \theta_{\mathrm{B}} \left[\Lambda(\tilde{Z}) \hat{w}_0^2 (Y_A, \tilde{Z}, T) + O(a^3, \mu a^2) \right] \mathcal{L} d\tilde{z} \\
&= -\mu \left\{ \frac{a^2 m \mathcal{N}^2 (\sigma^2 - f^2)}{k} \theta_{\mathrm{B}} \Lambda(Z) \left[\mathcal{F}(Y_A) \mathcal{H}(Z) \mathcal{G}(\hat{T}) \right]^2 + O(a^3, \mu a^2) \right\} \mathcal{L} \delta z. \qquad (2.5.22)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
&\iint_{\mathcal{S}_3} \theta_{\mathrm{B}} \left[\mathcal{J} \boldsymbol{\zeta}_{\mathrm{a}} + \mathbf{F} \times \left(\mathcal{N}^2 \hat{\mathbf{z}} + \check{\nabla} \Theta \right) \right] \cdot \hat{\mathbf{n}} dS \\
&\quad = \mu \left\{ \frac{a^2 m \mathcal{N}^2 (\sigma^2 - f^2)}{k} \theta_{\mathrm{B}} \Lambda(Z) \left[\mathcal{F}(Y_D) \mathcal{H}(Z) \mathcal{G}(\hat{T}) \right]^2 + O(a^3, \mu a^2) \right\} \mathcal{L} \delta z, \qquad (2.5.23)
\end{aligned}
$$

and in the limit $\delta y \to 0$,

$$
\left(\iint_{\mathcal{S}_2} + \iint_{\mathcal{S}_4} \right) \theta_{\mathrm{B}} \left[\mathcal{J} \boldsymbol{\zeta}_{\mathrm{a}} + \mathbf{F} \times \left(\mathcal{N}^2 \hat{\mathbf{z}} + \check{\nabla} \Theta \right) \right] \cdot \hat{\mathbf{n}} dS = \left[O(\mu a^3, \mu^2 a^2) \right] \theta_{\mathrm{B}} \mathcal{L} \delta y. \qquad (2.5.24)
$$

Note that for adiabatic motions the material surfaces $\mathcal{S}_2$ and $\mathcal{S}_4$ should always coincide with isentropic surfaces θ_{AD} and θ_{BC}, respectively (see Fig.2.4). Under such circumstances, the impermeability theorem ensures that there is no net PVS transport across $\mathcal{S}_2$ and $\mathcal{S}_4$, so that the surface integral in Eq.(2.5.24) should be zero exactly. Thus the $O(\mu a^3, \mu^2 a^2)$ term on the right-hand side of Eq.(2.5.24) must be due to diabatic processes. Eqs.(2.5.22)–(2.5.24) indicate that for weakly dissipating waves, the net PVS transport across $\mathcal{S}_2$ and $\mathcal{S}_4$ is negligible compared with the transport across $\mathcal{S}_1$ and $\mathcal{S}_3$.

Substituting Eqs.(2.5.22)–(2.5.24) into Eq.(2.5.5), correct to $O(\mu a^2)$ we obtain

$$
\begin{aligned}
\delta \mathcal{P}(t; Y, Z) &= \frac{\mu a^2 m \mathcal{N}^2 (\sigma^2 - f^2)}{k} \theta_{\mathrm{B}} \Lambda(Z) \mathcal{H}^2(Z) \left[\mathcal{F}^2(Y_D) - \mathcal{F}^2(Y_A) \right] \mathcal{L} \delta z \int_0^t \mathcal{G}^2(\hat{T}) dt \\
&= \frac{\mu a^2 m \mathcal{N}^2 (\sigma^2 - f^2)}{k} \theta_{\mathrm{B}} \Lambda(Z) \mathcal{H}^2(Z) \frac{d}{dY} \left[\mathcal{F}^2(Y) \right] \mathcal{K}(Z, T) \mathcal{L} \delta y \delta z, \qquad (2.5.25)
\end{aligned}
$$

where

$$
\mathcal{K}(Z, T) = \int_0^T \mathcal{G}^2(\tilde{T} - Z/c_{gz}) d\tilde{T}. \qquad (2.5.26)
$$

In addition, in the limit of $\delta y \to 0$ and $\delta z \to 0$, the total mass in the tube $\mathcal{V}$ is

$$
\mathcal{M} = \iiint_{\mathcal{V}} \rho_{\mathrm{B}} d\tau = \rho_{\mathrm{B}} \mathcal{L} \delta y \delta z, \qquad (2.5.27)
$$

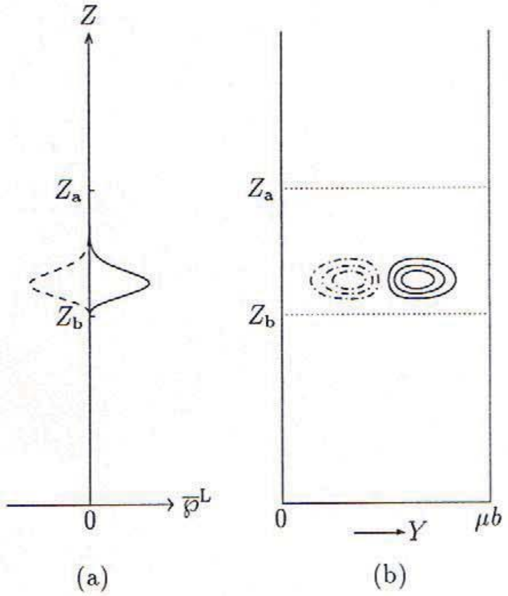


Figure 2.5: (a) Vertical structures of the PV anomalies, namely $\overline{\wp}^L$ defined by Eq.(2.5.32), at $Y = \frac{1}{3}\mu b$ (dashed) and $Y = \frac{2}{3}\mu b$ (solid). (b) Meridional distribution of $\overline{\wp}^L$. Negative contours are represented by dashdot lines and positive contours by solid lines. The zero contour is omitted.

and the initial PVS within the tube $\mathcal{V}$ is

$$\mathcal{P}(0; Y, Z) = \iiint_{\mathcal{V}} \rho_B P_B \, d\tau = \rho_B P_B \mathcal{L} \delta y \delta z, \tag{2.5.28}$$

where P_B is the non-dimensional background PV, defined by

$$P_B(Z) = f \mathcal{N}^2 \theta_B / \rho_B. \tag{2.5.29}$$

Now the GLM PV, defined by Eq.(2.5.8), can be written as

$$\overline{P}^L(Y, Z, T) = P_B(Z) + \frac{\mu a^2 \theta_B}{\rho_B} \Lambda(Z) \mathcal{H}^2(Z) \mathcal{K}(Z, T) \sum_{n=1}^{\infty} P_n \cos(\ell_n Y), \tag{2.5.30}$$

where

$$P_n = \frac{m \mathcal{N}^2 (\sigma^2 - f^2) \ell_n M_n}{k}, \tag{2.5.31}$$

with ℓ_n and M_n being defined by Eq.(2.4.14). Note that in the above calculation, only the leading-order wave solutions in μ are needed in order to obtain $\overline{P}^L$ correct to $O(\mu a^2)$.

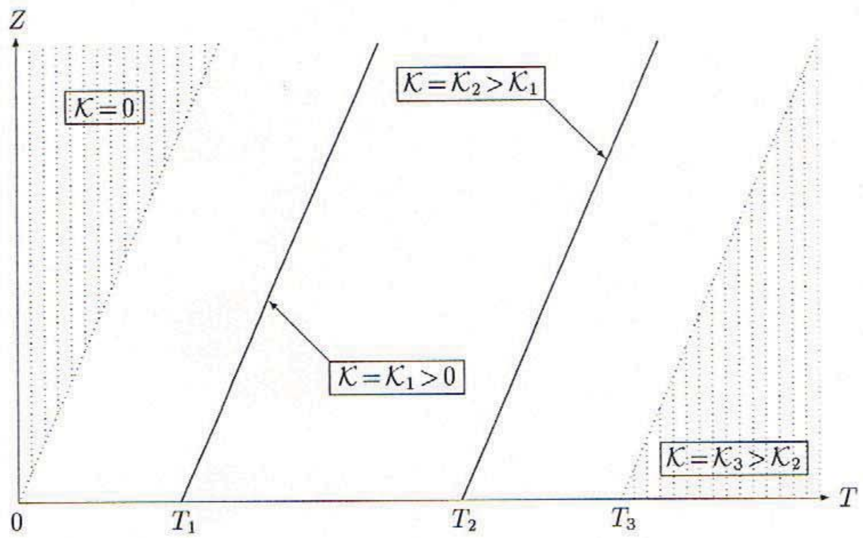


Figure 2.6: Schematic distribution of the function $\mathcal{K}(Z,T)$ defined by Eq.(2.5.26)

It may be convenient to replace the total GLM PV field $\overline{P}^{\mathrm{L}}$ with the deviation from its background value $P_{\mathrm{B}}(Z)$. Thus, we let

$$\overline{\wp}^{\mathrm{L}} = \frac{\rho_{\mathrm{B}}(\overline{P}^{\mathrm{L}} - P_{\mathrm{B}})}{\theta_{\mathrm{B}}} = \mu a^2 \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)\sum_{n=1}^{\infty} P_n \cos(\ell_n Y) \tag{2.5.32}$$

designate the normalised GLM PV anomalies. It is evident from Eq.(2.5.32) that these anomalies are non-zero only within the dissipation layer $\mathcal{D}$, where $\Lambda(Z) > 0$. Outside the layer $\mathcal{D}$, $\Lambda(Z) \equiv 0$, so that $\overline{\wp}^{\mathrm{L}} \equiv 0$, as shown in Fig.2.5.

As mentioned in Chapter 1, the PVS is neither created nor destroyed away from boundaries. Thus the effect of wave dissipation upon the distribution of PV in interior of a fluid can be thought of in terms of the transport of PVS exactly along isentropic surfaces (McIntyre and Norton, 1990; McIntyre, 1992). Fig.2.5 shows a typical dipole pattern of PV anomalies induced by an eastward propagating inertia-gravity wave ($\sigma/k > 0$). The dipole pattern consists of a cyclonic PV anomaly in the northern half and an anticyclonic PV anomaly of compensating strength in the southern half of the channel, indicating a northward PVS transport along the isentropes. It should be noticed incidentally that in our f-plane model the initial PV distribution on each isentrope is spatially uniform, so that the wave-induced PVS transport shown in Fig.2.5b is gradient-independent.

The structure of $\mathcal{K}(Z,T)$, which is defined by Eq.(2.5.26), is schematically depicted in Fig.2.6. At a fixed altitude, $Z = Z_\dagger$ say, $\mathcal{K}$ remains zero for $T \le Z_\dagger/c_{gz}$; it becomes a monotonic increasing function of T when $Z_\dagger/c_{gz} < T < T_3 + Z_\dagger/c_{gz}$ (for the definition of T_3 see Fig.2.2), and achieves its maximum value at the end of the thought experiment, i.e., at $T = T_3 + Z_\dagger/c_{gz}$. This implies that $|\overline{\wp}^L|$ within the dissipation layer $\mathcal{D}$ is a cumulatively increasing quantity in the presence of waves and achieves its maximum value at the end of the thought experiment.

2.5.2 The Eulerian-mean PV distribution

We have already derived the GLM PV distribution. For the purpose of applying the invertibility principle, it may be more convenient to deal with the Eulerian-mean PV, which can be easily interpreted either from a *coarse-grain perspective* as the scalar product of the Eulerian-averaged absolute vorticity and the Eulerian-averaged gradient of potential temperature divided by the background density (Shapiro, 1976, 1978; McIntyre and Palmer, 1983; Keyser and Rotunno, 1990), or from a *fine-grain perspective* as the Eulerian average of the scalar product of the absolute vorticity and the gradient of potential temperature divided by the background density (Danielsen and Hipskind, 1980; Danielsen *et al.*, 1987; Keyser and Rotunno, 1990). For example, when the Eulerian-mean operator $\overline{(\,)}$ is applied to Eq.(2.3.24), we find that

$$\begin{aligned}\overline{P} &= \frac{\theta_{\mathrm{B}}}{\rho_{\mathrm{B}}}\left[\overline{\left(f\hat{\mathbf{z}} + \dot{\nabla}\times\mathbf{u}\right)\cdot\left(\mathcal{N}^2\hat{\mathbf{z}} + \check{\nabla}\Theta\right)}\right] \\ &= P_{\mathrm{B}} + \mu\frac{\theta_{\mathrm{B}}}{\rho_{\mathrm{B}}}\left\{f\frac{\partial\overline{\Theta}}{\partial Z} - \mathcal{N}^2\frac{\partial\overline{u}}{\partial Y} + \frac{\partial}{\partial Y}\left[\overline{\Theta_0'\left(\frac{\partial u_0'}{\partial z} - \iota^2\frac{\partial w_0'}{\partial x}\right)}\right]\right. \\ &\qquad\left. + \frac{\partial}{\partial Y}\left[\overline{\Theta_0'\left(\frac{\partial v_0'}{\partial x} - \frac{\partial u_0'}{\partial y}\right)}\right] + O(\mu^2 a^2)\right\}. \qquad (2.5.33)\end{aligned}$$

In this fine-grain interpretation, $\overline{P}$ is decomposed into a mean contribution that involves only the mean fields, and an eddy contribution that involves Reynolds averages of wave fields. Thus, when P_{B}, $\overline{P}$ and wave solutions are available, with suitable boundary conditions and a specified balance condition, $\overline{u}$ and $\overline{\Theta}$ can be deduced from Eq.(2.5.33), as will be shown in Section 2.6. By contrast, it will be shown in Section 3.2 that the fine-grain interpretation of the GLM PV does not completely separate contributions from Lagrangian-mean and eddy fields under general circumstances, thereby bringing extra difficulties to the PV inversion problem (the barotropic, inviscid motion is an exception). Although for all practical purposes the coarse-grain interpretation of the mean PV (either Eulerian mean or GLM) is meaningful enough for large-scale dynamical motion (McIntyre and Palmer, 1983), for the sake of argument we would also like to examine the consequences produced by employing the fine-grain interpretation of mean PV. This task can be easily achieved in the conventional Eulerian mean formulation.

Unlike its GLM counterpart, the Eulerian-mean formulation involves taking averages over a set of coordinates fixed in (x, y, z)-space. Therefore in principle, $\overline{P}$ can be obtained by evaluating the PVS budget in a tube fixed in the space. But this procedure turns out to be very complicated, involving the advection due to the wave-induced mean meridional circulation, eddy flux, and diffusive transport. In particular, the determination of the net PVS transport across the boundary of the fixed tube involves careful consideration of the large cancellation between the mean advection and the eddy flux. In addition, in order to obtain $\overline{P}$ correct to $O(\mu a^2)$, one needs to know wave solutions correct to $O(\mu a)$, which are not given in Section 2.4 (although, in principle, these solutions can be determined by proceeding to higher order equations, they are expected to be much more complicated than their leading-order counterparts). Since $\overline{P}^{\mathrm{L}}$ is known already, one can evaluate $\overline{P}$ by examining its *Stokes correction* $\overline{P}^{\mathrm{S}}$, which, by definition, is the difference between $\overline{P}^{\mathrm{L}}$ and $\overline{P}$. It will be shown in Section 3.2 that the Stokes correction for an arbitrary field φ can be evaluated from linear wave solutions, as given by Eq.(3.2.10). Note that, in our present model it is assumed that wave amplitudes and mean quantities are all functions of (Y, Z, T) only. Therefore Eq.(3.2.10) reduces to

$$\overline{\varphi}^{\mathrm{S}} = \mu \left[\frac{\partial}{\partial Y} \left(\overline{\eta_0' \varphi_0'} \right) + \frac{\partial}{\partial Z} \left(\overline{\zeta_0' \varphi_0'} \right) \right] + O(a^2 \varphi_0') + O(\mu^2 a^2), \tag{2.5.34}$$

where $\overline{(\;)}$ denotes the Eulerian mean and $\overline{\varphi}^{\mathrm{S}}$ is the Stokes correction to $\overline{\varphi}$. Note that $\overline{\varphi}^{\mathrm{S}}$ is also called *Stokes drift* if φ is a velocity component.

Eq.(2.5.34) shows that all Stokes corrections are $O(\mu a^2, a^2 \varphi_0')$ or smaller in our model. Recalling that $\overline{P}^{\mathrm{L}}$ is of $O(\mu a^2)$ and $P_0' = O(a)$, we obtain the Eulerian-mean PV as

$$\overline{P} = \overline{P}^{\mathrm{L}} - \overline{P}^{\mathrm{S}} = \overline{P}^{\mathrm{L}} - \mu \left[\frac{\partial}{\partial Y} \left(\overline{\eta_0' P_0'} \right) + \frac{\partial}{\partial Z} \left(\overline{\zeta_0' P_0'} \right) \right] + O(a^2 P_0') + O(\mu^2 a^2), \tag{2.5.35}$$

where $\overline{P}^{\mathrm{L}}$ is given by Eq.(2.5.30), and P_0' is the leading order PV disturbance defined by

$$P_0' = \frac{\theta_{\mathrm{B}}}{\rho_{\mathrm{B}}} \left[f \frac{\partial \Theta_0'}{\partial z} + \mathcal{N}^2 \left(\frac{\partial v_0'}{\partial x} - \frac{\partial u_0'}{\partial y} \right) \right]. \tag{2.5.36}$$

Substituting the leading-order wave solutions (2.4.27) into Eq.(2.5.36), we see that

$$P_0' = 0. \tag{2.5.37}$$

Thus Eq.(2.5.35) reduces to

$$\overline{P} = \overline{P}^{\mathrm{L}} + O(\mu^2 a^2). \tag{2.5.38}$$

In other words, the Stokes correction to the Eulerian-mean PV is negligible to the order $O(\mu a^2)$ in our model.

2.6 The wave-induced balanced mean motions

Having obtained the distribution and evolution of the Eulerian-mean PV, we are now ready to apply the invertibility principle to find the wave-induced mean zonal velocity $\overline{u}$ and buoyancy acceleration $\overline{\Theta}$. For this purpose we need further to specify the total mass under each isentropic surface and a suitable balance condition. In addition, suitable boundary conditions should be prescribed.

2.6.1 The mass distribution, balance condition and boundary conditions

In the absence of diabatic heating, the potential temperature is a materially conserved quantity, so that the total mass under each isentrope is constant and then can be determined from the initial reference state. This simplicity may be lost when the effect of diabatic heating is considered. The presence of diabatic heating implies that there exists cross-isentropic transport of mass that may change the mass distribution under each isentropic surface. But it turns out that changes of the reference state due to this diabatic effect can be neglected under the assumption of weak dissipation. Therefore, for simplicity without loss of generality, we shall always assume that the total mass under each isentropic surface can be determined by the background density defined by Eq.(2.3.5).

The balance condition is chosen to be the simplest one, which consists of hydrostatic and geostrophic balance. Therefore the zonal component of velocity and the buoyancy acceleration must satisfy the thermal wind relation

$$f\frac{\partial \overline{u}}{\partial Z} = -\frac{\partial \overline{\Theta}}{\partial Y}. \tag{2.6.1}$$

In our model, the lower boundary coincides initially with an isentropic surface, namely $\theta(x, y, 0, 0) = \theta_{\mathrm{B}}(0)$. This coincidence will last for ever, because the motion near the lower boundary is assumed adiabatic. Therefore the lower boundary condition for our model can be written as

$$\overline{\Theta} = 0 \qquad \text{on} \quad Z = 0. \tag{2.6.2}$$

From Eqs.(2.6.1) and (2.6.2) it follows that

$$\frac{\partial \overline{u}}{\partial Z} = 0 \qquad \text{on} \quad Z = 0. \tag{2.6.3}$$

It is reasonable to suppose that the fluid at infinity ($z \to \infty$) is not forced by the waves excited at $Z = 0$. Therefore the upper boundary conditions are

$$\overline{u},\ \overline{\Theta} \to 0 \qquad \text{as} \quad Z \to \infty. \tag{2.6.4}$$

Since the fluid is assumed to be bounded by two vertical walls, the lateral boundary conditions can be written as

$$\overline{v} = 0, \qquad \text{at} \quad Y = 0, \mu b. \tag{2.6.5}$$

In addition, it can be easily checked that the forcing of the mean zonal flow due to the Reynolds stress associated with the inertia-gravity waves is, correct to $O(a^2)$, zero at the lateral boundaries in our model. Thus, the overall forcing of the mean zonal flow is, again correct to $O(a^2)$, zero at the lateral boundaries. This implies that, to $O(a^2)$,

$$\overline{u} = 0 \qquad \text{at} \quad Y = 0, \mu b; \tag{2.6.6}$$

because the fluid is initially at rest everywhere.

2.6.2 The balanced mean motions from a coarse-grain perspective

Since $\overline{P}$, P_{B}, and wave solutions in our model are all available from the above discussion, with the aid of balance condition and boundary conditions described in §2.6.1 we can deduce $\overline{u}$ and $\overline{\Theta}$ straightforwardly from Eq.(2.5.33), which provides a fine-grain interpretation of $\overline{P}$. In meteorological practice, however, small-scale fluctuations in the vorticity and potential-temperature gradients are not always resolvable by the observational data. Therefore, to a first approximation, the Reynolds averages on the right-hand side of

Eq.(2.5.33) usually have to be ignored. This neglect of eddy contribution is expected not to bring significant error to the inversion problem. As McIntyre and Palmer (1983) pointed out, if some such coarse-grain approximation were not dynamically meaningful, numerical model simulations of the large-scale behaviour of the atmosphere would hardly be practicable. To see how and why the large-scale dynamics survives in the coarse-grain approximation to the mean PV, we now first discuss the invertibility principle from the coarse-grain perspective.

Neglecting the Reynolds averages on the right-hand side of Eq.(2.5.33) and, for convenience, denoting the corresponding mean zonal velocity and buoyancy acceleration as $\tilde{u}$ and $\tilde{\Theta}$, respectively, we obtain

$$\overline{P} = P_{\mathrm{B}} + \mu \frac{\theta_{\mathrm{B}}}{\rho_{\mathrm{B}}} \left(f \frac{\partial \tilde{\Theta}}{\partial Z} - \mathcal{N}^2 \frac{\partial \tilde{u}}{\partial Y} \right), \tag{2.6.7}$$

correct to $O(\mu a^2)$. We further assume that $\tilde{u}$ and $\tilde{\Theta}$ also satisfy the balance condition and boundary conditions given in §2.6.1. In other words, $\tilde{u}$ and $\tilde{\Theta}$ are defined by Eqs.(2.6.1) and (2.6.7) with boundary conditions (2.6.2)–(2.6.4) and (2.6.6), rather than by applying the Eulerian-mean operator directly to u and Θ fields. Now taking $\partial(2.6.7)/\partial Y$ and using Eqs.(2.5.38), (2.5.30), and (2.6.1), after some manipulation we obtain

$$\mathcal{N}^2 \frac{\partial^2 \tilde{u}}{\partial Y^2} + f^2 \frac{\partial^2 \tilde{u}}{\partial Z^2} = a^2 \Lambda(Z) \mathcal{H}^2(Z) \mathcal{K}(Z,T) \sum_{n=1}^{\infty} \ell_n P_n \sin(\ell_n Y). \tag{2.6.8}$$

This is a (non-standardised) Poisson equation. To obtain a solution satisfying the boundary condition (2.6.6), we expand $\tilde{u}$ as

$$\tilde{u}(Y,Z,T) = a^2 \sum_{n=1}^{\infty} U_n(Z,T) \sin(\ell_n Y). \tag{2.6.9}$$

Inserting Eq.(2.6.9) into Eq.(2.6.8) and equating each component, yields

$$\frac{\partial^2 U_n}{\partial Z^2} - \Gamma_n^2 U_n = \frac{\ell_n P_n}{f^2} \Lambda(Z) \mathcal{H}^2(Z) \mathcal{K}(Z,T), \qquad n = 1,2,3,\cdots, \tag{2.6.10}$$

where

$$\Gamma_n = \mathcal{N} \ell_n / f. \tag{2.6.11}$$

By using the method of variation of parameters, the general solution of Eq.(2.6.10) can be written as

$$U_n = \left[A_n(T) + \frac{P_n}{2f\mathcal{N}} \mathcal{U}_{1n}(Z,T) \right] \exp(\Gamma_n Z) + \left[B_n(T) - \frac{P_n}{2f\mathcal{N}} \mathcal{U}_{2n}(Z,T) \right] \exp(-\Gamma_n Z), \tag{2.6.12}$$

where A_n and B_n are two arbitrary functions of T, and $\mathcal{U}_{1n}$ and $\mathcal{U}_{2n}$ are defined by

$$\mathcal{U}_{1n}(Z,T) = \int_0^Z \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)\exp(-\Gamma_n Z)\,dZ, \qquad (2.6.13)$$

$$\mathcal{U}_{2n}(Z,T) = \int_0^Z \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)\exp(\Gamma_n Z)\,dZ. \qquad (2.6.14)$$

The upper boundary condition (2.6.4) requires that

$$A_n(T) = -\frac{P_n}{2f\mathcal{N}}\mathcal{U}_{1n}(\infty,T) = -\frac{P_n}{2f\mathcal{N}}\mathcal{U}_{1n}(Z_a,T). \qquad (2.6.15)$$

The last expression in Eq.(2.6.15) comes from the fact that $\Lambda(Z)=0$ when $Z > Z_a$ (see Fig.2.3).

The lower boundary condition (2.6.3) requires that

$$B_n(T) = A_n(T). \qquad (2.6.16)$$

We now turn to evaluate $\tilde{\Theta}$. Using Eq.(2.6.9), it is a straightforward calculation to obtain $\tilde{\Theta}$ from the integration of Eq.(2.6.1) with respect to Y, namely

$$\tilde{\Theta}(Y,Z,T) = \tilde{\Theta}\Big|_{Y=0} + a^2\sum_{n=1}^{\infty}\frac{f}{\ell_n}\left\{\frac{\partial\tilde{U}_n}{\partial Z}[\cos(\ell_n Y)-1] - \frac{\partial\tilde{U}_n}{\partial Z}\sin(\ell_n Y)\right\}. \qquad (2.6.17)$$

At $Y=0$, Eq.(2.6.7) can be rewritten as

$$\frac{\partial\tilde{\Theta}}{\partial Z}\bigg|_{Y=0} = \frac{\mathcal{N}^2}{f}\frac{\partial\tilde{u}}{\partial Y}\bigg|_{Y=0} + \frac{\overline{\wp}^{\mathrm{L}}}{\mu f}\bigg|_{Y=0}. \qquad (2.6.18)$$

By substitution of Eqs.(2.5.32) and (2.6.9) into Eq.(2.6.18) and integration of the resulting equation with respect to Z, we obtain

$$\tilde{\Theta}\Big|_{Y=0} = a^2\sum_{n=1}^{\infty}\left[\mathcal{N}\Gamma_n\int_0^Z U_n(Z,T)dZ + \frac{P_n}{f}\int_0^Z \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)dZ\right]. \qquad (2.6.19)$$

Noticing that

$$\begin{aligned}&\int_0^Z \mathcal{U}_{1n}(Z,T)\exp(\Gamma_n Z)dZ \\ &\quad= \int_0^Z \exp(\Gamma_n Z)\left[\int_0^Z \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)\exp(-\Gamma_n Z)\,dZ\right]dZ \\ &\quad= \frac{1}{\Gamma_n}\left[\mathcal{U}_{1n}(Z,T)\exp(\Gamma_n Z) - \int_0^Z \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)dZ\right],\end{aligned} \qquad (2.6.20)$$

and, similarly,

$$\begin{aligned}&\int_0^Z \mathcal{U}_{2n}(Z,T)\exp(-\Gamma_n Z)dZ \\ &\quad= -\frac{1}{\Gamma_n}\left[\mathcal{U}_{2n}(Z,T)\exp(-\Gamma_n Z) - \int_0^Z \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)dZ\right],\end{aligned} \qquad (2.6.21)$$

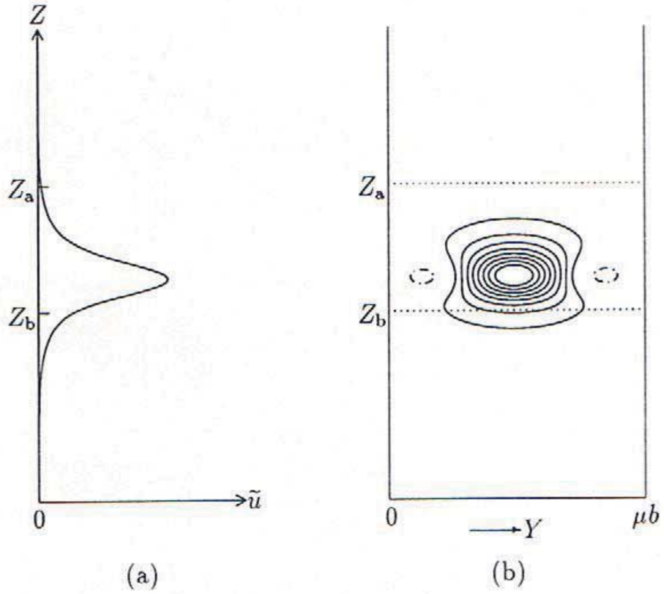


Figure 2.7: (a) Vertical structures of the wave-induced dissipation-dependent mean zonal flow, $\tilde{u}$, at $Y = \frac{1}{2}\mu b$. (b) As in Fig.2.5b, but for $\tilde{u}$.

then we see that

$$\int_0^Z U_n(Z,T)dZ = \frac{1}{\Gamma_n}\left[\mho_{an}(Z,T) - \mho_{bn}(Z,T) - \frac{P_n}{f\mathcal{N}}\int_0^Z \Lambda(Z)\mathcal{H}^2(Z)\mathcal{K}(Z,T)dZ\right], \quad (2.6.22)$$

where

$$\mho_{an}(Z,T) = \left[A_n(T) + \frac{P_n}{2f\mathcal{N}}\mathcal{U}_{1n}(Z,T)\right]\exp(\Gamma_n Z), \quad (2.6.23)$$

$$\mho_{bn}(Z,T) = \left[A_n(T) - \frac{P_n}{2f\mathcal{N}}\mathcal{U}_{2n}(Z,T)\right]\exp(-\Gamma_n Z). \quad (2.6.24)$$

Now, by using Eqs.(2.6.12)-(2.6.15), (2.6.17), (2.6.19), and (2.6.22)-(2.6.24), $\tilde{u}$ and $\tilde{\Theta}$ can finally be written as follows, respectively:

$$\tilde{u}(Y,Z,T) = a^2\sum_{n=1}^{\infty}[\mho_{an}(Z,T) + \mho_{bn}(Z,T)]\sin(\ell_n Y), \quad (2.6.25)$$

$$\tilde{\Theta}(Y,Z,T) = a^2\mathcal{N}\sum_{n=1}^{\infty}[\mho_{an}(Z,T) - \mho_{bn}(Z,T)]\cos(\ell_n Y). \quad (2.6.26)$$

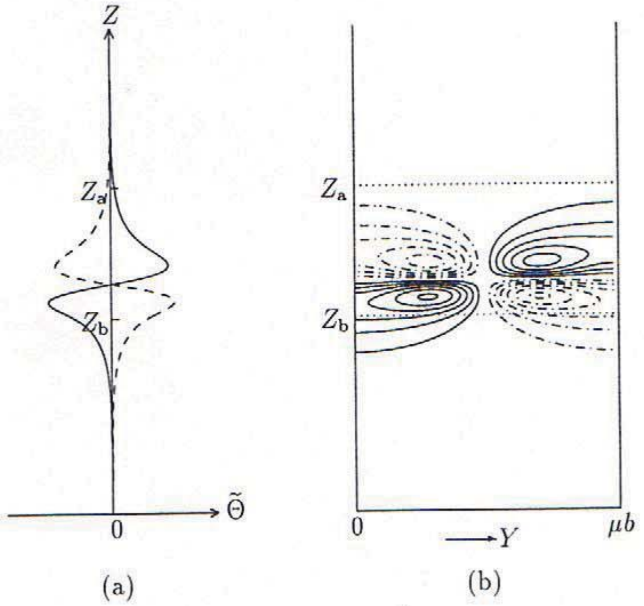


Figure 2.8: As in Fig.2.5, but for $\widetilde{\Theta}$.

Note first from Eqs.(2.6.13) and (2.6.14) that both $\mathcal{U}_{1n}(Z,T)$ and $\mathcal{U}_{2n}(Z,T)$ are zero below the dissipation layer $\mathcal{D}$ (i.e., $Z \leq Z_b$). Within the layer $\mathcal{D}$ (i.e., $Z_b < Z < Z_a$), both $\mathcal{U}_{1n}(Z,T)$ and $\mathcal{U}_{2n}(Z,T)$ remain zero if $T \leq Z/c_{gz}$, are monotonic increasing functions of T for $Z/c_{gz} < T < Z_a/c_{gz}$, and are constants independent of time for $T > Z_a/c_{gz}$. Above the layer $\mathcal{D}$ (i.e., $Z \geq Z_a$), they are zero if $T < Z_a/c_{gz}$ and otherwise are constants independent of time. Also note from Eq.(2.6.15) that $A_n(T)$ is a function that is zero for $T \leq Z_b/c_{gz}$, increases with T for $Z_b/c_{gz} < T < Z_a/c_{gz}$, and remains constant for $T > Z_a/c_{gz}$. These imply that both $|\tilde{u}|$ and $|\widetilde{\Theta}|$, where $\tilde{u}$ and $\widetilde{\Theta}$ are given respectively by Eqs.(2.6.25) and (2.6.26), are cumulative quantities and achieve their maximum values at the end of the thought experiment. As pointed out in §2.5.1, the wave-induced mean PV anomaly, i.e. $\overline{\wp}^L$, is also a cumulative quantity, whose magnitude increases with time within the dissipation layer $\mathcal{D}$ as long as the wave forcing remains.

The structures of $\tilde{u}$ and $\widetilde{\Theta}$ are shown in Fig.2.7 and Fig.2.8, respectively. In comparison with Fig.2.5 we see that, although the dissipating wave can only modify the the distribution of PVS within the dissipation layer $\mathcal{D}$, the wave-induced velocity and buoyancy acceleration fields can penetrate vertically out of the layer $\mathcal{D}$. Such feature has been pointed out by many authors (e.g., Thorpe, 1985; Hoskins *et al.*, 1985; Norton, 1988). From the general properties of the Poisson equation, it follows that $\tilde{u}$ and $\tilde{\Theta}$ are significantly different from zero only within a layer that is centred at the altitude of maximum PV anomaly and has a vertical extent

$$\Delta Z \sim \mu b f / \mathcal{N} \equiv H_{\mathrm{R}}, \tag{2.6.27}$$

provided that the vertical length scale of $\overline{\wp}^{\mathrm{L}}$, say D_δ, is much less than the *Rossby height* H_{R}. If D_δ is large compared to H_{R}, then $\tilde{u}$ and $\overline{\wp}^{\mathrm{L}}$ are significantly different from zero only within the layer where δP is significantly different from zero (Andrews, 1980).

2.6.3 The balanced mean motions from a fine-grain perspective

We now return to the fine-grain interpretation of $\overline{P}$. Substituting wave solutions (2.4.27) into Eqs.(2.5.33) and using Eqs.(2.4.38), (2.4.13), (2.4.30) and (2.5.31), we obtain

$$\begin{aligned}\overline{P} &= P_{\mathrm{B}} + \mu \frac{\theta_{\mathrm{B}}}{\rho_{\mathrm{B}}}\left[f\frac{\partial\tilde{\tilde{\Theta}}}{\partial Z} - \mathcal{N}^2\frac{\partial\tilde{\tilde{u}}}{\partial Y} + \frac{\mathcal{N}^2(\iota^2k^2+m^2)}{2k\sigma}\frac{\partial}{\partial Y}\left(\widehat{w}_0^2\right)\right] \\ &= P_{\mathrm{B}} + \mu \frac{\theta_{\mathrm{B}}}{\rho_{\mathrm{B}}}\left[f\frac{\partial\tilde{\tilde{\Theta}}}{\partial Z} - \mathcal{N}^2\frac{\partial\tilde{\tilde{u}}}{\partial Y} - \frac{a^2}{2}\mathcal{H}^2(Z)\mathcal{G}^2(\hat{T})\sum_{n=1}^{\infty}P_n\cos(\ell Y)\right],\end{aligned} \tag{2.6.28}$$

correct to $O(\mu a^2)$, where $\tilde{\tilde{u}}$ and $\tilde{\tilde{\Theta}}$ represent respectively the $O(a^2)$ mean zonal velocity and buoyancy acceleration deducible from Eq.(2.6.28) with the balance condition (2.6.1) and boundary conditions (2.6.2)–(2.6.4) and (2.6.6). In comparison with Eq.(2.6.7), we see that there is an extra term on the right-hand side of Eq.(2.6.28), representing the effect of Reynolds averages in Eq.(2.5.33). Therefore when the invertibility principle for $\overline{P}$ defined by Eq.(2.6.28) applies, the resulting balanced mean motions $\tilde{\tilde{u}}$ and $\tilde{\tilde{\Theta}}$ can be expected to be different from $\tilde{u}$ and $\tilde{\Theta}$ obtained from Eq.(2.6.7).

To see what results can be obtained from the fine-grain interpretation of $\overline{P}$, we now take $\partial(2.6.28)/\partial Y$ and use Eq.(2.5.32) and the thermal wind relation. After some manipulation we obtain

$$\mathcal{N}^2\frac{\partial^2\tilde{\tilde{u}}}{\partial Y^2} + f^2\frac{\partial^2\tilde{\tilde{u}}}{\partial Z^2} = \underbrace{a^2\mathcal{H}^2(Z)\left[\Lambda(Z)\mathcal{K}(Z,T) \qquad\qquad \right]\sum_{n=1}^{\infty}\ell_n P_n\sin(\ell_n Y)}_{\text{Dissipation-dependent effect}}$$

$$\underbrace{+\tfrac{1}{2}a^2\mathcal{H}^2(Z)\mathcal{G}^2(\hat{T})\sum_{n=1}^{\infty}\ell_n P_n \sin(\ell_n Y)}_{\text{Dissipation-independent effect}}. \tag{2.6.29}$$

In comparison with Eq.(2.6.8), we see that Eq.(2.6.29) includes on its right-hand side a dissipation-dependent forcing which is identical to the only forcing term on the right-hand side of Eq.(2.6.8), and a dissipation-independent forcing which does not appear in Eq.(2.6.8). With the aid of boundary conditions (2.6.3), (2.6.4), and (2.6.6), we obtain the solution of Eq.(2.6.29) as following:

$$\tilde{\tilde{u}}(Y,Z,T) = \tilde{u}(Y,Z,T) \underbrace{- a^2\sum_{n=1}^{\infty}\frac{P_n}{4\mathcal{N}fc_{gz}}\left[\mathcal{I}_{1n}(Z,T)+\mathcal{I}_{n2}(Z,T)\right]\sin(\ell_n Y)}_{\text{Dissipation-independent motion}}, \tag{2.6.30}$$

where $\tilde{u}(Y,Z,T)$ is given by Eq.(2.6.25), and $\mathcal{I}_1$ and $\mathcal{I}_2$ are defined respectively by

$$\mathcal{I}_{1n}(Z,T) = \exp(\Gamma_n Z)\int_Z^{\infty}\mathcal{H}^2(Z)\mathcal{G}^2(\hat{T})\exp(-\Gamma_n Z)dZ, \tag{2.6.31}$$

$$\begin{aligned}\mathcal{I}_{2n}(Z,T) = \exp(-\Gamma_n Z)\Big[&\int_0^{\infty}\mathcal{H}^2(Z)\mathcal{G}^2(\hat{T})\exp(-\Gamma_n Z)dZ \\ &+\int_0^{Z}\mathcal{H}^2(Z)\mathcal{G}^2(\hat{T})\exp(\Gamma_n Z)dZ\Big].\end{aligned} \tag{2.6.32}$$

Eq.(2.6.30) shows that $\tilde{\tilde{u}}$ deduced from the fine-grain perspective differs from $\tilde{u}$ deduced from the coarse-grain perspective by a dissipation-independent term, which does not vanish in the limit of zero dissipation. At first sight, the wave-induced dissipation-independent mean motion appears to be as important as its dissipation-dependent counterpart, i.e. $\tilde{u}$, because both of them appear to be $O(a^2)$ quantities. It should be remembered, however, that $\tilde{u}$ is a cumulative quantity whose magnitude increases irreversibly with time as long as the wave forcing remains, as pointed out in Section 2.6. In contrast, Eqs.(2.6.30)–(2.6.32) show that the wave-induced dissipation-independent mean motion is uniformly bounded by $O(a^2)$ and effectively vanishes after the waves propagate out of the region of interest. In other words, the dissipative mean effect of waves is irreversible and permanent, while the non-dissipative mean effect of waves is reversible and temporary. It should be noticed that the dissipation-independent mean motion in Eq.(2.6.30) may never vanish exactly after the last wave packet propagating out of the region of interest. However, it can be significantly different from zero only within two layers associated with the leading edge and the trailing edge of the wave packet. These layers have a vertical scale characterised by the Rossby height H_R defined by Eq.(2.6.27), and propagate away with the vertical group velocity c_{gz} (Andrews, 1980; Uryu, 1980). Therefore as time goes on the wave-induced dissipation-dependent mean motion can be expected to dominate its dissipation-independent counterpart.

Integrating the thermal wind relation (2.6.1) with respect to Y, gives

$$\tilde{\tilde{\Theta}}(Y,Z,T) = \tilde{\Theta}(Y,Z,T) - \underbrace{a^2 \sum_{n=1}^{\infty} \frac{P_n}{4fc_{gz}} \left[\mathcal{I}_{1n}(Z,T) - \mathcal{I}_{2n}(Z,T)\right] \sin(\ell_n Y)}_{\text{Dissipation-independent field}}, \quad (2.6.33)$$

where $\tilde{\Theta}(Y,Z,T)$ is defined by Eq.(2.6.26). Note again that $\tilde{\tilde{\Theta}}$ deduced from the fine-grain perspective differs from $\tilde{\Theta}$ deduced from the coarse-grain perspective by a dissipation-independent term, which does not vanish in the limit of zero dissipation but is uniformly bounded by $O(a^2)$ and vanishes effectively after the waves propagate out of the region of interest.

2.7 The wave-induced mean motions in the Eulerian-mean and GLM frameworks

In the preceding section, the wave-induced Eulerian-mean PV anomaly field was inverted to yield the wave-induced balanced mean motions. The analysis suggests a succinct yet general way of describing the wave–mean interaction in terms of the wave-induced PVS transport. This PV approach, as pointed out by McIntyre and Norton (1990), tries to complement, rather than compete with, existing theories of wave–mean interaction. This is because applying the PV invertibility principle is an inherently approximate process (Hoskins *et al.*, 1985; McIntyre and Norton, 1990, 1995), therefore in principle no claim can be made about the exactness of the approach, although in practice the approximations involved can be astonishingly good in comparison with what one might imagine from the standard approximate inversion theories that restrict attention to some small parameters (McIntyre and Norton, 1995).

In this section, we shall discuss the description of wave-induced mean motions in terms of the wave-induced momentum transport. Both the Eulerian-mean and GLM formulations are considered. Comparisons of the results obtained here with those obtained in Section 2.5 and Section 2.6 provide further insights into the general nature of the wave-induced mean motions and the usefulness, together with the ultimate limitations of applicability, of the PV approach to analysis of wave–mean interaction.

2.7.1 The wave-induced Eulerian-mean motion

Our starting point is the Eulerian-mean set of equations. Applying the Eulerian-mean operator $\overline{(\ )}$ to Eqs.(2.3.16)–(2.3.20) and assuming as before that all wave-induced mean quantities are functions of (Y, Z, T) only, we obtain (cf. Andrews, 1980):

$$\mu\frac{\overline{D}\overline{u}}{DT} - f\overline{v} = -\mu\left[\frac{\partial}{\partial Y}\left(\overline{v'u'}\right) + \frac{\partial}{\partial Z}\left(\overline{w'u'}\right)\right] + O(\mu^3\overline{u}), \tag{2.7.1}$$

$$\mu\frac{\overline{D}\overline{v}}{DT} + f\overline{u} + \mu\frac{\partial\overline{\Pi}}{\partial Y} = -\mu\left[\frac{\partial}{\partial Y}\left(\overline{v'^2}\right) + \frac{\partial}{\partial Z}\left(\overline{w'v'}\right)\right] + O(\mu^3\overline{v}), \tag{2.7.2}$$

$$\mu\iota^2\frac{\overline{D}\overline{w}}{DT} - \overline{\Theta} + \mu\frac{\partial\overline{\Pi}}{\partial Z} = -\mu\iota^2\left[\frac{\partial}{\partial Y}\left(\overline{v'w'}\right) + \frac{\partial}{\partial Z}\left(\overline{w'^2}\right)\right] + O(\mu^3\overline{w}), \tag{2.7.3}$$

$$\mu\frac{\overline{D}\overline{\Theta}}{DT} + \mathcal{N}^2\overline{w} = -\mu\left[\frac{\partial}{\partial Y}\left(\overline{v'\Theta'}\right) + \frac{\partial}{\partial Z}\left(\overline{w'\Theta'}\right)\right] + O(\mu^3\overline{\Theta}), \tag{2.7.4}$$

$$\frac{\partial\overline{v}}{\partial Y} + \frac{\partial\overline{w}}{\partial Z} = 0. \tag{2.7.5}$$

where

$$\frac{\overline{D}}{DT} \equiv \frac{\partial}{\partial T} + \overline{v}\frac{\partial}{\partial Y} + \overline{w}\frac{\partial}{\partial Z}. \tag{2.7.6}$$

The corresponding boundary conditions are (see Andrews, 1980)

$$\overline{v} = 0 \qquad \text{on} \quad Y = 0,\ \mu b, \tag{2.7.7}$$

$$\overline{w} = \mu\frac{\partial}{\partial Y}\left(\overline{v'h'}\right) \qquad \text{on} \quad Z = 0, \tag{2.7.8}$$

$$\left(\overline{u}, \overline{v}, \overline{w}, \overline{\Pi}, \overline{\Theta}\right) \to 0 \qquad \text{as} \quad Z \to \infty. \tag{2.7.9}$$

The initial conditions are as before, namely all wave-induced quantities vanish for $t < 0$.

Noticing that μ is a small parameter ($\mu \ll 1$), we now seek asymptotic solutions to Eqs.(2.7.1)–(2.7.5) by setting

$$\left.\begin{aligned}
\overline{u} &= \overline{u}_0(Y,Z,T) + \mu\overline{u}_1(Y,Z,T) + \mu^2\overline{u}_2(Y,Z,T) + \cdots \\
\overline{v} &= \overline{v}_0(Y,Z,T) + \mu\overline{v}_1(Y,Z,T) + \mu^2\overline{v}_2(Y,Z,T) + \cdots \\
\overline{w} &= \overline{w}_0(Y,Z,T) + \mu\overline{w}_1(Y,Z,T) + \mu^2\overline{w}_2(Y,Z,T) + \cdots \\
\overline{\Theta} &= \overline{\Theta}_0(Y,Z,T) + \mu\overline{\Theta}_1(Y,Z,T) + \mu^2\overline{\Theta}_2(Y,Z,T) + \cdots \\
\overline{\Pi} &= \mu^{-1}\overline{\Pi}_{-1}(Y,Z,T) + \overline{\Pi}_0(Y,Z,T) + \mu\overline{\Pi}_1(Y,Z,T) + \cdots
\end{aligned}\right\} \tag{2.7.10}$$

where $\overline{u}_0$, $\overline{u}_1$, etc. are all $O(a^2)$ and independent of the small parameter μ. Note that, for consistency with the fact that $\overline{u}$ and $\overline{\Theta}$ are $O(a^2)$, it is necessary to assume that $\overline{\Pi}$ is $O(\mu^{-1}a^2)$; also see Andrews (1980).

Substituting Eq.(2.7.10) into Eqs.(2.7.1)–(2.7.5) and equating coefficients of like powers of μ, we obtain

at μ^0:

$$\overline{v}_0 = 0, \quad \overline{w}_0 = 0, \tag{2.7.11}$$

$$f\overline{u}_0 + \frac{\partial \overline{\Pi}_{-1}}{\partial Y} = 0, \tag{2.7.12}$$

$$-\overline{\Theta}_0 + \frac{\partial \overline{\Pi}_{-1}}{\partial Z} = 0. \tag{2.7.13}$$

Eq.(2.7.11) indicates that both $\overline{v}$ and $\overline{w}$ are $O(\mu a^2)$ or smaller. Eqs.(2.7.12) and (2.7.13) show that the leading order fields are in geostrophic and hydrostatic balances. Combining these two equations leads to the thermal wind relation

$$f\frac{\partial \overline{u}_0}{\partial Z} = -\frac{\partial \overline{\Theta}_0}{\partial Y}. \tag{2.7.14}$$

At μ^1:

$$\frac{\partial \overline{u}_0}{\partial T} - f\overline{v}_1 = -\frac{\partial}{\partial Y}\left(\overline{v_0' u_0'}\right) - \frac{\partial}{\partial Z}\left(\overline{w_0' u_0'}\right), \tag{2.7.15}$$

$$f\overline{u}_1 + \frac{\partial \overline{\Pi}_0}{\partial Y} = -\frac{\partial}{\partial Y}\left(\overline{v_0'^2}\right) - \frac{\partial}{\partial Z}\left(\overline{w_0' v_0'}\right), \tag{2.7.16}$$

$$-\overline{\Theta}_1 + \frac{\partial \overline{\Pi}_0}{\partial Z} = -\iota^2\frac{\partial}{\partial Y}\left(\overline{v_0' w_0'}\right) - \iota^2\frac{\partial}{\partial Z}\left(\overline{w_0'^2}\right), \tag{2.7.17}$$

$$\frac{\partial \overline{\Theta}_0}{\partial T} + \mathcal{N}^2\overline{w}_1 = -\frac{\partial}{\partial Y}\left(\overline{v_0' \Theta_0'}\right) - \frac{\partial}{\partial Z}\left(\overline{w_0' \Theta_0'}\right), \tag{2.7.18}$$

$$\frac{\partial \overline{v}_1}{\partial Y} + \frac{\partial \overline{w}_1}{\partial Z} = 0, \tag{2.7.19}$$

with boundary conditions

$$\overline{v}_1 = 0 \qquad \text{on} \quad Y = 0,\ \mu b, \tag{2.7.20}$$

$$\overline{w}_1 = \frac{\partial}{\partial Y}\left(\overline{v_0' h'}\right) \qquad \text{on} \quad Z = 0, \tag{2.7.21}$$

$$\overline{(\,)} \to 0 \qquad \text{as} \quad Z \to \infty. \tag{2.7.22}$$

Now taking

$$\mathcal{N}^2\frac{\partial^2}{\partial Y^2}[(2.7.15)] - f^2\frac{\partial^2}{\partial Y \partial Z}[(2.7.18)] \tag{2.7.23}$$

and using Eqs.(2.7.12), (2.7.13), and (2.7.19), we obtain

$$\left(\mathcal{N}^2\frac{\partial^2}{\partial Y^2} + f^2\frac{\partial^2}{\partial Z^2}\right)\frac{\partial \overline{u}_0}{\partial T} = \mathcal{N}^2\frac{\partial^2}{\partial Y^2}(\nabla\cdot\mathbf{S}_0) + f\frac{\partial^3}{\partial Y \partial Z^2}\left(\overline{w_0' \Theta_0'}\right), \tag{2.7.24}$$

where $\mathbf{S}_0$ is the leading order EP flux defined by

$$\mathbf{S}_0 = -\left(\overline{u_0'v_0'}\right)\hat{\mathbf{y}} - \left(\overline{u_0'w_0'} - f\overline{v_0'\Theta_0'}/\mathcal{N}^2\right)\hat{\mathbf{z}}. \tag{2.7.25}$$

The last term in Eq.(2.7.24) will turn out to be zero, reflecting the fact that w' and Θ' are nearly in quadrature, with only $O(\mu)$ phase error due to the dissipation process.

The lateral and lower boundary conditions for Eq.(2.7.24) are

$$\frac{\partial\overline{u}_0}{\partial T} = -\frac{\partial}{\partial Y}\left(\overline{v_0'u_0'}\right) - \frac{\partial}{\partial Z}\left(\overline{w_0'u_0'}\right) \qquad \text{on} \quad Y = 0,\ \mu b, \tag{2.7.26}$$

$$\frac{\partial^2\overline{u}_0}{\partial Z\partial T} = \frac{1}{f}\frac{\partial}{\partial Y}\left[\frac{\partial}{\partial Y}\left(\overline{v_0'\Theta_0'} + \overline{v_0'h'}/\mathcal{N}^2\right) + \frac{\partial}{\partial Z}\left(\overline{w_0'\Theta_0'}\right)\right] \quad \text{on} \quad Z = 0. \tag{2.7.27}$$

The upper boundary condition is given by Eq.(2.7.22).

Substituting the wave solutions (2.4.27) and (2.4.38) into Eq.(2.7.24) and integrating the resulting equation with respect to T, we obtain

$$\begin{aligned}\mathcal{N}^2\frac{\partial^2\overline{u}_0}{\partial Y^2} + f^2\frac{\partial^2\overline{u}_0}{\partial Z^2} &= \underbrace{a^2\mathcal{H}^2(Z)\Lambda(Z)\mathcal{K}(Z,T)\sum_{n=1}^{\infty}\ell_n P_n\sin(\ell_n Y)}_{\text{Dissipation-dependent effect}} \\ &\quad \underbrace{-\frac{a^2\mathcal{H}^2(Z)}{2}\int_0^T\frac{\partial}{\partial Z}\left[\mathcal{G}^2(\tilde{T}-Z/c_{gz})\right]d\tilde{T}\sum_{n=1}^{\infty}\ell_n P_n\sin(\ell_n Y)}_{\text{Dissipation-independent effect}},\end{aligned} \tag{2.7.28}$$

where P_n is defined by Eq.(2.5.31). The corresponding boundary conditions are

$$\overline{u}_0 = 0 \qquad \text{on} \quad Y = 0, \mu b, \tag{2.7.29}$$

$$\frac{\partial\overline{u}_0}{\partial Z} = 0 \qquad \text{on} \quad Z = 0, \tag{2.7.30}$$

$$\overline{u}_0 \to 0 \qquad \text{as} \quad Z \to \infty. \tag{2.7.31}$$

The integral in the last term of Eq.(2.7.28) can be manipulated as follows:

$$\begin{aligned}\int_0^T\frac{\partial}{\partial Z}\left[\mathcal{G}^2(\tilde{T} - Z/c_{gz})\right]d\tilde{T} &= -\int_{-Z/c_{gz}}^{T-Z/c_{gz}}\frac{1}{c_{gz}}\frac{\partial}{\partial\tilde{T}}\left[\mathcal{G}^2(\tilde{T})\right]d\tilde{T} \\ &= -\frac{\mathcal{G}^2(T - Z/c_{gz}) - \mathcal{G}^2(-Z/c_{gz})}{c_{gz}} \\ &= -\frac{\mathcal{G}^2(\hat{T})}{c_{gz}},\end{aligned} \tag{2.7.32}$$

where $\hat{T}$ is defined by Eq.(2.4.40) and $\mathcal{G}(-Z/c_{gz}) = 0$ by definition of $\mathcal{G}$ in our model. Thus, it is evident that the dissipation-independent effect in Eq.(2.7.28) represents the effect of wave transience.

Substituting Eq.(2.7.32) into Eq.(2.7.28) will give an equation that is identical to Eq.(2.6.29). In addition, the boundary conditions (2.7.29)-(2.7.31) are also identical to those used to solve Eq.(2.6.29). Therefore the solution of Eq.(2.7.32) must be identical to that of Eq.(2.6.29), i.e.,

$$\overline{u}_0 = \tilde{\tilde{u}}, \tag{2.7.33}$$

where $\tilde{\tilde{u}}$ is given by Eq.(2.6.30). Similarly,

$$\overline{\Theta}_0 = \tilde{\tilde{\Theta}}, \tag{2.7.34}$$

where $\tilde{\tilde{\Theta}}$ is given by Eq.(2.6.33).

2.7.2 The wave-induced GLM motions

The wave-induced mean motions can also be examined by using the GLM formalism. It is possible, as shown in Appendix 2B, to derive the GLM motions directly from the GLM set of equations. Since the Eulerian-mean motions are known already from §2.7.1, an alternative approach to the GLM motions can be obtained by evaluating the corresponding Stokes corrections. As shown in Eq.(2.5.34), all Stokes corrections are $O(\mu a^2, a^2\varphi_0')$ or smaller in our model. In addition, from §2.7.1 we know that $\overline{u}$, $\overline{\Theta}$ are of $O(a^2)$ and $\overline{v}$, $\overline{w}$ are of $O(\mu a^2)$. Therefore we obtain

$$\overline{u}^{\mathrm{L}} = \overline{u} + \overline{u}^{\mathrm{S}} = \overline{u}_0 + O(\mu a^2, a^3), \tag{2.7.35}$$

$$\overline{\Theta}^{\mathrm{L}} = \overline{\Theta} + \overline{\Theta}^{\mathrm{S}} = \overline{\Theta}_0 + O(\mu a^2, a^3), \tag{2.7.36}$$

$$\overline{v}^{\mathrm{L}} = \overline{v} + \overline{v}^{\mathrm{S}} = O(\mu a^2, a^3), \tag{2.7.37}$$

$$\overline{w}^{\mathrm{L}} = \overline{w} + \overline{w}^{\mathrm{S}} = O(\mu a^2, a^3), \tag{2.7.38}$$

where $\overline{u}_0$ and $\overline{\Theta}_0$ are $O(a^2)$ and are given by Eqs.(2.7.33) and (2.7.34), respectively. Thus, Eqs.(2.7.35)-(2.7.38) show that in our model the Stokes corrections to the zonal velocity and buoyancy acceleration are negligible, but not the meridional and vertical Stokes drifts. Furthermore, from Eqs.(2.7.33)-(2.7.37) we see that the wave-induced mean motions described in terms of PVS transport from the fine-grain perspective consist, to the leading order, with those described in terms of momentum transport.

2.8 Summary and remarks

From the above discussions, we draw the following main conclusions:

1. The wave-induced mean motions can be separated into dissipation-dependent mean motions, which exist only in the presence of wave dissipation, and dissipation-independent mean motions, which are associated with various non-dissipative effects (such as wave transience, Stokes corrections, etc.). In the thought experiment described above, the wave-induced PV anomalies depend crucially on the wave dissipation. They are different from zero only within the dissipation layer. The dissipation-dependent mean zonal velocity and buoyancy acceleration can be nonzero outside the dissipation layer. They vanish, however, if the dissipation vanishes everywhere. In contrast, although the wave-induced dissipation-independent mean motions may also be modified by the effect of wave dissipation, they do not vanish in the absence of wave dissipation.

2. The $O(a^2)$ wave-induced mean motions can be understood in terms of the wave-induced PVS transport. From a fine-grain perspective, it is demonstrated in the above thought experiment that the mean motions derived from the PVS viewpoint are the same as the $O(a^2)$ mean motions derived from the momentum viewpoint. Even from a coarse-grain perspective, the PVS viewpoint still promises a correct description of the $O(a^2)$ dissipation-dependent mean motions, which, in the presence of wave dissipation, are cumulatively much larger than their dissipation-independent counterparts as time goes on.

3. The PV inversion is an inherently approximate process. In the above thought experiment, the wave-induced mean motions deduced from the invertibility principle are correct only to $O(a^2)$ in comparison with results obtained from the momentum viewpoint. In practice, however, the approximations involved can be astonishingly good, as pointed out by McIntyre and Norton (1990, 1995).

It would appear that the coarse-grain interpretation of the mean PV is the only practical option for diagnostic studies because observing analysis systems cannot capture correctly the small-scale fluctuations in velocity and temperature fields. Under such circumstances, only the dissipative type of wave-induced mean motions can be understood in terms of the wave-induced PVS transport. Nevertheless, since the dissipation-dependent mean motions are cumulative quantities, as time goes on they are much larger and more dynamically important than their dissipation-independent counterparts. In this sense, the PVS description of wave–mean interaction captures the essentials well. It not only complements previous descriptions of wave–mean interaction in terms of radiation stress, quasimomentum, etc., but also gives a succinct and general way of saying why certain contributions to the radiation stress are significant for mean flow, and others not (McIntyre and Norton, 1990).

In the above discussion, dissipation is prescribed in a layer above the bottom boundary. It has been shown that, from a viewpoint of PVS dynamics, the essential effect of dissipating waves on the mean motion is to cause irreversible PVS transport along isentropic surfaces. It is worth mentioning here that our above arguments rely on the validity of linear theory. The underlying assumption of linear theory allows us to specify and follow a material tube whose surface only undulates gently when waves propagate through it (Fig.2.4). In recent years, however, wave breaking processes have been frequently identified in the real atmosphere (Hines, 1972; Houghton, 1978; Lindzen, 1981; McIntyre and Palmer, 1983, 1984, 1985; Fritts, 1984, 1989; WMO, 1985; Palmer *et al.*, 1986; Andrews *et al.*, 1987; Hauchecorne *et al.*, 1987; McIntyre and Norton, 1990; Waugh *et al.*, 1994). With nonlinear dynamics being involved, the wave breaking processes are characterised by the rapid and irreversible deformation of otherwise wavy material surfaces. Under such circumstances, it appears that our above arguments are very difficult to justify. Nevertheless, McIntyre and Norton (1990) pointed out that, since the essential effect of the breaking waves is to transport irreversibly the PVS along isentropic surfaces as do the dissipative waves, it is convenient to widen the word 'dissipation' to include all cases of wave breaking. With this generalisation in mind, McIntyre (1992) argued that the effects of breaking Rossby and gravity waves upon the global distribution of PV, and hence the way in which they control the mean circulation, can be thought of in terms of the total transport of PVS exactly along the isentropic surfaces of the atmosphere's stable stratification, no matter how complicated the details.

Appendix 2A: The Boussinesq approximation in a compressible atmosphere

A natural starting point for the discussion of the Boussinesq approximation in a compressible, continuously-stratified atmosphere is the hydrostatic equilibrium state that is described by Eqs.(2.2.1) and (2.2.2). Substituting the second expression in Eq.(2.2.2) into the first expression yields

$$\frac{1}{p_{\mathrm{B}}}\frac{dp_{\mathrm{B}}}{dz} = -\frac{g}{RT_{\mathrm{B}}}. \tag{2A.1}$$

Between the Earth's surface and the 100 km level, the background temperature T_B for the Earth's atmosphere is within 20% of a constant value of $T_c = 240$K. Therefore, integrating Eq.(2A.1) with respect to z leads to, approximately,

$$p_B = p_s \exp(-gz/RT_c) = p_s \exp(-z/H), \tag{2A.2}$$

where

$$H \equiv RT_c/g \tag{2A.3}$$

is called the *scale height of the atmosphere*, i.e., the height at which the background pressure has fallen to e^{-1} of its surface value (Gill, 1982). In the middle atmosphere, $H \approx 7$ km.

For practical purposes we can also use Eq.(2A.2) to define the background pressure p_B. With this definition, the background density ρ_B and the background potential temperature θ_B are given respectively by

$$\rho_B(z) = \rho_s \exp(-z/H), \qquad \theta_B(z) = T_c \exp(\kappa z/H), \tag{2A.4}$$

where

$$\rho_s = p_s/RT_c. \tag{2A.5}$$

To describe motions which represent departures from the static state mentioned above, we now introduce the perturbation density $\delta\rho$, the perturbation pressure δp, and the perturbation potential temperature $\delta\theta$ defined by the equations

$$\rho = \rho_B(z) + \delta\rho, \quad p = p_B(z) + \delta p, \quad \theta = \theta_B(z) + \delta\theta, \tag{2A.6}$$

and assume that $\delta\rho$, δp, and $\delta\theta$ are much smaller than their background counterparts ρ_B, p_B, and θ_B, respectively. Thus, we can use the binomial expansion to approximate $1/\rho$ as

$$\frac{1}{\rho} = \frac{1}{\rho_B + \delta\rho} = \frac{1}{\rho_B(1 + \delta\rho/\rho_B)} \approx \frac{1}{\rho_B}\left(1 - \frac{\delta\rho}{\rho_B}\right). \tag{2A.7}$$

For most motions of meteorological interest, the Coriolis force associated with the horizontal component of the Earth's rotation vector can be consistently neglected. Therefore, substituting Eq.(2A.6) into Eqs.(1.5.1)–(1.5.5) and using conditions (2.3.9) and (2A.7), we obtain the following approximate equations:

$$\frac{D\mathbf{u}}{Dt} + f\hat{\mathbf{z}} \times \mathbf{u} + \frac{1}{\rho_B}\nabla\delta p + \frac{g\delta\rho}{\rho_B}\hat{\mathbf{z}} = \mathbf{F}, \tag{2A.8}$$

$$\frac{D\delta\theta}{Dt} + w\frac{d\theta_{\text{B}}}{dz} = Q, \tag{2A.9}$$

$$\frac{1}{\rho_{\text{B}}}\frac{\partial\delta\rho}{\partial t} + \nabla\cdot\mathbf{u} + \frac{w}{\rho_{\text{B}}}\frac{d\rho_{\text{B}}}{dz} = 0, \tag{2A.10}$$

$$\frac{\delta\rho}{\rho_{\text{B}}} = (1-\kappa)\frac{\delta p}{p_{\text{B}}} - \frac{\delta\theta}{\theta_{\text{B}}}, \tag{2A.11}$$

where $f = 2\Omega\sin\phi$ is the Coriolis parameter (ϕ being the latitude on the Earth.)

Further simplifications of above equations may be applied by using scaling arguments. Let the motion in question be characterised by a horizontal velocity V, a horizontal length scale L, and a vertical length scale D. The scales of $\delta\rho$ and δp are denoted by $\Delta\rho$ and Δp, respectively. We further assume that the scale of time is given by L/V, and the scale of vertical velocity is given by VD/L. Thus for the magnitudes of the terms in Eq.(2A.10) we have the estimates

$$\frac{1}{\rho_{\text{B}}}\frac{\partial\delta\rho}{\partial t} \sim \frac{V\Delta\rho}{L\rho_{\text{B}}}, \quad \nabla\cdot\mathbf{u} \sim \frac{V}{L}, \quad \frac{w}{\rho_{\text{B}}}\frac{d\rho_{\text{B}}}{dz} \sim \frac{DV}{HL}. \tag{2A.12}$$

Noticing the assumed conditions $D \ll H$ and $\Delta\rho/\rho_{\text{B}} \ll 1$, we see that Eq.(2A.10) can be safely replaced by Eq.(2.3.4), which is identical to the idealisation of incompressibility. It should be pointed out that Eq.(2.3.4) is valid only for those motions whose vertical length scale is much less than the atmospheric height scale. For motions with $D \sim H$, on the other hand, the compressibility associated with the height dependence of background density must be taken into account, and the corresponding results are usually referred to as the *anelastic approximation*[2] (Batchelor, 1953; Ogura and Phillips, 1962; Lipps and Hemler, 1982; Scinocca and Shepherd, 1992).

Now consider Eq.(2A.11). Note first that the order of magnitude of pressure perturbation, Δp, is indicated by the vertical component of Eq.(2A.8), which shows that

$$\frac{\Delta p}{\rho_{\text{B}}D} \sim \frac{g\Delta\rho}{\rho_{\text{B}}}, \qquad i.e.,\ \Delta p \sim gD\Delta\rho. \tag{2A.13}$$

From Eqs.(2A.13), (2A.2)–(2A.5), it follows that

$$\frac{\delta p}{p_{\text{B}}} \sim \frac{\Delta p}{p_{\text{B}}} = \frac{gD\Delta\rho}{p_{\text{B}}} = \frac{gD\rho_{\text{s}}\Delta\rho}{p_{\text{s}}\rho_{\text{B}}} = \frac{gD\Delta\rho}{RT_{\text{c}}\rho_{\text{B}}} = \frac{D\Delta\rho}{H\rho_{\text{B}}}. \tag{2A.14}$$

Therefore the order of magnitude of the second term in Eq.(2A.11) is given by

$$\frac{(1-\kappa)\delta p}{p_{\text{B}}} \sim \frac{(1-\kappa)D\Delta\rho}{H\rho_{\text{B}}}, \tag{2A.15}$$

[2] The anelastic approximation was first introduced by Batchelor (1953) in a discussion of dynamical similarity. The name "anelastic" was suggested by Ogura and Phillips (1962) to refer to the ability of the approximation to effectively filter out sound waves, which are of little meteorological interest and whose presence would require the use of very small time steps in a numerical integration.

while the order of magnitude of the first term in Eq.(2A.11) is given by $\Delta\rho/\rho_B$. Therefore, the second term in Eq.(2A.11) can be neglected compared with the first term in the same equation when

$$D \ll H/(1-\kappa) \approx 7H/5. \tag{2A.16}$$

Since D is assumed to satisfy Eq.(2.3.1), it satisfies Eq.(2A.16) automatically. Therefore, to a first approximation Eq.(2A.11) can be replaced by Eq.(2.3.5), which indicates that the changes of density perturbation resulting from the changes of pressure perturbation are negligible for atmospheric motions with $D \ll H$.

Using Eqs.(2.3.5), (2A.4) and (2.3.1), one can manipulate the vertical components of the third and fourth terms in Eq.(2A.8) as follows:

$$\begin{aligned} \frac{1}{\rho_B}\frac{\partial \delta p}{\partial z} + \frac{g\delta\rho}{\rho_B} &= \frac{\partial}{\partial z}\left(\frac{\delta p}{\rho_B}\right) + \frac{\delta p}{{\rho_B}^2}\frac{d\rho_B}{dz} - \frac{g\delta\theta}{\theta_B} \\ &= \frac{\partial}{\partial z}\left(\frac{\delta p}{\rho_B}\right) - \frac{\delta p}{H\rho_B} - \frac{g\delta\theta}{\theta_B} \\ &\approx \frac{\partial}{\partial z}\left(\frac{\delta p}{\rho_B}\right) - \frac{g\delta\theta}{\theta_B}. \end{aligned} \tag{2A.17}$$

Substituting Eq.(2A.17) into Eq.(2A.8) leads to Eq.(2.3.2). Note that in Eq.(2.3.2) the background density ρ_B has been brought under the gradient operator in the pressure-gradient force.

Using condition (2.3.1) and the definition of Θ given in Eq.(2.3.8), we can easily replace Eq.(2A.9) by Eq.(2.3.3).

Appendix 2B: An independent derivation of the GLM motion induced by inertia-gravity waves

Assuming that all wave-induced mean quantities are functions of (Y, Z, T) only, we can obtain the GLM versions of Eqs.(2.3.16)–(2.3.20) as follows (Andrews and McIntyre, 1978b; McIntyre, 1980a)

$$\mu\frac{\overline{D}^L\overline{u}^L}{DT} - f\overline{v}^L = \mu\left[\frac{\partial}{\partial Y}\left(\overline{\Pi'\frac{\partial\eta'}{\partial x}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi'\frac{\partial\zeta'}{\partial x}}\right)\right] + O(\mu^3\overline{u}^L), \tag{2B.1}$$

$$\mu\frac{\overline{D}^L\overline{v}^L}{DT} + f\overline{u}^L + \mu\frac{\partial\overline{\Pi}^L}{\partial Y} = \mu\left[\frac{\partial}{\partial Y}\left(\overline{\Pi'\frac{\partial\eta'}{\partial y}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi'\frac{\partial\zeta'}{\partial y}}\right)\right] + O(\mu^3\overline{v}^L), \tag{2B.2}$$

$$\mu\iota^2\frac{\overline{D}^{\mathrm{L}}\overline{w}^{\mathrm{L}}}{DT} - \overline{\Theta}^{\mathrm{L}} + \mu\frac{\partial\overline{\Pi}^{\mathrm{L}}}{\partial Z} = \mu\iota^2\left[\frac{\partial}{\partial Y}\left(\overline{\Pi'\frac{\partial\eta'}{\partial z}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi'\frac{\partial\zeta'}{\partial z}}\right)\right] + O(\mu^3\overline{w}^{\mathrm{L}}), \tag{2B.3}$$

$$\mu\frac{\overline{D}^{\mathrm{L}}\overline{\Theta}^{\mathrm{L}}}{DT} + \mathcal{N}^2\overline{w}^{\mathrm{L}} = O(\mu^3\overline{\Theta}^{\mathrm{L}}), \tag{2B.4}$$

$$\frac{\partial\overline{v}^{\mathrm{L}}}{\partial Y} + \frac{\partial\overline{w}^{\mathrm{L}}}{\partial Z} = O(\mu^2 a^2), \tag{2B.5}$$

where

$$\frac{\overline{D}^{\mathrm{L}}}{DT} \equiv \frac{\partial}{\partial T} + \overline{v}^{\mathrm{L}}\frac{\partial}{\partial Y} + \overline{w}^{\mathrm{L}}\frac{\partial}{\partial Z}. \tag{2B.6}$$

The corresponding boundary conditions are

$$\overline{v}^{\mathrm{L}} = 0 \qquad \text{on} \quad Y = 0,\ \mu b, \tag{2B.7}$$

$$\overline{w}^{\mathrm{L}} = 0 \qquad \text{on} \quad Z = 0, \tag{2B.8}$$

$$\left(\overline{u}^{\mathrm{L}}, \overline{v}^{\mathrm{L}}, \overline{w}^{\mathrm{L}}, \overline{\Pi}^{\mathrm{L}}, \overline{\Theta}^{\mathrm{L}}\right) \to 0 \qquad \text{as} \quad Z \to \infty. \tag{2B.9}$$

The initial conditions require that all wave-induced quantities vanish for $t < 0$.

Since μ is a small parameter ($\mu \ll 1$), we can let

$$\left.\begin{aligned}
\overline{u}^{\mathrm{L}} &= \overline{u}_0^{\mathrm{L}}(Y,Z,T) + \mu\overline{u}_1^{\mathrm{L}}(Y,Z,T) + \mu^2\overline{u}_2^{\mathrm{L}}(Y,Z,T) + \cdots \\
\overline{v}^{\mathrm{L}} &= \overline{v}_0^{\mathrm{L}}(Y,Z,T) + \mu\overline{v}_1^{\mathrm{L}}(Y,Z,T) + \mu^2\overline{v}_2^{\mathrm{L}}(Y,Z,T) + \cdots \\
\overline{w}^{\mathrm{L}} &= \overline{w}_0^{\mathrm{L}}(Y,Z,T) + \mu\overline{w}_1^{\mathrm{L}}(Y,Z,T) + \mu^2\overline{w}_2^{\mathrm{L}}(Y,Z,T) + \cdots \\
\overline{\Theta}^{\mathrm{L}} &= \overline{\Theta}_0^{\mathrm{L}}(Y,Z,T) + \mu\overline{\Theta}_1^{\mathrm{L}}(Y,Z,T) + \mu^2\overline{\Theta}_2^{\mathrm{L}}(Y,Z,T) + \cdots \\
\overline{\Pi}^{\mathrm{L}} &= \mu^{-1}\overline{\Pi}_{-1}^{\mathrm{L}}(Y,Z,T) + \overline{\Pi}_0^{\mathrm{L}}(Y,Z,T) + \mu\overline{\Pi}_1^{\mathrm{L}}(Y,Z,T) + \cdots
\end{aligned}\right\} \tag{2B.10}$$

where $\overline{u}_0^{\mathrm{L}}$, $\overline{u}_1^{\mathrm{L}}$, etc. are all $O(a^2)$ and independent of the small parameter μ. Substituting Eq.(2B.10) into Eqs.(2B.1)–(2B.5) and equating coefficients of like powers of μ, we obtain:

at μ^0:

$$\overline{v}_0^{\mathrm{L}} = 0, \quad \overline{w}_0^{\mathrm{L}} = 0, \tag{2B.11}$$

$$f\overline{u}_0^{\mathrm{L}} + \frac{\partial\overline{\Pi}_{-1}^{\mathrm{L}}}{\partial Y} = 0, \tag{2B.12}$$

$$-\overline{\Theta}_0^{\mathrm{L}} + \frac{\partial\overline{\Pi}_{-1}^{\mathrm{L}}}{\partial Z} = 0. \tag{2B.13}$$

Eqs.(2B.11)–(2B.13) are identical to their Eulerian-mean counterparts (2.7.11)–(2.7.13).

At μ^1:

$$\frac{\partial\overline{u}_0^{\mathrm{L}}}{\partial T} - f\overline{v}_1^{\mathrm{L}} = \left[\frac{\partial}{\partial Y}\left(\overline{\Pi_0'\frac{\partial\eta_0'}{\partial x}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi_0'\frac{\partial\zeta_0'}{\partial x}}\right)\right], \tag{2B.14}$$

$$f\overline{u}_1^{\mathrm{L}} + \frac{\partial \overline{\Pi}_0^{\mathrm{L}}}{\partial Y} = \left[\frac{\partial}{\partial Y}\left(\overline{\Pi_0'\frac{\partial \eta_0'}{\partial y}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi_0'\frac{\partial \zeta_0'}{\partial y}}\right)\right], \tag{2B.15}$$

$$-\overline{\Theta}_1^{\mathrm{L}} + \frac{\partial \overline{\Pi}_0^{\mathrm{L}}}{\partial Z} = \iota^2\left[\frac{\partial}{\partial Y}\left(\overline{\Pi_0'\frac{\partial \eta_0'}{\partial z}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi_0'\frac{\partial \zeta_0'}{\partial z}}\right)\right], \tag{2B.16}$$

$$\frac{\partial \overline{\Theta}_0^{\mathrm{L}}}{\partial T} + \mathcal{N}^2\overline{w}_1^{\mathrm{L}} = 0, \tag{2B.17}$$

$$\frac{\partial \overline{v}_1^{\mathrm{L}}}{\partial Y} + \frac{\partial \overline{w}_1^{\mathrm{L}}}{\partial Z} = 0, \tag{2B.18}$$

with boundary conditions

$$\overline{v}_1^{\mathrm{L}} = 0 \qquad \text{on} \quad Y = 0,\ \mu b, \tag{2B.19}$$

$$\overline{w}_1^{\mathrm{L}} = 0 \qquad \text{on} \quad Z = 0, \tag{2B.20}$$

$$\overline{(\,)}^{\mathrm{L}} \to 0 \qquad \text{as} \quad Z \to \infty. \tag{2B.21}$$

From Eqs.(2B.12)–(2B.14), (2B.17), and (2B.18), it follows that

$$\left(\mathcal{N}^2\frac{\partial^2}{\partial Y^2} + f^2\frac{\partial^2}{\partial Z^2}\right)\frac{\partial \overline{u}_0^{\mathrm{L}}}{\partial T} = \mathcal{N}^2\frac{\partial^2}{\partial Y^2}\left[\frac{\partial}{\partial Y}\left(\overline{\Pi_0'\frac{\partial \eta_0'}{\partial x}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi_0'\frac{\partial \zeta_0'}{\partial x}}\right)\right]. \tag{2B.22}$$

The lateral and lower boundary conditions for Eq.(2B.22) are

$$\frac{\partial \overline{u}_0^{\mathrm{L}}}{\partial T} = \left[\frac{\partial}{\partial Y}\left(\overline{\Pi_0'\frac{\partial \eta_0'}{\partial x}}\right) + \frac{\partial}{\partial Z}\left(\overline{\Pi_0'\frac{\partial \zeta_0'}{\partial x}}\right)\right] \qquad \text{on} \quad Y = 0,\ \mu b, \tag{2B.23}$$

$$\frac{\partial^2 \overline{u}_0^{\mathrm{L}}}{\partial Z \partial T} = 0, \qquad \text{on} \quad Z = 0. \tag{2B.24}$$

The upper boundary condition is given by Eq.(2B.21).

Substituting the wave solutions (2.4.27) and (2.4.38) into Eq.(2B.22) and integrating the resulting equation with respect to T, after some manipulation we obtain

$$\begin{aligned}\mathcal{N}^2\frac{\partial^2 \overline{u}_0^{\mathrm{L}}}{\partial Y^2} + f^2\frac{\partial^2 \overline{u}_0^{\mathrm{L}}}{\partial Z^2} &= \underbrace{a^2\mathcal{H}^2(Z)\Lambda(Z)\mathcal{K}(Z,T)\sum_{n=1}^{\infty}\ell_n P_n \sin(\ell_n Y)}_{\text{Dissipation-dependent effect}} \\ &\quad \underbrace{-\frac{a^2\mathcal{H}^2(Z)}{2}\int_0^T \frac{\partial}{\partial Z}\left[\mathcal{G}^2(\tilde{T} - Z/c_{gz})\right] d\tilde{T}\sum_{n=1}^{\infty}\ell_n P_n \sin(\ell_n Y)}_{\text{Dissipation-independent effect}},\end{aligned} \tag{2B.25}$$

where P_n is defined by Eq.(2.5.31). The corresponding boundary conditions are

$$\overline{u}_0^{\mathrm{L}} = 0 \qquad \text{on} \quad Y = 0, \mu b, \tag{2B.26}$$

$$\frac{\partial \overline{u}_0^{\mathrm{L}}}{\partial Z} = 0 \qquad \text{on} \quad Z = 0, \tag{2B.27}$$

$$\overline{u}_0^{\mathrm{L}} \to 0 \qquad \text{as} \quad Z \to \infty. \tag{2B.28}$$

Eq.(2B.25) and the boundary conditions (2B.26)–(2B.28) are identical to their Eulerian-mean counterparts (2.7.28)–(2.7.31). Therefore their solutions must be identical, i.e.,

$$\overline{u}_0^L = \overline{u}_0. \tag{2B.29}$$

In addition, use of the thermal wind relation yields

$$\overline{\Theta}_0^L = \overline{\Theta}_0. \tag{2B.30}$$

CHAPTER 3

WAVE-INDUCED TRANSPORT OF PV-SUBSTANCE AND THE RATE OF QUASIMOMENTUM DISSIPATION: THE GENERALIZED TAYLOR IDENTITY

It was demonstrated in Chapter 2 that wave-induced mean motions in a simple thought-experiment can be understood in a unified way by viewing the phenomena in terms of the wave-induced PV-substance (PVS) transport. As a further step towards developing the theoretical tools needed to quantify these ideas and extend the argument to more complicated fluid systems, in this chapter we derive some general relationships between the wave-induced momentum transport and the wave-induced PVS transport.

3.1 Introduction

As mentioned in Section 1.2, the idea that transport of momentum by disturbances (wave-like or turbulence motions) may be related to the flux of PV or vorticity can be traced back to G. I. Taylor (1915). Taking one of the simplest cases in which incompressible, non-rotating fluid motions are limited to two dimensions, say x and z, Taylor found that the effect of disturbances on the mean flow can be examined in terms of vorticity transport. The most important result of his investigation was the discovery of the well-known relation (known as the Taylor identity):

$$I = \rho \int w' \left(\frac{\partial u'}{\partial z} - \frac{\partial w'}{\partial x} \right) dx, \tag{3.1.1}$$

where I is the rate at which x-momentum leaves a layer of unit thickness, defined by

$$I = \frac{\partial}{\partial z} \int \rho(\overline{U} + u')w' dx = \rho \frac{\partial}{\partial z} \int u'w' dx, \tag{3.1.2}$$

with ρ being the fluid density, $\overline{U}$ a mean zonal velocity that is assumed constant, u' and w' the x-component and z-component eddy velocities, respectively. Since the term in the parentheses on the right-hand side of Eq.(3.1.1) represents the eddy vorticity, the Taylor identity states that in an incompressible, two-dimensional, non-rotating fluid, the momentum flux divergence is equal to the eddy flux of vorticity.

It worth emphasising here that, although waves in fluids can transfer momentum from where they are created to where they are dissipated, they do not generally have a uniquely defined mean momentum, or, in other words, the momentum of the fluid may not be spatially distributed in the same way as the waves (McIntyre, 1973, 1981). In practice, one can define a quasimomentum (also known as pseudomomentum), which is a wave property evaluable from linear wave solutions, but is not physically the same thing as momentum (Andrews and McIntyre, 1978b,c; McIntyre, 1981). In most cases, however, there is a fairly close relationship (although not equality) between the momentum and quasimomentum fluxes (Brillouin, 1925; Gordon, 1973; Andrews and McIntyre, 1978b,c; McIntyre, 1981, 1993). Thus, the Taylor identity may also be interpreted as stating that the rate of dissipation of quasimomentum is equal to the eddy flux of vorticity in the incompressible, two-dimensional, non-rotating fluid.[1]

Bretherton (1966a,b) first applied Taylor's idea of eddy transport of vorticity to geophysical fluid dynamics within the framework of quasigeostrophic theory. He found an important equation, which can be written as (also see Green, 1970; Edmon *et al.*, 1980; Dunkerton *et al.*, , 1981; Killworth and McIntyre, 1985):

$$\nabla\cdot\mathbf{S}_{\mathrm{g}} = \rho\overline{v'_{\mathrm{g}}q'_{\mathrm{g}}}, \tag{3.1.3}$$

where $\mathbf{S}_{\mathrm{g}}$ is the quasi-geostrophic version of EP flux, v'_{g} the y-component quasi-geostrophic eddy velocity, and q'_{g} the quasi-geostrophic potential vorticity. Thus, Eq.(3.1.3) shows that the divergence of the guasi-geostrophic EP flux, $\nabla\cdot\mathbf{S}_{\mathrm{g}}$, is equal to the northward eddy flux of the quasi-geostrophic potential vorticity. On the other hand, according to the quasi-geostrophic version of the generalised Charney-Drazin theorem (Edmon *et al.*, 1980), $\nabla\cdot\mathbf{S}_{\mathrm{g}}$ represents the sole forcing of the mean state by the disturbances, and is closely related (but not equal) to the quasimomentum flux divergence (Andrews, 1987). Therefore Eq.(3.1.3) can be regarded as a quasi-geostrophic generalisation of the Taylor identity.

However, Eq.(3.1.3) cannot be applied near the equator, where geostrophic approximation breaks down, nor is it applicable in regions with sharp meridional variation of the static stability (Hoskins, 1991). In addition, the EP flux divergence that is expected to express quasimomentum flux divergence unfortunately includes the difference between the residual-mean and the Lagrangian-mean advection of angular momentum (Iwasaki, 1989). It is desirable, therefore, to derive a more general relation which may avoid such problem. Such an effort has been made by McIntyre and Norton (1990). In favour of the

[1] Here we assume that the wave in question is steady.

usefulness of PV, McIntyre and Norton argued that to any wave-induced contribution to the PVS transport, $\overline{\mathbf{J}^{\mathrm{w}}}$ say, there always corresponds an 'effective mean force' per unit mass

$$\overline{\mathbf{F}}_{\mathrm{eff}} = \overline{\mathbf{J}^{\mathrm{w}}} \times \nabla\bar{\theta}/|\nabla\bar{\theta}|^2, \text{ say,} \tag{3.1.4}$$

that would have the same effect as the waves upon the PVS transport. Moreover, they gave some examples to show that the effective mean force defined by Eq.(3.1.4) is equal to the horizontal projection of the rate of dissipation of quasimomentum.

Motivated by the examples given in McIntyre and Norton's paper, the objective of the present chapter is to develop some general relations between the wave-induced PVS transport and wave-induced momentum transport. Since the most convenient framework for our discussion is the generalised Lagrangian-mean (GLM) formulation in which the quasimomentum can be defined exactly (Andrews and McIntyre, 1978b,c), in Section 3.2 we first recall some basic properties of the GLM theory. Some of these properties also have application in other chapters. In Section 3.3 we demonstrate how the wave-induced PVS transport across a material surface is related to the rate of dissipation of quasimomentum. In Section 3.4, a transport equation for the mean PV is constructed, and a generalised Taylor identity is given within the framework of non-geostrophic theory for a stably stratified, rapidly rotating fluid.

3.2 Some basic properties of the GLM theory

Since this chapter concentrates exclusively on the PVS transport in the GLM framework, it is desirable to recall briefly some basic properties of the GLM description. Those who desire a more detailed discussion are referred to the excellent papers by Andrews and McIntyre (1978b,c).

The GLM description is in fact a hybrid Eulerian–Lagrangian description in which the GLM flow is described by equations in Eulerian form, with spatial position $\mathbf{x}$ and time t as independent variables rather than initial particle position and time. The essential idea is to average over positions displaced by the disturbance motion. Let Ξ be the position vector of particles displaced by the waves. The particle displacement $\boldsymbol{\xi}'$ associated with the waves is defined as a function of $\mathbf{x}$ and t by the mapping $\mathbf{x} \to \mathbf{x} + \boldsymbol{\xi}'(\mathbf{x}, t)$, i.e.,

$$\boldsymbol{\xi}'(\mathbf{x}, t) = \Xi(\mathbf{x}, t) - \mathbf{x}. \tag{3.2.1}$$

The GLM operator $\overline{(\,)}^{\mathrm{L}}$ entails averaging over particles at the displaced positions $\mathbf{x} + \boldsymbol{\xi}'(\mathbf{x}, t)$, namely

$$\overline{\varphi(\mathbf{x},t)}^{\mathrm{L}} \equiv \overline{\varphi^{\xi}(\mathbf{x},t)}, \tag{3.2.2}$$

where

$$\varphi^{\xi}(\mathbf{x},t) \equiv \varphi(\mathbf{x} + \boldsymbol{\xi}'(\mathbf{x},t), t), \tag{3.2.3}$$

and $\overline{(\,)}$ denotes a corresponding Eulerian mean at $\mathbf{x}$ in any of the usual senses (time, space, ensemble, etc., depending on the problem to be examined). The difference between φ^{ξ} and $\overline{\varphi}^{\mathrm{L}}$, denoted by φ^{l}, is called the Lagrangian disturbance of φ, i.e.,

$$\varphi^{l} \equiv \varphi^{\xi} - \overline{\varphi}^{\mathrm{L}}. \tag{3.2.4}$$

It is evident that

$$\overline{\varphi^{l}} = 0. \tag{3.2.5}$$

In addition, the parcel displacement vector $\boldsymbol{\xi}'$ can be equivalently defined by

$$\frac{\overline{D}^{\mathrm{L}}\boldsymbol{\xi}'}{Dt} = \mathbf{u}^{l}, \tag{3.2.6}$$

where $\overline{D}^{\mathrm{L}}/Dt$ is the Lagrangian-mean material derivative, defined as

$$\frac{\overline{D}^{\mathrm{L}}}{Dt} = \frac{\partial}{\partial t} + \overline{\mathbf{u}}^{\mathrm{L}} \cdot \nabla. \tag{3.2.7}$$

The difference between the Eulerian mean and the Lagrangian mean of φ is referred to as the Stokes correction $\overline{\varphi}^{\mathrm{S}}$:

$$\overline{\varphi(\mathbf{x},t)}^{\mathrm{S}} \equiv \overline{\varphi(\mathbf{x},t)}^{\mathrm{L}} - \overline{\varphi(\mathbf{x},t)}. \tag{3.2.8}$$

When φ is velocity $\mathbf{u}$, the Stokes correction $\overline{\mathbf{u}}^{\mathrm{S}}$ is sometimes referred to as the Stokes 'drift'.

For small-amplitude waves, Taylor expansion of φ^{ξ} shows that

$$\begin{aligned}\varphi^{\xi}(\mathbf{x},t) &= \varphi(\mathbf{x},t) + \boldsymbol{\xi}'\cdot\nabla\left[\varphi(\mathbf{x},t)\right] + \tfrac{1}{2}\boldsymbol{\xi}'\cdot(\boldsymbol{\xi}'\cdot\nabla)\nabla\left[\varphi(\mathbf{x},t)\right] + O(a^3\varphi)\\ &= \overline{\varphi} + \varphi' + \boldsymbol{\xi}'\cdot\nabla\left(\overline{\varphi}+\varphi'\right) + \tfrac{1}{2}\boldsymbol{\xi}'\cdot(\boldsymbol{\xi}'\cdot\nabla)\nabla\overline{\varphi} + O(a^2\varphi') + O(a^3\overline{\varphi}),\end{aligned} \tag{3.2.9}$$

since $\overline{\varphi} + \varphi' = \varphi(\mathbf{x}, t)$ by definition. Application of the Eulerian-mean operator $\overline{(\;)}$ to Eq.(3.2.9) gives the following result immediately,

$$\overline{\varphi}^{\mathrm{S}} = \overline{(\boldsymbol{\xi}' \cdot \nabla)\varphi'} + \tfrac{1}{2}\overline{\boldsymbol{\xi}' \cdot (\boldsymbol{\xi}' \cdot \nabla)\nabla\overline{\varphi}} + O(a^2\varphi'). \tag{3.2.10}$$

Substituting Eq.(3.2.9) into Eq.(3.2.4) and using Eqs.(3.2.8) and (3.2.10), gives

$$\varphi^l = \varphi' + \boldsymbol{\xi}' \cdot \nabla\overline{\varphi} + O(a\varphi'). \tag{3.2.11}$$

In the GLM description, the quasimomentum per unit mass can be defined exactly as

$$\mathrm{p}_i(\mathbf{x}, t) = -\overline{\xi'_{j,i}\{u^l_j + (\boldsymbol{\Omega} \times \boldsymbol{\xi}')_j\}}, \tag{3.2.12}$$

where the Cartesian tensor and vector notations have been used interchangeably, with $\mathbf{A} = A_j$, $(\;)_{,j} = \partial/\partial x_j$, and summation of repeated indices over the values 1, 2, and 3.

Three basic results of the GLM theory for any $\varphi(\mathbf{x}, t)$ can be expressed as

$$\overline{\left(\frac{D\varphi}{Dt}\right)^{\xi}} = \frac{\overline{D}^{\mathrm{L}}\overline{\varphi}^{\mathrm{L}}}{Dt}, \tag{3.2.13}$$

$$\left(\frac{D\varphi}{Dt}\right)^{l} = \frac{\overline{D}^{\mathrm{L}}\varphi^{l}}{Dt}, \tag{3.2.14}$$

$$\left(\varphi_{,i}\right)^{\xi} = \left(\varphi^{\xi}\right)_{,j} K_{ij}/J, \tag{3.2.15}$$

where J is the Jacobian of the mapping $\mathbf{x} \to \mathbf{x} + \boldsymbol{\xi}'$, defined as

$$J \equiv \det\left\{\delta_{ij} + \xi'_{i,j}\right\}, \tag{3.2.16}$$

and K_{ij} is the (i, j)th cofactor of J.

Using the definitions and properties described above, Andrews and McIntyre (1978b) derived the exact equations of GLM motion, which, after some manipulation, can be written as

$$\frac{\partial}{\partial t}(\overline{\mathbf{u}}^{\mathrm{L}} - \mathbf{p}) + \left[2\boldsymbol{\Omega} + \nabla \times (\overline{\mathbf{u}}^{\mathrm{L}} - \mathbf{p})\right] \times \overline{\mathbf{u}}^{\mathrm{L}}$$
$$+\nabla\left[\overline{\mathbf{u}}^{\mathrm{L}} \cdot (\overline{\mathbf{u}}^{\mathrm{L}} - \mathbf{p}) + \overline{\Pi}\right] - \overline{T(\overline{\theta}^{\mathrm{L}}, p^{\xi})\nabla\overline{\theta}^{\mathrm{L}}} = \overline{\mathbf{F}}^{\mathrm{L}} + \mathcal{F}, \tag{3.2.17}$$

$$\frac{\overline{D}^{\mathrm{L}}\overline{\theta}^{\mathrm{L}}}{Dt} = \overline{Q}^{\mathrm{L}}, \tag{3.2.18}$$

$$\frac{\overline{D}^{\mathrm{L}}\tilde{\rho}}{Dt} + \tilde{\rho}\nabla \cdot \overline{\mathbf{u}}^{\mathrm{L}} = 0, \tag{3.2.19}$$

where

$$\Pi \equiv \overline{\mathcal{E}(\overline{\theta}^{\mathrm{L}}, p^{\xi})} + \overline{\Phi}^{\mathrm{L}} - \overline{\mathbf{u}^{\xi} \cdot \left[\tfrac{1}{2}\mathbf{u}^{\xi} + (\Omega \times \boldsymbol{\xi}')\right]}, \tag{3.2.20}$$

$$\mathcal{F} = \mathcal{F}_i = \overline{\xi'_{j,i} F^l_j} + \overline{(p^l)_{,i}\, q}, \tag{3.2.21}$$

$$\tilde{\rho} = \rho^{\xi} J, \tag{3.2.22}$$

with $\mathcal{E}$ representing the enthalpy per unit mass and

$$q \equiv -\frac{1}{\rho(\theta^{\xi}, p^{\xi})} + \frac{1}{\rho(\overline{\theta}^{\mathrm{L}}, p^{\xi})}. \tag{3.2.23}$$

Other symbols are the same as in Section 1.5. It can be checked that $q = 0$ if the fluid motion is adiabatic. Thus $\mathcal{F}$ is a wave property explicitly involving departures from conservative motion. It was recognised by Andrews and McIntyre (1978b,c) as the rate of dissipation of quasimomentum.

Finally, we note that for any vector field $\mathbf{A}(\mathbf{x}, t)$, an associated vector $\tilde{\mathbf{A}}$ can be defined as

$$\tilde{\mathbf{A}} = \tilde{A}_j = K_{ij} A^{\xi}_i. \tag{3.2.24}$$

If $\mathbf{A}$ is governed by an equation

$$\frac{\partial \mathbf{A}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{A}) = 0, \tag{3.2.25}$$

then $\tilde{\mathbf{A}}$ satisfies the following equation (Moffatt, 1978):

$$\frac{\partial \tilde{\mathbf{A}}}{\partial t} - \nabla \times (\overline{\mathbf{u}}^{\mathrm{L}} \times \tilde{\mathbf{A}}) = 0, \tag{3.2.26}$$

and $\tilde{\mathbf{A}}$ can be regarded as a mean quantity in the sense that

$$\overline{\tilde{\mathbf{A}}} = \tilde{\mathbf{A}} \tag{3.2.27}$$

(Andrews and McIntyre, 1978b). When the right-hand side of Eq.(3.2.25) is different from zero, $\tilde{\mathbf{A}}$ is not generally a mean quantity.

3.3 Wave-induced contribution to the PVS transport: Integral properties

In Section 2.5 we have demonstrated how to evaluate the PVS budget following a material element. This section considers a slightly different situation, in which the PVS budget

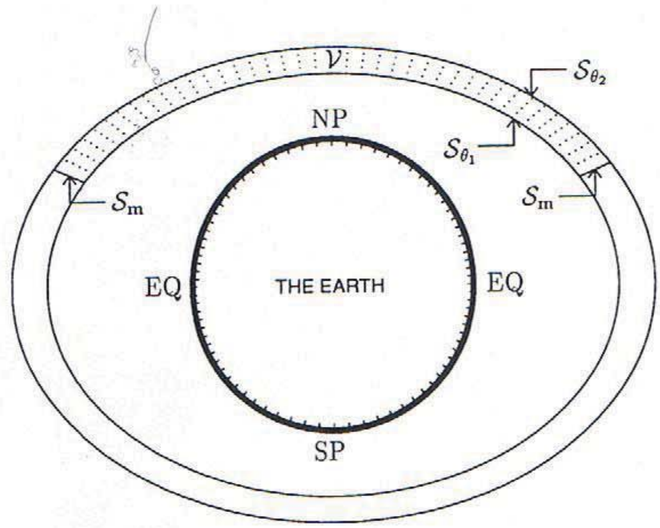


Figure 3.1: A specified volume element of fluid V in the Earth's atmosphere. V is bounded by two isentropic surfaces S_{θ_1}, S_{θ_2} and a lateral material surface S_{m} that is initially parallel to a fixed latitude.

within a volume $\mathcal{V}$ shown in Fig.3.1 is examined. Here $\mathcal{V}$ is chosen to be enclosed by two nearby isentropic surfaces S_{θ_1}, S_{θ_2}, and a lateral material surface S_{m} which is initially parallel to a latitude. If the diabatic heating is zero everywhere, then $\mathcal{V}$ can be identified as a material mass of fluid since mass cannot cross the isentropic surfaces under such circumstances.

The total amount of PVS within the volume $\mathcal{V}$, denoted by $\mathcal{P}(t)$, is given by,

$$\mathcal{P}(t) = \iiint_{\mathcal{V}} \rho P d\tau = \iiint_{\mathcal{V}} \boldsymbol{\zeta}_{\mathrm{a}} \cdot \nabla\theta \, d\tau = \iiint_{\mathcal{V}} \nabla\cdot(\boldsymbol{\zeta}_{\mathrm{a}}\theta) d\tau = \iint_{\mathcal{S}} \theta \boldsymbol{\zeta}_{\mathrm{a}} \cdot \hat{\mathbf{n}} \, dS, \qquad (3.3.1)$$

where P is the PV defined by Eq.(1.1.1). With reference to Fig.3.1, Eq.(3.3.1) may be written as

$$\mathcal{P}(t) = \delta\theta \iint_{\mathcal{S}_{\theta_2}} \boldsymbol{\zeta}_{\mathrm{a}} \cdot \hat{\mathbf{n}} dS + O\left[(\delta\theta)^2\right] = C_{\mathrm{a}}\delta\theta + O\left[(\delta\theta)^2\right], \qquad (3.3.2)$$

where the $O\left[(\delta\theta)^2\right]$ term represents contribution from the side surface $\mathcal{S}_{\mathrm{m}}$, C_{a} is the absolute circulation around the circuit $\Gamma^{\dagger}$, which is the intersection of the lateral material surface $\mathcal{S}_{\mathrm{m}}$ with the isentropic surface $\mathcal{S}_{\theta_2}$. Eq.(3.3.2) shows that in the limit $\delta\theta \to 0$, the contribution from the side surface $\mathcal{S}_{\mathrm{m}}$ to the surface integral in Eq.(3.3.1) is negligible compared to those from the isentropic surfaces $\mathcal{S}_{\theta_1}$ and $\mathcal{S}_{\theta_2}$.

For weak thermal dissipation, $\Gamma^{\dagger}$ can be regarded approximately as a material curve Γ^{ξ}, and then

$$\begin{aligned}
C_{\mathrm{a}} &= \oint_{\Gamma^{\xi}} (\mathbf{u} + \boldsymbol{\Omega} \times \boldsymbol{\Xi}) \cdot d\boldsymbol{\Xi} \\
&= \oint_{\Gamma} \left\{ \overline{u}_j^{\mathrm{L}} + u_j^l + [\boldsymbol{\Omega} \times (\mathbf{x} + \boldsymbol{\xi})]_j \right\} \Xi_{j,i} d\ell_i \\
&= \oint_{\Gamma} \left\{ \overline{u}_j^{\mathrm{L}} + u_j^l + [\boldsymbol{\Omega} \times (\mathbf{x} + \boldsymbol{\xi})]_j \right\} (\delta_{ij} + \xi'_{i,j}) d\ell_i \\
&= \oint_{\Gamma} \left[\overline{u}_i^{\mathrm{L}} - \mathrm{p}_i + (\boldsymbol{\Omega} \times \mathbf{x})_i \right] d\ell_i,
\end{aligned} \tag{3.3.3}$$

where p_i is the ith component of the quasimomentum per unit mass defined in Eq.(3.2.12), Γ is the image of Γ^{ξ} under the inverse mapping $\boldsymbol{\Xi} \rightarrow \mathbf{x}$, and $d\ell_i$ represents a displacement vector locally tangent to Γ.

Differentiation of Eq.(3.3.2) with respect to time t and use of Eq.(3.3.3), gives

$$\begin{aligned}
\frac{d\mathcal{P}(t)}{dt} &= \delta\theta \frac{dC_{\mathrm{a}}}{dt} + O\left[(\delta\theta)^2\right] \\
&= \delta\theta \frac{d}{dt} \oint_{\Gamma} \left[\overline{u}_i^{\mathrm{L}} - \mathrm{p}_i + (\boldsymbol{\Omega} \times \mathbf{x})_i \right] d\ell_i + O\left[(\delta\theta)^2\right] \\
&= \delta\theta \oint_{\Gamma} \left\{ \frac{\overline{D}^{\mathrm{L}}}{Dt} \left[\overline{u}_i^{\mathrm{L}} - \mathrm{p}_i + (\boldsymbol{\Omega} \times \mathbf{x})_i \right] + \overline{u}_{j,i}^{\mathrm{L}} \left[\overline{u}_j^{\mathrm{L}} - \mathrm{p}_j + (\boldsymbol{\Omega} \times \mathbf{x})_j \right] \right\} d\ell_i + O\left[(\delta\theta)^2\right],
\end{aligned} \tag{3.3.4}$$

where the last step follows immediately from the standard identity (Andrews and McIntyre, 1978b)

$$\frac{d}{dt} \oint_{\Gamma} \varphi_i d\ell_i = \oint_{\Gamma} \left(\frac{\overline{D}^{\mathrm{L}} \varphi_i}{Dt} + \overline{u}_{j,i}^{\mathrm{L}} \varphi_j \right) d\ell_i. \tag{3.3.5}$$

Now substituting Eq.(3.2.17) into Eq.(3.3.4), with some manipulation we obtain

$$\frac{d\mathcal{P}(t)}{dt} = \delta\theta \oint_{\Gamma} \left(\overline{F}_i^{\mathrm{L}} + \mathcal{F}_i \right) d\ell_i + O\left[(\delta\theta)^2\right], \tag{3.3.6}$$

where the fact that Γ lies in a surface of constant $\overline{\theta}^{\mathrm{L}}$ has been used. Note that $\mathcal{F}_i$ is the rate of dissipation of quasimomentum defined by Eq.(3.2.21).

According to the impermeability theorem, there can be no transport of PVS across isentropic surfaces. This implies that the total amount of PVS within $\mathcal{V}$ can change only as a result of non-advective PVS transport across the lateral material surface S_{m}. Therefore Eq.(3.3.6) shows that the wave-induced contribution to the PVS transport across a material surface is related closely to the rate of dissipation of quasimomentum. It thus generalizes the result obtained by Taylor (1915) for an incompressible, two-dimensional, non-rotating fluid.

3.4 Wave-induced contribution to the PVS transport: Differential properties

In the preceding section we discussed the integral properties of PVS transport. The majority of problems in fluid dynamics, however, require use of equations in differential form. Eq.(3.1.3) is such a differential relation. It shows that within the framework of quasi-geostrophic theory, the EP flux divergence, which is related closely to the rate of dissipation of quasimomentum, is equal to the meridional eddy flux of potential vorticity. Since the constraint of the geostrophic approximation may be a severe limitation in practice, it is desirable to derive a more general differential equation, which may relate the wave-induced PVS transport directly to the rate of dissipation of quasimomentum. The existence of such a relation is suggested by examples discussed in McIntyre and Norton (1990) and hinted in the integral equation (3.3.6).

First, let us have a quick look at the exact GLM PV transport equation,

$$\frac{\partial \tilde{\rho}\overline{P}^{\mathrm{L}}}{\partial t} + \nabla \cdot (\tilde{\rho}\overline{P}^{\mathrm{L}}\overline{\mathbf{u}}^{\mathrm{L}}) = \tilde{\rho}\overline{\rho^{-1}\nabla \cdot (\,Q\boldsymbol{\zeta}_{\mathrm{a}} + \mathbf{F} \times \nabla\theta\,)}^{\mathrm{L}}, \tag{3.4.1}$$

which is obtained by applying the GLM operator $\overline{(\,)}^{\mathrm{L}}$ directly to Eq.(1.5.14) and using Eqs.(3.2.13) and (3.2.19). The most important information shown in Eq.(3.4.1) is that there are no advective eddy flux terms in the GLM PV transport equation, so that only dissipative processes can be responsible for the wave-induced PVS transport in the GLM framework.

It is difficult to see any way of simplifying Eq.(3.4.1) further. In particular, the wave-induced contribution to the PVS transport cannot be easily separated from the total nonadvective PVS transport on the right-hand side of this equation.

To show how the wave-induced contribution to the PVS transport is connected to the rate of dissipation of quasimomentum, we now turn to consider a useful alternative to Eq.(3.4.1). Let us start by noting that the vorticity equation can be written as

$$\frac{\partial \boldsymbol{\zeta}_{\mathrm{a}}}{\partial t} - \nabla \times (\mathbf{u} \times \boldsymbol{\zeta}_{\mathrm{a}}) = \frac{\nabla\rho \times \nabla p}{\rho^2} + \nabla \times \mathbf{F}. \tag{3.4.2}$$

It shows that if the fluid is barotropic and inviscid, the absolute vorticity $\boldsymbol{\zeta}_{\mathrm{a}}$ satisfies Eq.(3.2.25). Under such circumstances, the associated absolute vorticity $\widetilde{\boldsymbol{\zeta}}_{\mathrm{a}}$, defined by Eq.(3.2.24) with $\boldsymbol{\zeta}_{\mathrm{a}}$ replacing $\mathbf{A}$, should be a mean quantity, and then $\overline{P}^{\mathrm{L}}$ by definition can be manipulated as

$$\overline{P}^{\mathrm{L}} = \overline{\left[\rho^{-1}(\zeta_{\mathrm{a}})_i\, \theta_{,i}\right]^{\xi}} = \overline{\tilde{\rho}^{-1} K_{ij}(\zeta_{\mathrm{a}}^{\xi})_i\, \theta^{\xi}_{,j}} = \overline{\tilde{\rho}^{-1}(\widetilde{\zeta}_{\mathrm{a}})_j\, \theta^{\xi}_{,j}} = \tilde{\rho}^{-1}\overline{\widetilde{\boldsymbol{\zeta}}}_{\mathrm{a}} \cdot \nabla\overline{\theta}^{\mathrm{L}}, \tag{3.4.3}$$

using Eqs.(3.2.2), (3.2.15), (3.2.22) and (3.2.24), and (3.2.27).

In the presence of baroclinicity and/or viscous force, the last expression in Eq.(3.4.3) is not justified because there is no guarantee that the associated absolute vorticity $\widetilde{\boldsymbol{\zeta}}_{\mathrm{a}}$ is a mean quantity under such circumstances. In any case, nevertheless, we can define a 'quasi-GLM PV', $\widetilde{P}$ say, by

$$\widetilde{P} \equiv \tilde{\rho}^{-1}\overline{\widetilde{\boldsymbol{\zeta}}}_{\mathrm{a}} \cdot \nabla\overline{\theta}^{\mathrm{L}}, \tag{3.4.4}$$

which is always a mean quantity in the sense that $\overline{\widetilde{P}} = \widetilde{P}$. Note that $\widetilde{P}$ is approximately equal to $\overline{P}^{\mathrm{L}}$ whenever $\widetilde{\boldsymbol{\zeta}}_{\mathrm{a}}$ is approximately a mean quantity; in particular, $\widetilde{P}$ is identical to $\overline{P}^{\mathrm{L}}$ in the case of barotropic and inviscid fluids, as shown in Eq.(3.4.3).

In addition, Andrews and McIntyre (1978b) showed that

$$\overline{\widetilde{\boldsymbol{\zeta}}}_{\mathrm{a}} = 2\boldsymbol{\Omega} + \nabla \times (\overline{\mathbf{u}}^{\mathrm{L}} - \mathbf{p}), \tag{3.4.5}$$

where $\mathbf{p}$ is the quasimomentum density defined in Eq.(3.2.12).

Now taking the curl of Eq.(3.2.17) yields

$$\frac{\partial\overline{\widetilde{\boldsymbol{\zeta}}}_{\mathrm{a}}}{\partial t} + \nabla \times \left(\overline{\widetilde{\boldsymbol{\zeta}}}_{\mathrm{a}} \times \overline{\mathbf{u}}^{\mathrm{L}}\right) - \nabla \times \left[\overline{T}(\overline{\theta}^{\mathrm{L}}, p^{\xi})\nabla\overline{\theta}^{\mathrm{L}}\right] = \nabla \times \left(\overline{\mathbf{F}}^{\mathrm{L}} + \boldsymbol{\mathcal{F}}\right). \tag{3.4.6}$$

Forming the scalar product of Eq.(3.4.6) with $\nabla\overline{\theta}^{\mathrm{L}}$ leads to

$$\frac{\partial}{\partial t}\left(\tilde{\rho}\widetilde{P}\right) + \nabla \cdot \left[\tilde{\rho}\widetilde{P}\overline{\mathbf{u}}^{\mathrm{L}} - \overline{Q}^{\mathrm{L}}\overline{\widetilde{\boldsymbol{\zeta}}}_{\mathrm{a}} - \overline{\mathbf{F}}^{\mathrm{L}} \times \nabla\overline{\theta}^{\mathrm{L}}\right] = -\nabla \cdot \overline{\mathbf{J}^{\mathrm{w}}}, \tag{3.4.7}$$

where

$$\overline{\mathbf{J}^{\mathrm{w}}} = -\boldsymbol{\mathcal{F}} \times \nabla\overline{\theta}^{\mathrm{L}}. \tag{3.4.8}$$

In the absence of baroclinicity and viscous force, Eq.(3.4.7) is equivalent to Eq.(3.4.1). Under general circumstances, it represents a transport equation for the quasi-GLM PV. Note that no eddy quantity appears on the left-hand side of this equation. Thus, the wave-induced contribution to the PVS transport is represented solely by the eddy quantity $\overline{\mathbf{J}^{\mathrm{w}}}$ on the right-hand side of Eq.(3.4.7).

The relationship between the wave-induced PVS transport ($\overline{\mathbf{J}^{\mathrm{w}}}$) and the rate of dissipation of quasimomentum ($\mathcal{F}$) is expressed by Eq.(3.4.8). This is a general relation, valid in a three-dimensional, compressible, rotating fluid without the constraint of the geostrophic approximation. Thus it, together with the integral relation (3.3.6), generalizes straightforwardly the well-known result of Taylor (1915) for an incompressible, two-dimensional, non-rotating fluid and the result of Bretherton (1966a,b) which is valid only within the framework of quasi-geostrophic theory.

McIntyre and Norton (1990) pointed out that, to any wave-induced contribution to the PVS transport, there always corresponds an effective mean force per unit mass, which may be defined by Eq.(3.1.4) and would have the same effect as the waves upon the PVS transport. When Eq.(3.4.8) is substituted into Eq.(3.1.4), we see that the effective mean force may be given as

$$\overline{\mathbf{F}}_{\mathrm{eff}} = \frac{\overline{\mathbf{J}^{\mathrm{w}}} \times \nabla\overline{\theta}^{\mathrm{L}}}{|\nabla\overline{\theta}^{\mathrm{L}}|^2} = -\frac{\left(\mathcal{F} \times \nabla\overline{\theta}^{\mathrm{L}}\right) \times \nabla\overline{\theta}^{\mathrm{L}}}{|\nabla\overline{\theta}^{\mathrm{L}}|^2} = \mathcal{F} - \frac{\left(\mathcal{F} \cdot \nabla\overline{\theta}^{\mathrm{L}}\right) \nabla\overline{\theta}^{\mathrm{L}}}{|\nabla\overline{\theta}^{\mathrm{L}}|^2} = \mathcal{F}_{\|}, \qquad (3.4.9)$$

where "$\|$" denotes projection parallel to the $\overline{\theta}^{\mathrm{L}}$-surface.

Eq.(3.4.9) shows that the effective mean force defined by Eq.(3.1.4) is equal to the rate of dissipation of quasimomentum projected on the mean isentropic surface. This result has been confirmed by examples analyzed in McIntyre and Norton (1990).

3.5 Discussion

The invertibility principle has been widely recognised as a powerful simplifying principle in geophysical fluid dynamics. It enables us to deduce all the relevant dynamical information from the knowledge of a single scalar field (PV) with suitable boundary and balance conditions. Applying this principle to the wave-mean interaction theory, we have shown in Chapter 2 that the description of the wave-induced mean flow from a PVS viewpoint is, at least to the leading order, equivalent to the description of the same phenomenon from a momentum viewpoint. This drops a strong hint of an inherent relation between the wave-induced PVS transport and the wave-induced momentum transport. In this chapter, we demonstrated that such a relation has its origin in the Taylor identity derived for an incompressible, two-dimensional, non-rotating fluid (Taylor, 1915). The generalised Taylor identity, i.e. the integral relation (3.3.6) or the differential relation (3.4.8), strongly support the idea that the general nature of the dissipative type of wave-induced mean motion can be understood in a unified and succinct way by viewing all the phenomena in terms of the wave-induced PV transport.

CHAPTER 4

ROSSBY WAVES AND THEIR MEAN EFFECTS IN A CHARNEY–DRAZIN MODEL ON THE MID-LATITUDE β-PLANE

In this and the next chapter, extratropical Rossby waves and their mean effects will be examined. The basic equation for dynamical analysis is the quasi-geostrophic potential vorticity (QGPV) equation. It is understood that the QGPV equation has fallen out of favour as the equation for modelling the global atmospheric circulation since it is not uniformly valid as one approaches the equator, or sloping tropopause transitions in static stability. However, it remains of great value in diagnosing and gaining insight into the dominant dynamical processes in the extratropical regions.

4.1 Introduction

Rossby waves, which are also frequently referred to as *planetary waves*, have been generally recognised as the most important waves for large-scale meteorological and oceanographical processes. These waves owe their existence to the latitudinal variation of PV on isentropic surfaces (Hoskins *et al.*, 1985). In atmospheric studies there has been considerable emphasis on the vertical propagation of Rossby waves. The primary motivation has been the importance of Rossby waves for the dynamics of the middle atmosphere, as reviewed in Holton (1975, 1992), Dickinson (1978), and Andrews *et al.* (1987).

Observations have long suggested that planetary-scale Rossby waves in the troposphere can be traced into the winter stratosphere. The simplest and most illuminating theory for the upward propagation of Rossby waves is due to Charney and Drazin (1961). In a quasi-geostrophic model with constant mean zonal flow and constant static stability, Charney and Drazin demonstrated that stationary Rossby waves can only occur for eastward zonal winds greater than zero and less than a cutoff velocity defined in Eq.(1.4.1). They found in particular that the range criteria in Eq.(1.4.1) were so restrictive for higher zonal wavenumbers that only the largest scale, essentially zonal wavenumbers one and two, would be expected to propagate significantly into the winter stratosphere. This is in qualitative agreement with observations (see Holton, 1975; Andrews *et al.* 1987).

In the present chapter we consider how Rossby waves in the Charney–Drazin model are modified by dissipative processes, and how these dissipating waves modify the mean flow. According to Charney and Drazin's theory, substantial vertical energy propagation should occur during the equinoctial seasons when the mean winds are weak westerlies in the stratosphere and mesosphere. However, there is little evidence to support this argument. After examining Rossby-wave behaviour in the presence of radiative dissipation, Dickinson (1969b) pointed out that the absence of large internal Rossby waves at the equinoxes could be ascribed in part to the damping of Rossby waves by Newtonian cooling as they propagate upward through the weak westerlies in the middle atmosphere. Considerable sensitivity of the wave amplitude to the strength of the dissipation was also reported by Schoeberl and Geller (1977) in a spherical model. In addition, the importance of wave dissipation has been further emphasized in studies of the interactions between waves and zonal flow. According to the generalized Charney–Drazin theorem, the divergence of the EP flux, which serves as the effective force per unit mass acting on the zonal mean flow due to disturbances, is directly related to wave dissipation and transience (Andrews and McIntyre, 1976a,b, 1978a; Boyd, 1976). Andrews and McIntyre (1976a, 1978b) pointed out that the wave transience associated with a temporary conservative disturbance can force only a temporary mean-flow change. In other words, the mean effect of wave transience tends to average out as time goes on, leaving systematic modification of the mean flows due to wave dissipation as the dominant effect in the long time average (McIntyre and Norton, 1990).

In Section 4.2, a QGPV equation is derived on the mid-latitude β-plane. Based on this equation and the assumptions of Rayleigh friction and Newtonian cooling, linear and steady solutions for dissipating Rossby waves are examined in Section 4.3. Their mean effects are considered in Section 4.4. Our emphasis will be on possibilities for the existence of divergent EP fluxes, which have frequently emerged in the winter stratosphere and mesosphere in observational studies (e.g., Hamilton, 1982; Geller *et al.*, 1983, 1984; Hartmann *et al.*, 1984; Mechoso *et al.*, 1985; Marks, 1989; Rosenlof and Holton, 1993), and have caused some interesting arguments (Mahlman and Umscheid, 1984; Robinson, 1986; Andrews, 1987; Marks, 1989).

Another objective of the present chapter (Section 4.5) is to re-examine the difference between the transformed Eulerian-mean (TEM) and the generalized Lagrangian-mean

(GLM) meridional circulations. This problem is directly relevant to tracer transport in the middle atmosphere. It has been widely recognized that tracer transport is closely related to the GLM flow (Dunkerton, 1978; Andrews *et al.*, 1987). Dunkerton (*op. cit.*) further suggested that the GLM meridional circulation could be approximated by the TEM meridional circulation. In contrast to this argument, Rood & Schoeberl (1983) found that using the TEM circulation instead of the GLM circulation would significantly overestimate the advection by Rossby waves. It will be demonstrated in Section 4.5 that, in the presence of thermal dissipation, the difference between the TEM and GLM circulation is unlikely to be negligible. In fact, they are not only different, but also opposite-signed in some situations.

4.2 The β-plane approximation in the mid-latitudes

In order to avoid the complexity of spherical geometry, in Chapter 2 we considered inertia-gravity waves on an f-plane, in which the Coriolis parameter f was taken as constant. To examine the dynamics of a Rossby wave whose typical horizontal length scale is over 1,000 km, it is necessary to retain the dynamical effect of the variation of f with latitude in the Coriolis force term in the momentum equation. This variation can be approximated by expanding the latitudinal dependence of f in a Taylor series about a reference latitude ϕ_0 and retaining up to the linear term. This approximation is usually referred to as the *β-plane approximation* (Rossby, 1939; Phillips, 1963).

In this section, we invoke the mid-latitude β-plane approximation to obtain the quasi-geostrophic equations. Our starting point is the *primitive equations* in spherical coordinates.

4.2.1 The primitive equations of atmosphere in spherical coordinates

Most motions in the atmosphere can be described by the primitive equations, which consist of the horizontal momentum equation, the hydrostatic approximation, the continuity equation, and the thermodynamic energy equation. To analyze Rossby waves in the stratosphere it is convenient to write these equations in the log-pressure coordinate system, in which the vertical coordinate is defined by (see Holton, 1975; Andrews *et al.*, 1987),

$$z \equiv -H \ln(p/p_s), \tag{4.2.1}$$

where p is pressure and p_s a standard reference pressure, H is the scale height of the atmosphere. Now the primitive equations in spherical coordinates can be expressed as follows,

$$\frac{Du}{Dt} - \left(f + \frac{u}{a_E}\tan\phi\right)v + \frac{1}{a_E\cos\phi}\frac{\partial\Phi}{\partial\lambda} = X, \tag{4.2.2}$$

$$\frac{Dv}{Dt} + \left(f + \frac{u}{a_E}\tan\phi\right)u + \frac{1}{a_E}\frac{\partial\Phi}{\partial\phi} = Y, \tag{4.2.3}$$

$$\frac{\partial\Phi}{\partial z} = \frac{RT}{H}, \quad \frac{d\Phi_B}{dz} = \frac{RT_B}{H}, \tag{4.2.4}$$

$$\frac{1}{a_E\cos\phi}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial}{\partial\phi}(v\cos\phi)\right] + \frac{1}{\rho_B}\frac{\partial}{\partial z}(\rho_B w) = 0, \tag{4.2.5}$$

$$\frac{DT}{Dt} + \left[\frac{dT_B}{dz} + \frac{\kappa(T_B + T)}{H}\right]w = \frac{J}{c_p}, \tag{4.2.6}$$

where λ is the longitude, ϕ is the latitude, and $(u,\, v,\, w)$ are velocity components defined by

$$(u,\, v,\, w) \equiv \left[(a_E\cos\phi)\frac{D\lambda}{Dt},\, a_E\frac{D\phi}{Dt},\, \frac{Dz}{Dt}\right], \tag{4.2.7}$$

with the material derivative D/Dt defined by,

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \frac{u}{a_E\cos\phi}\frac{\partial}{\partial\lambda} + \frac{v}{a_E}\frac{\partial}{\partial\phi} + w\frac{\partial}{\partial z}. \tag{4.2.8}$$

In addition, $f = 2\Omega\sin\phi$ is the Coriolis parameter, R is the gas constant for dry air, c_p is the specific heat at constant pressure, κ is the ratio of R to c_p, a_E is the radius of the earth. The basic density $\rho_B(z)$ is defined by $\rho_B(z) \equiv \rho_s\exp(-z/H)$ with ρ_s being a standard reference density. $T_B(z)$ and $\Phi_B(z)$ are basic state temperature and geopotential, respectively, while T and Φ are departures of local temperature from T_B and of local geopotential from Φ_B, respectively. J is the diabatic heating rate per unit mass. X and Y are horizontal components of friction.

When the hydrostatic approximation, namely Eq.(4.2.4), is substituted into Eq.(4.2.6) and the fact that $T \ll T_B$ has been used as a basis for neglecting $\kappa wT/H$ compared to $\kappa wT_B/H$, we obtain a further simplified thermodynamic energy equation:

$$\frac{D}{Dt}\left(\frac{\partial\Phi}{\partial z}\right) + N^2 w = \frac{\kappa J}{H}, \tag{4.2.9}$$

where

$$N^2(z) \equiv \frac{R}{H}\left(\frac{dT_B}{dz} + \frac{\kappa T_B}{H}\right), \tag{4.2.10}$$

is the buoyancy frequency squared.

4.2.2 The β-plane approximation and quasi-geostrophic theory

Despite a few simplifications having been used in deriving it, the set of primitive equations is still very complicated. To focus on the large-scale motions in the mid-latitude region around some central latitude ϕ_0 distant from both the equator and the poles, we can introduce the mid-latitude β-plane approximation to simplify our analysis. Following Pedlosky (1987), we define x and y by

$$x = a_E \lambda \cos\phi_0, \qquad y = a_E(\phi - \phi_0). \tag{4.2.11}$$

Expanding the trigonometric functions in the primitive equations about the latitude ϕ_0, we obtain,

$$\sin\phi = \sin\phi_0 + \frac{y\cos\phi_0}{a_E} - \frac{y^2\sin\phi_0}{2{a_E}^2} + \cdots \tag{4.2.12}$$

$$\cos\phi = \cos\phi_0 - \frac{y\sin\phi_0}{a_E} - \frac{y^2\cos\phi_0}{2{a_E}^2} + \cdots \tag{4.2.13}$$

$$\tan\phi = \tan\phi_0 + \frac{y}{a_E\cos^2\phi_0} + \frac{y^2\tan\phi_0}{{a_E}^2\cos^2\phi_0} + \cdots \tag{4.2.14}$$

$$\frac{1}{\cos\phi} = \frac{1}{\cos\phi_0} + \frac{y\tan\phi_0}{a_E\cos\phi_0} + \frac{y^2(1+\sin^2\phi_0)}{2{a_E}^2\cos^3\phi_0} + \cdots. \tag{4.2.15}$$

Further dynamical simplifications involve systematically using scaling arguments. Our primary purpose is to deduce quantitative rules and simplified equations for large-scale motion of limited meridional extent. The resulting equations should consistently describe the essential dynamics of the assumed motion, while disregard the unnecessary complications inherent in the original equations. It should be kept in mind, however, that it may not be possible to find a single scaling approach which is appropriate for all the phenomena of interest. For one thing, the time scales of the large-scale motions in the atmosphere vary from several hours to many months, and the horizontal space scales from several hundred kilometers to global dimensions (Holton, 1975). For another thing, the scaling arguments assume that a partial derivative of a quantity, α say, will have an order of magnitude at most equal to the magnitude of α divided by the appropriate scale length (Phillips, 1963). Such assumption may fail in regions where variables have steep gradients, as the potential vorticity does at the edge of the polar vortex (McIntyre and Palmer, 1983, 1984). Nevertheless, it suffices for us to derive the model and find those conditions under which quasi-geostrophic solutions can exist (Charney, 1947, Phillips, 1963, Holton, 1975).

Let V, L, D and N_0 represent the typical magnitudes, or the scales, of horizontal velocity, horizontal length, vertical length, and buoyancy frequency, respectively. We further assume that geostrophic balance is the leading approximation of the motion, so that the typical magnitude of the geopotential Φ is $2\Omega VL$. Then we can define the following non-dimensional variables (denoted by an over-check):

$$\left.\begin{array}{lll} (\check{x},\check{y})=(x,y)/L, & \check{z}=z/D, & \check{t}=t/(L/V), \quad \check{\Phi}=\Phi/(2\Omega VL), \\ (\check{u},\check{v})=(u,v)/V, & \check{w}=w/(VD/L), & \check{N}=N/N_0, \quad \check{\rho}_{\rm B}=\rho_{\rm B}/\rho_{\rm S}, \\ (\check{X},\check{Y})=(X,Y)/(V^2/L), & & \check{J}=J/[2\Omega HV^2/(\kappa D)]. \end{array}\right\} \quad (4.2.16)$$

Substituting Eqs.(4.2.11)–(4.2.16) into Eqs.(4.2.2), (4.2.3), (4.2.5), and (4.2.9), and dropping the over-check from the non-dimensional variables, we obtain the following non-dimensional equations

$$\mathcal{R}\frac{Du}{Dt}-(f_0+\mathcal{R}\beta y+\cdots)v+\left(1+\mathcal{R}\frac{\beta^\dagger y}{f_0}+\cdots\right)\frac{\partial\Phi}{\partial x}=\mathcal{R}X, \quad (4.2.17)$$

$$\mathcal{R}\frac{Dv}{Dt}+(f_0+\mathcal{R}\beta y+\cdots)u+\frac{\partial\Phi}{\partial y}=\mathcal{R}Y, \quad (4.2.18)$$

$$\left(1+\mathcal{R}\frac{\beta^\dagger y}{f_0}+\cdots\right)\left\{\frac{\partial u}{\partial x}+\frac{\partial}{\partial y}\left[v\left(1-\mathcal{R}\frac{\beta^\dagger y}{f_0}+\cdots\right)\right]\right\}+\frac{1}{\rho_{\rm B}}\frac{\partial}{\partial z}(\rho_{\rm B}w)=0, \quad (4.2.19)$$

$$\mathcal{R}\frac{D}{Dt}\left(\frac{\partial\Phi}{\partial z}\right)+\mathcal{B}N^2w=\mathcal{R}J, \quad (4.2.20)$$

where

$$\frac{D}{Dt}\equiv\frac{\partial}{\partial t}+u\left(1+\mathcal{R}\frac{\beta^\dagger y}{f_0}+\cdots\right)\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}+w\frac{\partial}{\partial z}, \quad (4.2.21)$$

and the non-dimensional parameters $\mathcal{R}$, $\mathcal{B}$, f_0, β, and $\beta^\dagger$ are defined as follows, respectively,

$$\mathcal{R}=\frac{V}{2\Omega L},\quad \mathcal{B}=\left(\frac{N_0D}{2\Omega L}\right)^2,\quad f_0=\sin\phi_0,\quad \beta=\frac{2\Omega L^2\cos\phi_0}{a_{\rm E}V},\quad \beta^\dagger=\beta\tan^2\phi_0. \quad (4.2.22)$$

Note that $\mathcal{R}$ and $\mathcal{B}$ are usually called the Rossby number and Burger number, respectively. For the planetary-scale motion in the stratosphere, we choose specific magnitudes for the scalings as follows,

$$V=30\,{\rm ms}^{-1},\quad L=1.5\times10^6\,{\rm m},\quad D=10^4\,{\rm m},\quad N_0^2=5\times10^{-4}{\rm s}^{-2}. \quad (4.2.23)$$

In addition, $H=7\times10^3\,{\rm m}$, $\phi_0=45°$, $a_{\rm E}=6.371\times10^6\,{\rm m}$, $g=9.8\,{\rm ms}^{-2}$, and $2\Omega=1.4584\times10^{-4}{\rm s}^{-1}$. Substituting these values into Eq.(4.2.22), we see that

$$\mathcal{R}=0.1371,\quad \mathcal{B}=1.0448,\quad f_0=0.7071,\quad \beta=1.2140,\quad \beta^\dagger=1.2140. \quad (4.2.24)$$

Eq.(4.2.24) shows that $\mathcal{R} \ll 1$ while other dimensionless parameters, i.e. $\mathcal{B}$, f_0, β and $\beta^\dagger$, are all $O(1)$ in our model. Therefore we can expand all the variables in asymptotic series in $\mathcal{R}$, namely,

$$\left.\begin{array}{l} u(x,y,z,t,\mathcal{R}) = \mathcal{R}^0 u_g(x,y,z,t) + \mathcal{R}^1 u_a(x,y,z,t) + \cdots \\ v(x,y,z,t,\mathcal{R}) = \mathcal{R}^0 v_g(x,y,z,t) + \mathcal{R}^1 v_a(x,y,z,t) + \cdots \\ \vdots \end{array}\right\} \tag{4.2.25}$$

where u_g, u_a, etc., are independent of $\mathcal{R}$. Substituting these expansions into Eqs.(4.2.17)–(4.2.20) and equating coefficients of $\mathcal{R}^0$, we obtain the leading order equations

$$v_g = \frac{1}{f_0}\frac{\partial \Phi_g}{\partial x} = \frac{\partial \psi_g}{\partial x}, \qquad u_g = -\frac{1}{f_0}\frac{\partial \Phi_g}{\partial y} = -\frac{\partial \psi_g}{\partial y}, \qquad w_g = 0, \tag{4.2.26}$$

where $\psi_g \equiv \Phi_g / f_0$ is the geostrophic stream function. Eq.(4.2.26) shows that the leading-order fields are in geostrophic balance and do not contain enough information to complete the dynamical determination of the motion in the sense that Φ_g remains undetermined. In order to obtain the equation governing the evolution of Φ_g, we must resort to the first order equations in $\mathcal{R}$. Substituting Eq.(4.2.25) into Eqs.(4.2.17)–(4.2.20) and equating coefficients of $\mathcal{R}^1$, we obtain

$$\frac{D_g u_g}{Dt} - (\beta - \beta^\dagger) y v_g - X_g = -\frac{\partial \Phi_a}{\partial x} + f_0 v_a, \tag{4.2.27}$$

$$\frac{D_g v_g}{Dt} + \beta y u_g - Y_g = -\frac{\partial \Phi_a}{\partial y} - f_0 u_a, \tag{4.2.28}$$

$$\left[\frac{D_g}{Dt}\left(\frac{\partial \Phi_g}{\partial z}\right) - J_g\right] = -\mathcal{B} N^2 w_a, \tag{4.2.29}$$

$$\frac{\beta^\dagger}{f_0}\frac{\partial}{\partial y}(y v_g) = \frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} + \frac{1}{\rho_B}\frac{\partial}{\partial z}(\rho_B w_a), \tag{4.2.30}$$

where

$$\frac{D_g}{Dt} = \frac{\partial}{\partial t} + u_g \frac{\partial}{\partial x} + v_g \frac{\partial}{\partial y}. \tag{4.2.31}$$

Now taking $\partial(4.2.28)/\partial x - \partial(4.2.27)/\partial y$, yields

$$\frac{D_g}{Dt}\left(\beta y + \frac{\partial v_g}{\partial x} - \frac{\partial u_g}{\partial y}\right) - \beta^\dagger \frac{\partial}{\partial y}(y v_g) - \left(\frac{\partial Y_g}{\partial x} - \frac{\partial X_g}{\partial y}\right) = -f_0\left(\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y}\right). \tag{4.2.32}$$

Combining Eq.(4.2.32) with Eqs.(4.2.30) and (4.2.29), we obtain the QGPV equation as follows,

$$\frac{D_g q_g}{Dt} = Z_g, \tag{4.2.33}$$

where

$$q_g = \beta y + \frac{\partial^2 \psi_g}{\partial x^2} + \frac{\partial^2 \psi_g}{\partial y^2} + \frac{1}{\rho_B}\frac{\partial}{\partial z}\left(\epsilon \rho_B \frac{\partial \psi_g}{\partial z}\right), \tag{4.2.34}$$

$$Z_g = \frac{\partial Y_g}{\partial x} - \frac{\partial X_g}{\partial y} + \frac{1}{f_0 \rho_B}\frac{\partial}{\partial z}(\epsilon \rho_B J_g), \tag{4.2.35}$$

with

$$\epsilon(z) = \frac{f_0^2}{BN^2(z)}. \tag{4.2.36}$$

Note that q_g is the QGPV and Z_g represents the effect of nonconservative processes. In the QGPV equation, the only effect of the spherical shape of the earth on the leading-order fields of quasi-geostrophic motions is the term βy. This β effect obviously comes from the approximation of using the Cartesian geometry but letting the Coriolis parameter vary linearly with y. Such an approximation is referred to as the β-plane approximation. Application of the β-plane approximation is limited to flow fields for which the meridional extent of the motion is small compared to the radius of the earth. Note that the QGPV equation with the appropriate boundary conditions forms the complete dynamical basis for the calculation of the leading-order fields of motion. In principle, higher-order fields can be calculated in a similar fashion (e.g., Zeytounian, 1990). In this chapter we shall concentrate exclusively on quasi-geostrophic motions and need not invoke the first-order ageostrophic motions.

4.3 Forced Rossby waves on the β-plane

The most important large-scale wave disturbances in the middle atmosphere are identified as forced Rossby waves that have their origin in the troposphere. In this section, we shall discuss linear Rossby waves forced from below within the framework of quasi-geostrophic theory in the β-plane geometry.

4.3.1 Linearised equations for Rossby waves

Suppose that small-amplitude disturbances are superimposed upon a basic zonal flow $[\overline{U}(y, z), 0, 0]$. Following standard practice, we use an overbar to denote a zonal mean and a prime to represent a departure from the zonal mean. Then the equation for the disturbance QGPV can be written as

$$\frac{\overline{D} q_g'}{Dt} + v_g' \overline{q}_y = Z_g' + O(a^2), \tag{4.3.1}$$

where

$$q'_{\rm g} \equiv \frac{\partial^2 \psi'_{\rm g}}{\partial x^2} + \frac{\partial^2 \psi'_{\rm g}}{\partial y^2} + \frac{1}{\rho_{\rm B}}\frac{\partial}{\partial z}\left(\epsilon\rho_{\rm B}\frac{\partial \psi'_{\rm g}}{\partial z}\right), \tag{4.3.2}$$

$$\overline{q}_y \equiv \frac{\partial \overline{q}_{\rm g}}{\partial y} = \beta - \frac{\partial^2 \overline{U}}{\partial y^2} - \frac{1}{\rho_{\rm B}}\frac{\partial}{\partial z}\left(\epsilon\rho_{\rm B}\frac{\partial \overline{U}}{\partial z}\right), \tag{4.3.3}$$

$$Z'_{\rm g} = \frac{\partial Y'_{\rm g}}{\partial x} - \frac{\partial X'_{\rm g}}{\partial y} + \frac{1}{f_0\rho_{\rm B}}\frac{\partial}{\partial z}\left(\epsilon\rho_{\rm B} J'_{\rm g}\right), \tag{4.3.4}$$

and

$$\frac{\overline{D}}{Dt} \equiv \frac{\partial}{\partial t} + \overline{U}\frac{\partial}{\partial x}. \tag{4.3.5}$$

In addition, a is a dimensionless amplitude parameter that is assumed much less than 1, namely

$$a \ll 1. \tag{4.3.6}$$

Under such circumstances, the nonlinear effect represented by the $O(a^2)$ terms in Eq.(4.3.1) can be omitted.

$X'_{\rm g}$, $Y'_{\rm g}$, and $J'_{\rm g}$ in Eq.(4.3.4) are unspecified forcing terms in the linearised zonal momentum, meridional momentum and thermodynamic energy equations, respectively. In general, these arbitrary forcing terms may represent either wave dissipation or excitation, depending on whether their correlations with the corresponding disturbance motions are positive or negative (Andrews and McIntyre, 1976a). In what follows, only wave dissipation is concerned. Thus, in our model $X'_{\rm g}$, $Y'_{\rm g}$, and $J'_{\rm g}$ are assumed to be correlated negatively with $u'_{\rm g}$, $v'_{\rm g}$, and $\partial\Phi'_{\rm g}/\partial z$, respectively. In the absence of other effects, they tend to bring the corresponding disturbance motions to zero. This, however, does not imply automatically that $Z'_{\rm g}$ defined by Eq.(4.3.4) will tend to bring $q'_{\rm g}$ to zero; see the discussion in §4.4.3.

Owing to radiative and photochemical effects, temperature perturbations in the stratosphere and mesosphere tend to relax towards a state of radiative equilibrium. The dynamical consequences of the relaxing are usually considered approximately by parameterising the diabatic heating rate as proportional to the perturbation temperature, i.e.,

$$J'_{\rm g} = -\mu\alpha_{\rm N}\frac{\partial \Phi'_{\rm g}}{\partial z}, \tag{4.3.7}$$

where $\alpha_{\rm N}$ is the non-dimensional Newtonian cooling coefficient and μ a non-dimensional parameter that characterises the dissipation strength. We shall assume that the dissipation processes are weak in the sense that μ is a small number, i.e.,

$$\mu \ll 1. \tag{4.3.8}$$

The mechanical dissipation terms in the atmosphere are poorly understood. For simplicity, $X'_{\rm g}$ and $Y'_{\rm g}$ are parameterised by the Rayleigh friction approximation, that is

$$X'_{\rm g} = -\mu\alpha_{\rm R} u'_{\rm g}, \quad Y'_{\rm g} = -\mu\alpha_{\rm R} v'_{\rm g}, \tag{4.3.9}$$

where $\alpha_{\rm R}$ is the non-dimensional Rayleigh friction coefficient. Though Rayleigh friction is, generally speaking, unrealistic in the real atmosphere, it may be used to estimate some mechanical dissipation such as that caused by the breaking of gravity waves (Holton and Wehrbein, 1980; Chen and Robinson, 1992).

To retain the maximum simplicity we shall assume that both $\alpha_{\rm N}$ and $\alpha_{\rm R}$ are constants, although they are, in general, functions of all the independent variables.

We now specialise the motion by assuming that the flow is confined to a channel bounded by two rigid vertical walls at $y = -\Delta y/2,\ \Delta y/2$. Thus, the lateral boundary conditions are

$$v'_{\rm g} = \frac{\partial \psi'_{\rm g}}{\partial x} = 0, \qquad \text{at } y = -\Delta y/2 \text{ and } \Delta y/2. \tag{4.3.10}$$

To model the stationary Rossby waves that are forced in the troposphere and propagate into the stratosphere and mesosphere, we specify the disturbance geopotential at a given log-pressure level $z = z_0$ as

$$\psi'_{\rm g}(x, y, z_0, t) = \mathrm{Re}\left\{a\cos(\ell y)\exp(ikx)\right\}, \tag{4.3.11}$$

where

$$k = \frac{sL}{a_{\rm E}\cos\phi_0}, \qquad \ell = \frac{\pi L}{\Delta y}, \tag{4.3.12}$$

with s being the zonal wavenumber ($s = 1, 2, 3, \cdots$). In this chapter, we choose $\ell = 1.0595$, which corresponds to a channel centred at $\phi_0 = 45°$ latitude and bounded respectively by walls at 25° and 65° latitudes.

As for the upper boundary condition, we require that the density of wave-energy remains finite at great heights. This corresponds to

$$\rho_{\rm B}|\overline{\psi'^2_{\rm g}}| = \text{bounded}, \qquad \text{as } z \to \infty. \tag{4.3.13}$$

We now look for steady, stationary solutions to Eq.(4.3.1) of the form

$$\psi'_{\rm g} = \mathrm{Re}\left\{\hat{\psi}(y, z)\exp(ikx)\right\}. \tag{4.3.14}$$

Substituting Eq.(4.3.14) into Eq.(4.3.1) and neglecting the nonlinear effect, we obtain the following second-order partial differential equation for $\widehat{\psi}(y,z)$,

$$\frac{1}{\rho_{\text{B}}}\frac{\partial}{\partial z}\left(\epsilon\rho_{\text{B}}\frac{\partial\widehat{\psi}}{\partial z}\right)+\frac{\left(ik\overline{U}+\mu\alpha_{\text{R}}\right)}{\left(ik\overline{U}+\mu\alpha_{\text{N}}\right)}\left(\frac{\partial^2\widehat{\psi}}{\partial y^2}-k^2\widehat{\psi}\right)+\frac{ik\overline{q}_y}{\left(ik\overline{U}+\mu\alpha_{\text{N}}\right)}\widehat{\psi}=0. \tag{4.3.15}$$

For realistic flows $\overline{U}(y,z)$, Eq.(4.3.15) with suitable boundary conditions can only be solved by numerical methods (e.g., Matsuno, 1970). To gain some insight into the nature of the solutions, in what follows we assume that $\overline{U}=\overline{U}(z)$ and

$$\widehat{\psi}(z,y)=\widehat{\Psi}(z)\cos(\ell y). \tag{4.3.16}$$

Then, Eq.(4.3.15) reduces

$$\frac{d^2\widehat{\Psi}}{dz^2}+\left(\frac{d\ln\epsilon}{dz}-2n_H\right)\frac{d\widehat{\Psi}}{dz}+\left(C+n_H{}^2\right)\widehat{\Psi}=0, \tag{4.3.17}$$

where

$$n_H=\frac{D}{2H}, \tag{4.3.18}$$

and

$$C=\frac{(k^2+\ell^2)\left[k(u_{\text{c}}-\overline{U})+i\mu\alpha_{\text{R}}\right]}{\epsilon(k\overline{U}-i\mu\alpha_{\text{N}})}-n_H{}^2=C_{\text{r}}+iC_{\text{i}}, \tag{4.3.19}$$

with

$$u_{\text{c}}=\frac{\overline{q}_y}{k^2+\ell^2}, \tag{4.3.20}$$

$$C_{\text{r}}=\frac{(k^2+\ell^2)\left[k^2\overline{U}(u_{\text{c}}-\overline{U})-\mu^2\alpha_{\text{N}}\alpha_{\text{R}}\right]}{\epsilon(k^2\overline{U}^2+\mu^2\alpha_{\text{N}}{}^2)}-n_H{}^2, \tag{4.3.21}$$

$$C_{\text{i}}=\frac{\mu k(k^2+\ell^2)\left[\alpha_{\text{N}}(u_{\text{c}}-\overline{U})+\alpha_{\text{R}}\overline{U}\right]}{\epsilon(k^2\overline{U}^2+\mu^2\alpha_{\text{N}}{}^2)}. \tag{4.3.22}$$

Following Charney and Drazin (1961), we shall call u_{c} the *Rossby critical velocity.*[1] The corresponding boundary conditions are

$$\widehat{\Psi}=a, \qquad \text{at } z=z_0, \tag{4.3.23}$$

$$\rho_{\text{B}}|\widehat{\Psi}|^2=\text{bounded}, \qquad \text{as } z\to\infty. \tag{4.3.24}$$

[1] In Charney and Drazin's original paper, u_{c} defined by Eq.(4.3.20) is called the Rossby critical velocity. In some standard text books of dynamic meteorology (e.g., Holton, 1975; Andrews *et al.* 1987), however, this terminology usually refers to another quantity U_{c} defined by Eq.(4.3.33), which is called the *modified Rossby critical velocity* in Charney and Drazin's paper. Note that $U_{\text{c}}<u_{\text{c}}$, provided that $\overline{q}_y>0$.

4.3.2 Wave solutions for constant $\overline{U}$ and N^2

Let us consider first the simplest situation in which both $\overline{U}$ and N^2 are taken to be constant. Under such circumstances, the solution of Eq.(4.3.17) satisfying the boundary conditions (4.3.23) and (4.3.24) can be written as

$$\widehat{\Psi}(z) = a \exp\left[(im_\dagger - n_\dagger)z\right], \tag{4.3.25}$$

where $m_\dagger$, $n_\dagger$ are real numbers. Substituting Eq.(4.3.25) into Eq.(4.3.17) yields (note that $d\ln\epsilon/dz = 0$ in this case)

$$m_\dagger{}^2 - (n_\dagger + n_H)^2 = C_\mathrm{r}, \tag{4.3.26}$$

$$2m_\dagger\,(n_\dagger + n_H) = C_\mathrm{i}. \tag{4.3.27}$$

The upper boundary condition (4.3.24) requires $(n_\dagger + n_H) \geq 0$. With this constraint $m_\dagger$ and $n_\dagger$ can be solved from Eqs.(4.3.26) and (4.3.27) as

$$m_\dagger = \mathrm{sgn}(C_\mathrm{i})\sqrt{(|C| + C_\mathrm{r})/2}, \tag{4.3.28}$$

$$n_\dagger = -n_H + \sqrt{(|C| - C_\mathrm{r})/2}. \tag{4.3.29}$$

It is demonstrated in Appendix 4A that the vertical wave-energy flux due to the dissipating Rossby waves, whose solutions are described by Eq.(4.3.25) with Eqs.(4.3.28) and (4.3.29), is always positive. This is in accordance with the assumption that the waves are forced from below.

Now letting $\mu \to 0$, we see that

$$\lim_{\mu\to 0} m_\dagger = m = \begin{cases} \sqrt{C_{\mathrm{r}0}}, & \text{if } 0 < \overline{U} < U_\mathrm{c} \text{ (i.e., } C_{\mathrm{r}0} > 0), \\ 0, & \text{if otherwise;} \end{cases} \tag{4.3.30}$$

$$\lim_{\mu\to 0} n_\dagger = n = \begin{cases} -n_H, & \text{if } 0 < \overline{U} < U_\mathrm{c}, \\ -n_H + \sqrt{-C_{\mathrm{r}0}}, & \text{if otherwise;} \end{cases} \tag{4.3.31}$$

where

$$C_{\mathrm{r}0} = \frac{(k^2 + \ell^2)(u_\mathrm{c} - \overline{U})}{\epsilon\overline{U}} - n_H{}^2, \tag{4.3.32}$$

$$U_\mathrm{c} = \frac{u_\mathrm{c}(k^2 + \ell^2)}{(k^2 + \ell^2 + \epsilon n_H{}^2)} = \frac{\overline{q}_y}{(k^2 + \ell^2 + \epsilon n_H{}^2)}. \tag{4.3.33}$$

U_c is referred to as the *modified Rossby critical velocity* (Charney and Drazin, 1961). m is chosen to be positive to guarantee an upward wave-energy flux, as is demonstrated explicitly in Appendix 4A.

Eqs.(4.3.30) and (4.3.31) show that $m \neq 0$ (and $n = -n_H$) if and only if the basic zonal flow $\overline{U}$ satisfies the following equation

$$0 < \overline{U} < U_c. \tag{4.3.34}$$

This criterion was first derived by Charney and Drazin (1961) to illustrate the fact that there exists a "window" (to be referred to as the CD window), $0 < \overline{U} < U_c$, within which the vertical structure of a conservative, stationary Rossby wave is of oscillating shape ($m \neq 0, n = -n_H$) and the wave is free to propagate vertically. It should be noted that, owing to the exponential decrease of density with altitude, the amplitude of disturbance geopotential for the linear, conservative, stationary Rossby wave grows with altitude exponentially within the CD window ($n = -n_H < 0$). At some height, therefore, the nonlinear terms that have been neglected will become important, and the linear theory will break down.

Outside the CD window, where $m = 0$ and $n > -n_H$, the amplitude of the disturbance geopotential (and disturbance velocities as well) for the linear, conservative, stationary Rossby wave decays or grows with altitude; depending on whether $n > 0$ or $n < 0$. For positive $\overline{q}_y$, it can be easily seen from Eqs.(4.3.31) and (4.3.32) that there exists another "window",

$$0 < \overline{U} < u_c, \tag{4.3.35}$$

within which $n < 0$, and then the amplitude of the disturbance geopotential grows with height; otherwise the waves are vertically evanescent. For convenience, the domain defined by Eq.(4.3.35) will be referred to as the *geopotential-growth* (*GG*) *window* for the linear, conservative, stationary Rossby wave. Note that, within the GG window, disturbance geostrophic velocities u'_g and v'_g also grow with height, as can be seen from Eqs.(4.2.26) and (4.3.14). Since $u_c > U_c$, as indicated by Eq.(4.3.33), the GG window is somewhat wider than the CD window.

How m and n depend on $\overline{U}$ is shown in Fig.4.1. It is seen that both the CD window and the GG window become smaller as the zonal wavenumber s increases, in agreement with observational evidence that only planetary-scale quasi-stationary disturbances can be observed in the winter stratosphere (see Holton, 1975; Andrews *et al.*, 1987).

Variations of $m_\dagger$ and $n_\dagger$ with $\overline{U}$ for dissipating Rossby waves are shown in Fig.4.2 and Fig.4.3. Here $\mu = 0.0827$ is chosen. This value corresponds to a dimensional Newtonian

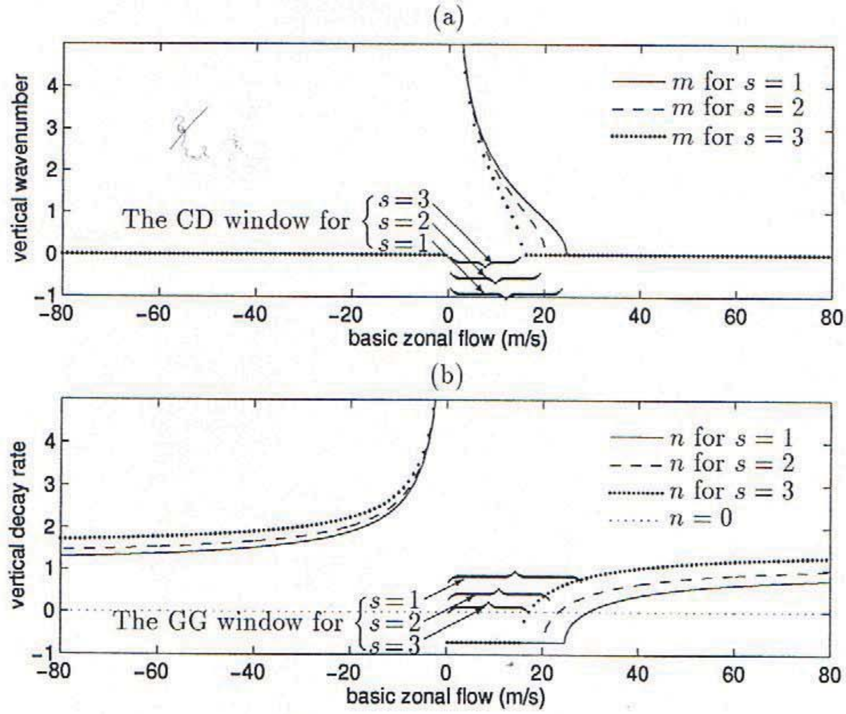


Figure 4.1: The dependence of (a) the vertical wavenumber m and (b) the vertical decay rate n on the basic zonal flow $\overline{U}$ ($\overline{U}$ is given in ms^{-1}) for linear, conservative, stationary Rossby waves with $s = 1, 2, 3$.

cooling rate of (7 days)$^{-1}$. In comparison with Fig.4.1 we see that dissipation produces noticeable amplitude attenuation and phase modification of the Rossby waves in some neighbourhood of $\overline{U} = 0$. Thus, the waves in the stratosphere and mesosphere should be strongly damped during the equinoctial seasons when the mean winds are weak westerlies (Dickinson, 1969b). The wave amplitudes are least attenuated for westerly winds with velocities near the Rossby critical velocity U_c. It should also be noted that the sign of $m_\dagger$ depends on the sign of C_i. If $\alpha_N = \alpha_R \neq 0$, Eq.(4.3.22) shows that C_i is positive, so that $m_\dagger$ is positive, which implies that the wave phase line tilts westward with increasing altitude. If, on the other hand, $\alpha_N \neq \alpha_R$, we see from Eqs.(4.3.22) and (4.3.28) that $m_\dagger$ is negative if

$$\alpha_N(u_c - \overline{U}) + \alpha_R \overline{U} < 0, \tag{4.3.36}$$

(Takahashi and Uryu, 1981). In particular, if the waves are dissipated by Newtonian cooling alone, $m_\dagger$ is negative if $\overline{U} > u_c$, as shown in Fig.4.3a. Note that a negative $m_\dagger$ corresponds to an eastward tilt of phase line with increasing altitude.

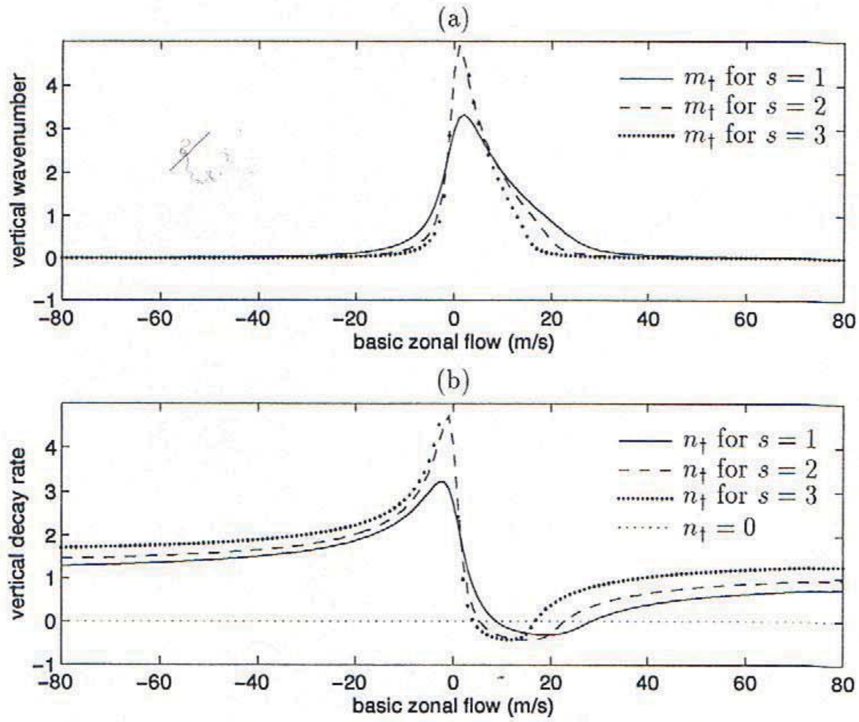


Figure 4.2: The dependence of (a) the vertical wavenumber $m_\dagger$ and (b) the vertical decay rate $n_\dagger$ on the basic zonal flow $\overline{U}$ for linear, dissipative, stationary Rossby waves with $s = 1, 2, 3$. $\alpha_{\rm N} = \alpha_{\rm R} = 0.5$.

We have seen that $m_\dagger$ and $n_\dagger$ can be solved exactly from Eqs.(4.3.26) and (4.3.27). For small μ ($\mu \ll 1$), it appears that approximations to $m_\dagger$ and $n_\dagger$ may be obtained by expanding the solutions of Eqs.(4.3.26) and (4.3.27) in perturbation series in powers of μ. To seek such asymptotic series we note first that, provided both $\overline{U}$ and $C_{\rm r0}$ are of order $O(1)$, $C_{\rm r}$ and $C_{\rm i}$ can be written as

$$C_{\rm r} = C_{\rm r0} + O(\mu^2), \tag{4.3.37}$$

$$C_{\rm i} = \mu C_{\rm i0} + O(\mu^3), \tag{4.3.38}$$

where $C_{\rm r0}$ is defined by Eq.(4.3.32) and $C_{\rm i0}$ by

$$C_{\rm i0} = \frac{(k^2 + \ell^2)\left[\alpha_{\rm N}(u_{\rm c} - \overline{U}) + \alpha_{\rm R}\overline{U}\right]}{\epsilon k \overline{U}^2}. \tag{4.3.39}$$

Then we assume that the solutions of Eqs.(4.3.26) and (4.3.27) can be written as

$$m_\dagger = m + \mu m_1 + \mu^2 m_2 + \cdots, \tag{4.3.40}$$

$$n_\dagger = n + \mu n_1 + \mu^2 n_2 + \cdots, \tag{4.3.41}$$

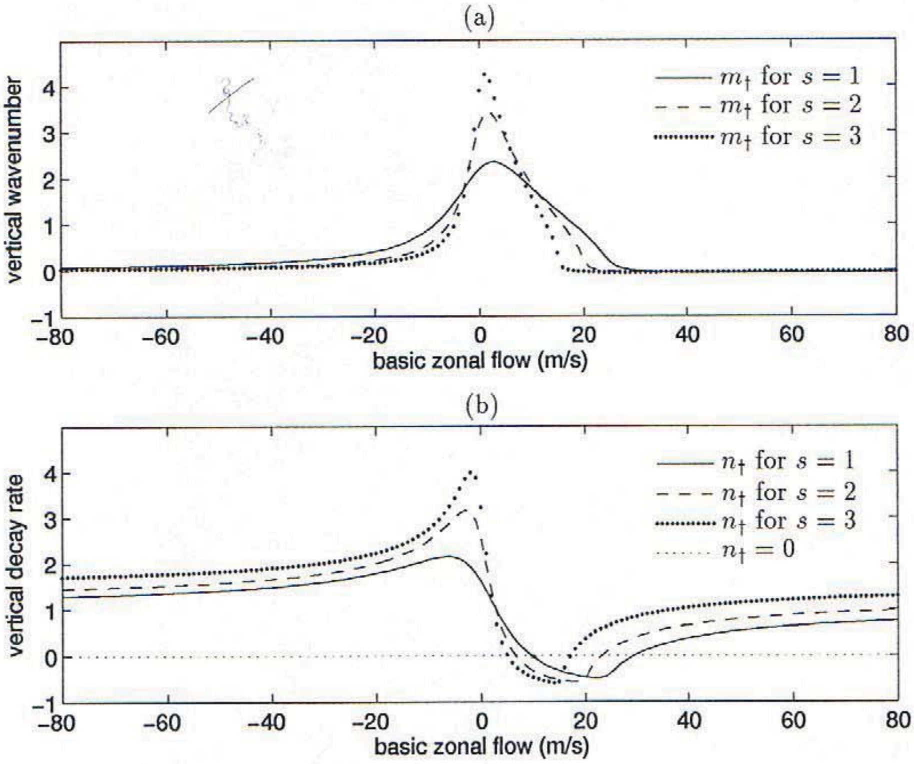


Figure 4.3: Same as Fig.4.2, but for $\alpha_N = 1, \alpha_R = 0$.

respectively. Substituting Eqs.(4.3.37), (4.3.38), (4.3.40) and (4.3.41) into Eqs.(4.3.26) and (4.3.27), and equating coefficients of equal powers of μ on both sides of equations, we obtain

$$m^2 - (n + n_H)^2 = C_{r0}, \tag{4.3.42}$$

$$m\,(n + n_H) = 0, \tag{4.3.43}$$

$$m_1 m - n_1\,(n + n_H) = 0, \tag{4.3.44}$$

$$2\,[mn_1 + m_1(n + n_H)] = C_{i0}. \tag{4.3.45}$$

Eqs.(4.3.30) and (4.3.31) can be recovered immediately from Eqs.(4.3.42) and (4.3.43). When $\mu \neq 0$, the dissipation terms produce a modification factor $\exp\,[\mu(im_1 - n_1)z]$, where m_1 and n_1 can be obtained from Eqs.(4.3.30), (4.3.31), (4.3.44) and (4.3.45) as follows,

$$m_1 = \begin{cases} 0, & \text{if } 0 < \overline{U} < U_c, \\ C_{i0}/[2(n + n_H)], & \text{if otherwise;} \end{cases} \tag{4.3.46}$$

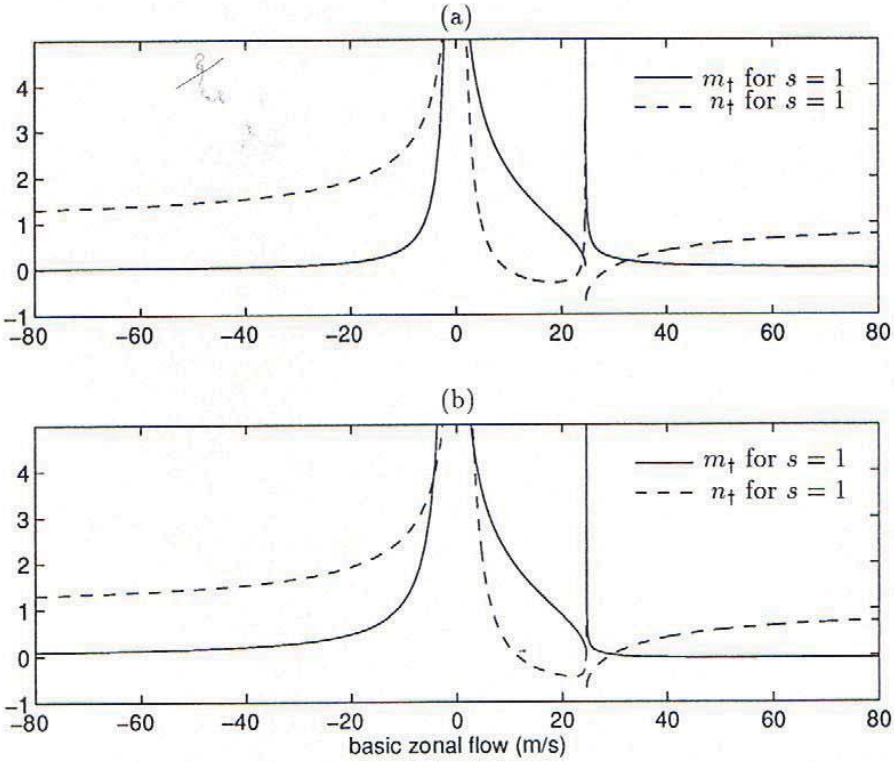


Figure 4.4: Dependences of the vertical wavenumber $m_\dagger$ (solid line) and decay rate $n_\dagger$ (dashed line) on the basic zonal flow $\overline{U}$ for $s = 1$. $m_\dagger$ and $n_\dagger$, calculated respectively from Eqs.(4.3.40) and (4.3.41) to the first order in μ. (a) $\alpha_N = \alpha_R = 0.5$. (b) $\alpha_N = 1, \alpha_R = 0$.

$$n_1 = \begin{cases} C_{i0}/(2m), & \text{if } 0 < \overline{U} < U_c, \\ 0, & \text{if otherwise.} \end{cases} \tag{4.3.47}$$

Substituting Eqs.(4.3.31) and (4.3.47) into Eq.(4.3.41) shows that the extra damping of waves due to weak dissipation is important within the CD window, where $n_\dagger = n + \mu n_1 + O(\mu^2)$, and is negligible outside the CD window, where $n_\dagger = n + O(\mu^2)$. In contrast, the modification of wave phase due to weak dissipation is important outside the CD window, where $m = 0$ but $m_1 \neq 0$, and is negligible within the CD window, where $m \neq 0$ but $m_1 = 0$. Because $(n + n_H) > 0$ outside the CD window, m_1 given by Eq.(4.3.46) has the same sign as C_{i0}. Therefore it is straightforward to show that m_1 is negative and m zero when $\overline{U}$ satisfies Eq.(4.3.36). This implies that $m_\dagger = \mu m_1 + O(\mu^2)$ is negative, in accordance with Eq.(4.3.28). In addition, $n_1 \geq 0$ is required by the upper boundary condition (4.3.24).

It should be emphasized that the asymptotic expansions of Eqs.(4.3.37), (4.3.38), (4.3.40) and (4.3.41) are based on the assumption that both C_{r0} and C_{i0}, given respectively by Eq.(4.3.32) and Eq.(4.3.39), are of order unity. Since $C_{r0} \to \infty$ and $C_{i0} \to \infty$ as $\overline{U} \to 0$, there must be some neighbourhood of $\overline{U} = 0$ such that the expansion based on Eqs.(4.3.37) and (4.3.38) fails to be uniformly valid (as $\overline{U}$ varies). On the other hand, when $\overline{U} \to U_c$, Eq.(4.3.32) shows that $C_{r0} \to 0$. This also violates the assumption mentioned above. Therefore there must also be some neighbourhood of $\overline{U} = U_c$ on which the expansion based on Eq.(4.3.37) fails to be uniformly valid.

The asymptotic approximations for $m_\dagger$ and $n_\dagger$, given respectively by Eqs.(4.3.40) and (4.3.41) to the first order in μ for zonal wavenumber $s = 1$, are shown in Fig.4.4. In comparison with Fig.4.2 and Fig.4.3 we see that $m_\dagger$ and $n_\dagger$ are appropriately represented by these approximations except on some small neighbourhood of $\overline{U} = 0$ and $\overline{U} = U_c$, in agreement with the above argument.

4.3.3 Wave solutions for $\overline{U}$ and N^2 varying with z

In §4.3.2 the characteristics of forced Rossby waves in an atmosphere with constant $\overline{U}$ and N^2 were discussed. However, the restriction to constant $\overline{U}$ and N^2 is rather unsatisfactory in the real atmosphere. To gain further insight into the structure of forced Rossby waves in the real atmosphere, we now consider the effects of vertical variations of $\overline{U}$ and N^2.

Consider a channel model centred at 45°N with a meridional extent equal to 40° latitude (i.e., 25°N—65°N). The basic zonal winds $\overline{U}(z)$ and temperature $T_B(z)$ in January for this model are plotted in Fig.4.5a and Fig.4.5b, respectively. Data for the diagrams are obtained from the 1986 COSPAR International Reference Atmosphere compilation (see Fleming *et al.*, 1990), as shown in Fig.1.2 and Fig.1.3. These data are averaged across the channel to represent the winter mean conditions in middle latitudes. Fig.4.5c shows $N^2(z)$ calculated from Eq.(4.2.10) using the temperature profile given in Fig.4.5b.

Now Eq.(4.3.17) can be solved by numerical methods. Following Matsuno (1970), we impose a radiation condition at the top of the model ($z = z_T = 85$ km) by assuming that $\overline{U}$ and N^2 are independent of z above z_T. This allows analytical solutions to be found above z_T, and the solutions with upward wave-energy flux are chosen to supply the upper boundary condition required by the numerical method of solution.

Fig.4.6 shows the vertical structures of the amplitude $|\widehat{\Psi}(z)|$ for waves with $s = 1$–3.

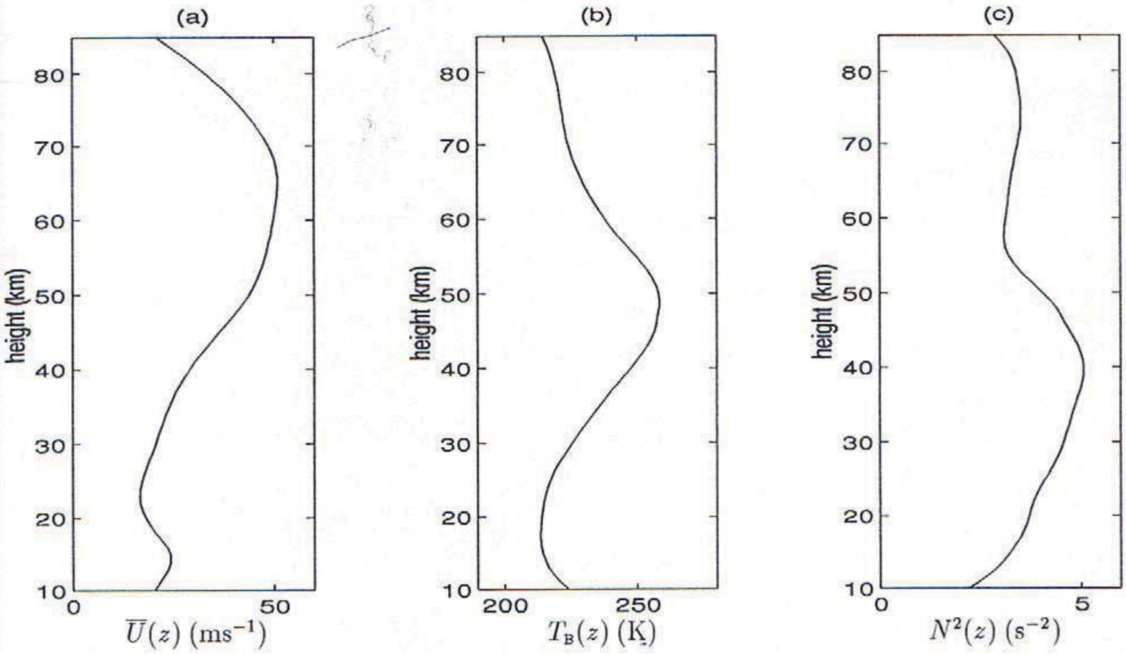


Figure 4.5: Vertical structures of (a) the basic zonal flow $\overline{U}(z)$, (b) the basic state temperature $T_B(z)$, and (c) the buoyancy frequency squared $N^2(z)$ in the winter Northern Hemisphere. $\overline{U}(z)$ and $T_B(z)$ are calculated by using data from 1986 COSPAR International Reference Atmosphere compilation (see Fleming *et al.*, 1990). $N^2(z)$ is calculated from Eq.(4.2.10) using the temperature profile given in Fig.4.5b.

Waves with $s \geq 4$ were also calculated but omitted from this discussion, because their amplitudes are much smaller than that of wave 3. For conservative waves ($\alpha_N = \alpha_R = 0$, Fig.4.6a), the amplitude of wave 1 ($s = 1$) increases with height rapidly in the stratosphere, reaching its maximum near the stratopause. In contrast, the amplitude of wave 2 in the same figure is approximately constant in the stratosphere, and that of wave 3 decreases with height monotonically. When wave dissipation is included in the model, wave 1 is heavily attenuated, while the damping of waves 2 and 3 due to dissipative effects is insignificant (Fig.4.6b,c).

An investigation into the vertical structure of the wave phase indicates that the behaviour of the Rossby wave largely depends on the dissipation mechanisms. For $\alpha_N = \alpha_R = 0.5$, all phase lines tilt westward (phase angle increases with height), as shown in in Fig.4.7a. For $\alpha_N = 1$, $\alpha_R = 0$, on the other hand, eastward inclination (phase angle decreases with height) occurs in the mesosphere for all wave components and in

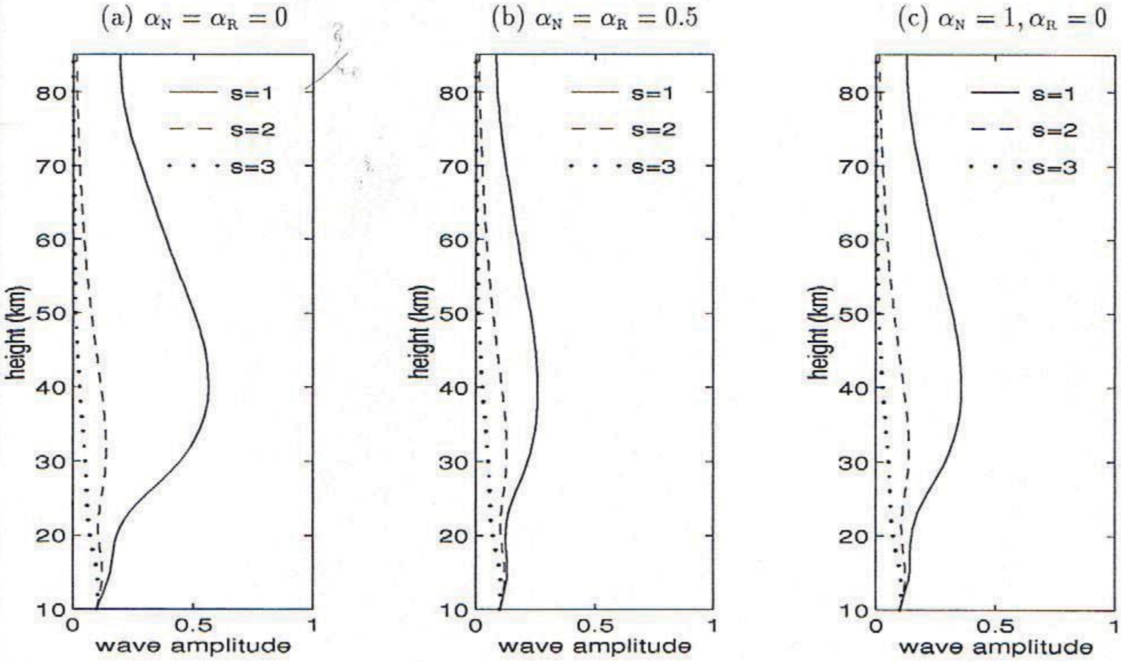


Figure 4.6: Vertical structures of wave amplitude $|\widehat{\Psi}(z)|$ for waves with $s = 1$–3.

the stratosphere for waves 2 and 3 as well, as shown in Fig.4.7b.

For constant $\overline{U}$ and N^2, it was shown in §4.3.2 that the disturbance geopotential associated with conservative, stationary Rossby waves grows with height within the GG window (i.e., $0 < \overline{U} < u_c$), and decays with height otherwise. In addition, conservative, stationary Rossby waves are free to propagate vertically with phase line tilting westward within the CD window (i.e., $0 < \overline{U} < U_c$). Outside the CD window, these waves are not of oscillating shape, so that there is no phase tilt with height. For dissipating Rossby waves, the extra damping of wave amplitude due to weak dissipation is significant within the CD window and negligible otherwise, while the modification of wave phase due to weak dissipation is negligible within the CD window and significant otherwise. In particular, if Rossby waves are dissipated by the Newtonian cooling alone, the phase line tilts eastward. To see whether these results are still valid when $\overline{U}$ and N^2 vary with height, we now plot the vertical structures of $(\overline{U} - U_c)$ and $(\overline{U} - u_c)$ in Fig.4.8. In comparison with Fig.4.6 and Fig.4.7 we see that the amplitudes for conservative waves grow with height when $\overline{U} < u_c$ and decay otherwise. Moreover, the damping of waves due to weak dissipation

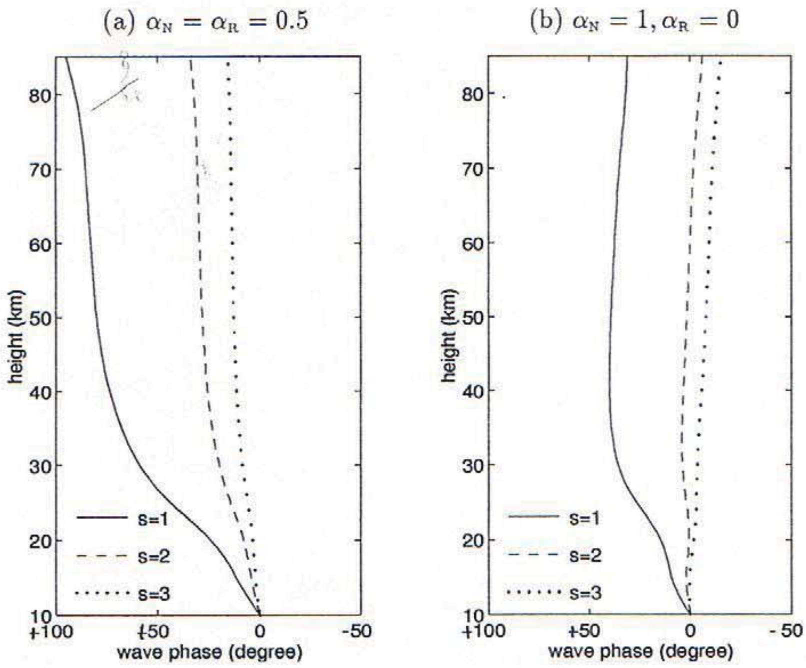
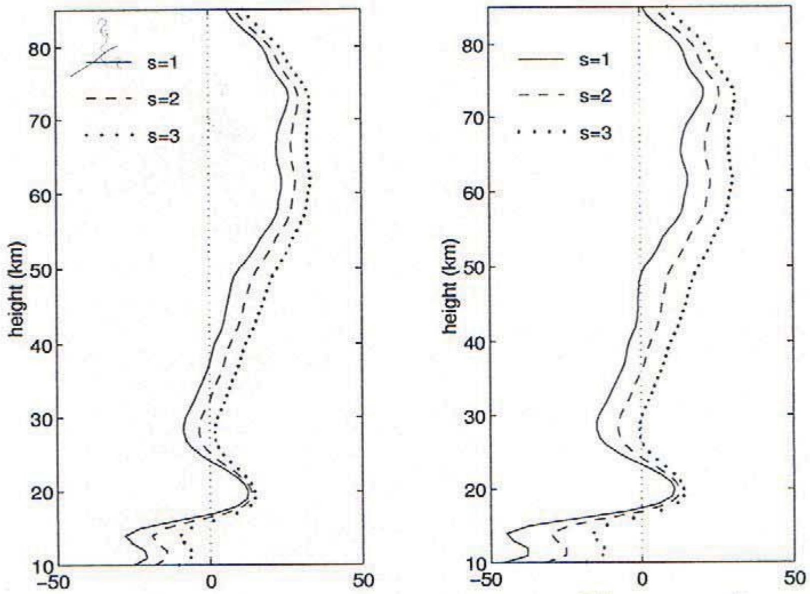


Figure 4.7: Vertical structures of wave phase for waves with $s = 1$—3.

is significant only when $\overline{U} < U_c$. In addition, when waves are dissipated by Newtonian cooling alone, their phase lines tilt westward if $\overline{U} < u_c$ and eastward if otherwise (an exception can be seen in the lower stratosphere where the phase line of wave 1 tilts westward, though the corresponding u_c is less than $\overline{U}$). These phenomena indicate that the conclusions obtained in §4.3.2 for constant $\overline{U}$ and N^2 are loosely applicable for the case in which $\overline{U}$ and N^2 vary with height.

4.4 The quasi-geostrophic EP flux

Many phenomena observed in the stratosphere involve the interaction of the mean flow with planetary-scale Rossby waves. The discussion of Section 4.3 illustrates explicitly how the mean flow can modify the disturbances in the model atmosphere. The question of how the Rossby waves modify the mean flow is investigated in this and the next section. The focus of this section is the quasi-geostrophic EP flux, the divergence of which represents the sole forcing of the mean state by large-scale disturbances (Edmon *et al.*, 1980).



(a) $\overline{U} - U_c\ (\mathrm{m\,s^{-1}})$ (b) $\overline{U} - u_c\ (\mathrm{m\,s^{-1}})$

Figure 4.8: (a) Vertical structure of the difference between $\overline{U}(z)$ and $u_c(z)$, where $\overline{U}(z)$ is the basic zonal wind and $u_c(z)$ the Rossby critical velocity defined by Eq.(4.3.20). (b) Vertical structure of the difference between $\overline{U}(z)$ and $U_c(z)$, where $U_c(z)$ is the modified Rossby critical velocity defined by Eq.(4.3.33).

4.4.1 The TEM equations and the generalized Eliassen–Palm theorem

The dynamical role of eddies in maintaining the zonal mean motion can be elucidated by utilizing the so-called *transformed Eulerian-mean* (TEM) equations introduced by Andrews and McIntyre (1976a, 1978a). In the TEM description, a new mean meridional circulation, sometimes known as the *residual circulation*, is defined as the sum of the conventional mean meridional circulation and some additional wave disturbance quadratic terms, in order that the wave effects on the zonal-mean flow can be separated from the mean properties. In the quasi-geostrophic case, the residual circulation $(0, \overline{v}^*_a, \overline{w}^*_a)$ is defined by

$$\overline{v}^*_a = \overline{v}_a - \frac{1}{f_0\rho_B}\frac{\partial}{\partial z}\left(\epsilon\rho_B \overline{v'_g \frac{\partial\Phi'_g}{\partial z}}\right), \qquad \overline{w}^*_a = \overline{w}_a + \frac{\epsilon}{f_0{}^2}\frac{\partial}{\partial y}\left(\overline{v'_g\frac{\partial \Phi'_g}{\partial z}}\right). \tag{4.4.1}$$

Now applying the conventional Eulerian-mean operator to Eqs.(4.2.27), (4.2.29), and (4.2.30), and then substituting Eq.(4.4.1) into the resulting mean equations to eliminate

$\overline{v}_{\mathrm{a}}$ and $\overline{w}_{\mathrm{a}}$, we obtain the following TEM equations:

$$\frac{\partial \overline{u}_{\mathrm{g}}}{\partial t} - f_0 \overline{v}_{\mathrm{a}}^* = \frac{1}{\rho_{\mathrm{B}}} \nabla \cdot \mathbf{S}_{\mathrm{g}} + \overline{X}_{\mathrm{g}}, \tag{4.4.2}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \overline{\Phi}_{\mathrm{g}}}{\partial z}\right) + \mathcal{B}N^2 \overline{w}_{\mathrm{a}}^* = \overline{J}_{\mathrm{g}}, \tag{4.4.3}$$

$$\frac{\partial \overline{v}_{\mathrm{a}}^*}{\partial y} + \frac{1}{\rho_{\mathrm{B}}}\frac{\partial}{\partial z}\left(\rho_{\mathrm{B}} \overline{w}_{\mathrm{a}}^*\right) = 0, \tag{4.4.4}$$

where $\overline{X}_{\mathrm{g}}$ represents all contributions to the mean zonal force per unit mass associated with gravity waves and other small-scale disturbances, $\overline{J}_{\mathrm{g}}$ is the zonal-mean radiative heating rate, and $\mathbf{S}_{\mathrm{g}}$ is the quasi-geostrophic Eliassen–Palm flux, defined by,

$$\mathbf{S}_{\mathrm{g}} = \rho_{\mathrm{B}}\left(0,\, -\overline{v_{\mathrm{g}}' u_{\mathrm{g}}'},\, \epsilon f_0^{-1} \overline{v_{\mathrm{g}}' \partial \Phi_{\mathrm{g}}' / \partial z}\right). \tag{4.4.5}$$

In Eqs.(4.4.2)–(4.4.4), the eddy forcing due to Rossby waves is expressed entirely in terms of the divergence of EP flux in the mean zonal momentum equation; in particular, no eddy quantity appears in the thermodynamic energy equation at all. Note that the EP flux divergence has a useful interpretation in terms of QGPV. Using the chain rule of differentiation and the fact that x derivatives of zonal-mean quantities vanish, one can find that (see Bretherton, 1966a; Edmon *et al.*, 1980)

$$\overline{v_{\mathrm{g}}' q_{\mathrm{g}}'} = \frac{1}{\rho_{\mathrm{B}}} \nabla \cdot \mathbf{S}_{\mathrm{g}}. \tag{4.4.6}$$

Thus, from Eqs.(4.4.2) and (4.4.6) it follows that the net eddy forcing of the mean flow can be interpreted as the meridional eddy flux of QGPV. This generalizes straightforward the well-known result of Taylor (1915), which was derived for the incompressible, two-dimensional, non-rotating fluid, to the geophysical fluid within the framework of quasi-geostrophic theory (see the discussion in Section 3.1). Incidentally, Edmon *et al.* (1980) pointed out that for Rossby wavelike disturbances, $\mathbf{S}_{\mathrm{g}}$ is parallel to the meridional projection of the local group velocity in situations where the group-velocity concept is applicable (e.g., Lighthill, 1978). Thus, the quasi-geostrophic EP flux is also a measure of net wave propagation from one height and latitude to another.

To gain further insight into the nature of the quasi-geostrophic EP flux divergence, we now multiply (4.3.1) by $\rho_{\mathrm{B}} q_{\mathrm{g}}'$ and apply the Eulerian-mean operator to the resulting equation. With the aid of Eq.(4.4.6), we obtain

$$\frac{\partial \mathcal{A}_{\mathrm{g}}}{\partial t} + \nabla \cdot \mathbf{S}_{\mathrm{g}} = \mathcal{D}_{\mathrm{g}} + O(a^3), \tag{4.4.7}$$

where

$$\mathcal{A}_g \equiv \tfrac{1}{2}\rho_B\, \overline{q_g'^2}\big/\,\overline{q}_y, \qquad (4.4.8)$$

$$\mathcal{D}_g \equiv \rho_B\, \overline{Z_g' q_g'}\big/\,\overline{q}_y; \qquad (4.4.9)$$

provided that $\overline{q}_y$ is non-zero everywhere. A useful alternative form for Eq.(4.4.7) can be given as follows (Edmon *et al.*, 1980; McIntyre and Norton, 1990)

$$\frac{\partial \widehat{\mathcal{A}}_g}{\partial t} + \nabla\cdot\mathbf{S}_g = \widehat{\mathcal{D}}_g + O(a^3), \qquad (4.4.10)$$

where

$$\widehat{\mathcal{A}}_g \equiv -\rho_B\left(\tfrac{1}{2}\overline{q}_y\overline{\eta_g'^2} + \overline{\eta_g' q_g'}\right), \qquad (4.4.11)$$

$$\widehat{\mathcal{D}}_g \equiv -\rho_B\overline{\eta_g' Z_g'}, \qquad (4.4.12)$$

with η_g' being the northward geostrophic particle displacement defined by Eq.(4.5.3). Eq.(4.4.10) can be derived by multiplying (4.3.1) by $\rho_B\eta_g'$, then applying the Eulerian-mean operator to the resulting equation. Note that Eq.(4.4.10) is valid even for $\overline{q}_y = 0$.

Eqs.(4.4.7) and (4.4.10) are two quasi-geostrophic versions of the generalized EP theorem. They make explicit the dependence of the EP flux divergence on the so-called "wave transience", namely the $\partial\overline{(\;)}/\partial t$ terms, and nonconservative wave effects. Thus, if wave dissipation and transience vanish in Eqs.(4.4.7) and (4.4.10), the EP flux divergence vanishes (Eliassen and Palm, 1961; Andrews and McIntyre, 1976a; Boyd, 1976). Under such circumstances, Eq.(4.4.2) shows that there is no forcing of the mean flow by the waves. This is the 'non-acceleration' theorem of Charney and Drazin (1961).

4.4.2 The mean effect of wave transience: Linear versus nonlinear Rossby waves

Since the paper of Edmon *et al.* (1980), EP flux has frequently been used in observational and theoretical studies as a powerful diagnostic of wave propagation and wave-induced forcing of the mean flow in the atmosphere (Dunkerton *et al.*, 1981; McIntyre, 1982; Andrews *et al.*, 1987, 1983; Gao *et al.*, 1990). According to the generalized EP theorem, the EP flux divergence depends explicitly on the wave transience and wave dissipation in a way expressed by Eq.(1.2.1). In general, the explicit expressions for $\mathcal{A}$ and $\mathcal{D}$ in Eq.(1.2.1) are very complicated in form and then it is difficult to make general statements about their signs, except in the region where quasi-geostrophic theory is a valid approximation for the flow under consideration (Andrews, 1987). For quasi-geostrophic motion, Eq.(4.4.8) shows that $\mathcal{A}_g$ takes the same sign as $\overline{q}_y$. In the region of positive $\overline{q}_y$, therefore, $\partial \mathcal{A}_g/\partial t$ is positive if the wave is 'growing' in the sense that $\partial \overline{q_g'^2}/\partial t > 0$, and is negative if the wave is decaying. Likewise, in the region of negative $\overline{q}_y$, $\partial \mathcal{A}_g/\partial t$ is negative if the wave is growing and is positive if the wave is decaying. Thus, Eq.(4.4.7) shows that in the region of positive (negative) $\overline{q}_y$, $\nabla \cdot \mathbf{S}_g$ is negative (positive) if the conservative wave is growing, and *vice versa*, provided that the assumption of small-amplitude theory remains valid. These properties imply that the wave transience[2] associated with a conservative, nonbreaking disturbance can force only a temporary mean-flow change, because the $\partial\overline{(\;)}/\partial t$ terms in Eqs.(4.4.7) and (4.4.10) tend to average out as time goes on.

In reference to long-term averages (monthly or seasonal), Robinson (1986) discounted the possibility that wave transience can account for the regions of divergent EP flux in the middle and upper stratosphere, which were observed in the climatological studies of Geller *et al.* (1983, 1984) for the Northern Hemisphere winter and of Hartmann *et al.* (1984) and Mechoso *et al.* (1985) for Southern Hemisphere winter. Robinson (*op. cit.*) further argued and demonstrated that these divergent regions were spurious and resulted from the overestimation of the horizontal momentum flux divergence when it was calculated from geostrophic winds. When the exact EP flux for a system described by the primitive equations (instead of the quasi-geostrophic model) was computed, he found that the divergent region essentially disappeared except at high latitudes in the upper stratosphere, where a region of weak divergence remained in his calculation. Similar results were also pointed out by Randel (1987) and Boville (1987). Robinson suspected the region of weak divergent EP flux remaining in his model as the inaccuracy of the numerical calculation. Using three-dimensional fields of wind derived from global satellite temperature data, Marks (1989) found that the positive EP flux divergences disappeared altogether from the stratosphere but appeared more strongly in the mesosphere. In addition, some regions of positive divergences observed by Hartmann *et al.* (1984, 1985) were found to be associated with mean westerly acceleration, suggesting their existence may be physically possible (Andrews, 1987).

The existence of regions of positive EP flux divergences in the middle atmosphere suggests that some dissipative mechanisms may act to generate wave activity under certain

[2] In general, the wave transience refers to the effect of a $\partial\overline{(\;)}/\partial t$ term, $\overline{(\;)}$ being an $O(a^2)$ mean quantity, in a relevant mean equation.

conditions (Robinson, 1986; Andrews, 1987). This possibility will be examined in detail in §4.4.3. Another possibility is that the effect of nonlinearity may contribute to positive values of $\nabla\cdot\mathbf{S}_g$ under certain conditions. We have mentioned that the contribution of wave transience to the long-term average distribution of EP flux divergences is negligible under the assumption that the waves are of small amplitude. For finite-amplitude disturbances, the issue deserves a nonlinear investigation: Firstly, the finite-amplitude generation of $\mathcal{A}_g$ is not directly related to the current value of $\overline{q}_y(y,z,t)$ (Andrews, 1987), so that there is no obvious *a priori* reason why the time-averaged $\nabla\cdot\mathbf{S}_g$ need be negative or positive. Secondly, if the transient, finite-amplitude Rossby waves are breaking, as commonly observed in the winter middle atmosphere (McIntyre and Palmer, 1983, 1984), the idea that the time-integrated effects of the $\partial\overline{(\;)}/\partial t$ terms are zero becomes invalid. In fact, even $\overline{(\;)}$ itself may not return to zero when it involves Lagrangian displacement fields that need not return to zero — and are very unlikely to return to zero if the waves have broken. In addition, when waves break, the $O(a^2)$ theory becomes invalid, too, in the sense that the $O(a^3)$ term is no longer negligible.

It has been suggested that the Rossby-wave breaking may lead to positive $\nabla\cdot\mathbf{S}_g$ in the regions of negative $\overline{q}_y$ (Mahlman and Umscheid, 1984; Mechoso *et al.*, 1985; Andrews, 1987). The fact that the sign of $\nabla\cdot\mathbf{S}_g$ is positive is made plausible by the evidence emerging in theoretical, observational and numerical studies, which show that breaking Rossby waves tend to rearrange the Rossby-Ertel potential vorticity downgradient in the north-south direction along isentropic surfaces (Stewartson, 1978; Warn and Warn, 1978; McIntyre and Palmer, 1983, 1984; Juckes and McIntyre, 1987; Juckes, 1989; Haynes, 1989; Norton, 1994). If the same holds qualitatively for quasi-geostrophic quantities on z surfaces, then $\overline{v'_g q'_g} > 0$ in the regions of negative $\overline{q}_y$, and hence $\nabla\cdot\mathbf{S}_g > 0$ by Eq.(4.4.6). On the other hand, the presence of negative $\overline{q}_y$ in some regions implies that instabilities of quasi-geostrophic disturbances might be possible (Charney and Stern, 1962). If these instabilities do develop, small-amplitude waves can be expected to grow to finite amplitude, and then a wave-breaking process is likely to follow.

Based on $\overline{U}(y,z)$ shown in Fig.1.3 and $\epsilon(z)$ defined in Eq.(4.2.36), $\overline{q}_y$ is calculated from Eq.(4.3.3) and its cross sections in the extratropical stratosphere and mesosphere (25°—65° latitudes, 10 km—85 km) for January in the Northern Hemisphere and July in the Southern Hemisphere are shown in Fig.4.9. It can be seen that $\overline{q}_y$ has its largest values in the regions of the strongest winds. This is due primarily to the second term on

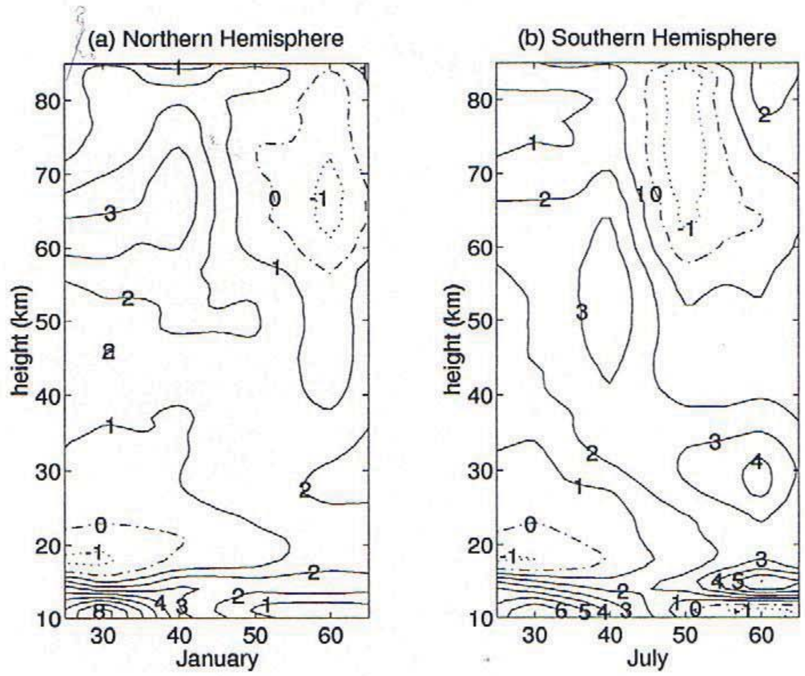


Figure 4.9: Meridional gradients (dimensionless) of the zonal mean quasi-geostrophic potential vorticity for (a) Northern Hemisphere in January and (b) Southern Hemisphere in July.

the right-hand side of Eq.(4.3.3). Above the tropopause, $\overline{q}_y$ is normally positive, except in the high-latitude mesosphere and in the middle and low-latitude lower stratosphere. Regions of negative $\overline{q}_y$ on the poleward flank of the mesospheric jet were also noted by, e.g., Dunkerton (1987) and Marks (1989). These regions roughly overlap the regions of positive EP flux divergences in the mesosphere documented by Marks (1989), suggesting that these divergent regions are likely to be related to the breaking transient Rossby waves.

4.4.3 EP flux divergences induced by steady, nonbreaking, dissipating Rossby waves

The discussion in §4.4.2 suggests that the breaking transient Rossby waves may be responsible for some positive EP flux divergences that occur in the region of negative $\overline{q}_y$.

However, Mechoso *et al.* (1985) found that some divergent EP fluxes observed in the stratosphere are actually associated with positive $\overline{q}_y$. The most likely mechanism for this phenomenon is wave dissipation. In order to concentrate on this mechanism, in this section we analyze an idealized case in which the steady, stationary Rossby waves are dissipating without breaking, and the waves are of small amplitude so that the nonlinear effect is negligible. Under such circumstances, Eq.(4.4.7) reduces to

$$\nabla\cdot\mathbf{S}_{\mathrm{g}} = \mathcal{D}_{\mathrm{g}} = \rho_{\mathrm{B}}\,\overline{Z'_{\mathrm{g}}q'_{\mathrm{g}}}\big/\,\overline{q}_y. \tag{4.4.13}$$

Eq.(4.4.13) shows that, under the assumed conditions, the value of the EP flux divergence depends on the value of $\overline{q}_y$ and the correlation of q'_{g} with Z'_{g}. The correlation of q'_{g} with Z'_{g} should be negative if the Z'_{g} term, defined by Eq.(4.3.4), represents the dissipation of q'_{g}. For example, if the Rossby waves are dissipated by Rayleigh friction and Newtonian cooling with equal constant relaxation coefficients, namely, $\alpha_{\mathrm{R}} = \alpha_{\mathrm{N}} = \alpha \neq 0$, we can see that $Z'_{\mathrm{g}} = -\mu\alpha q'_{\mathrm{g}}$, which follows from substituting Eqs.(4.3.7) and (4.3.9) into Eq.(4.3.4). Under such circumstances, Eq.(4.4.13) further reduces to

$$\overline{v'_{\mathrm{g}}q'_{\mathrm{g}}} = \frac{1}{\rho_{\mathrm{B}}}\nabla\cdot\mathbf{S}_{\mathrm{g}} = -\mu\alpha\,\overline{q'^{\,2}_{\mathrm{g}}}\big/\,\overline{q}_y. \tag{4.4.14}$$

Eq.(4.4.14) indicates that $\nabla\cdot\mathbf{S}_{\mathrm{g}}$ is negative provided that $\overline{q}_y$ is positive (Holton, 1983a; Robinson, 1986; Andrews 1987). Conversely, $\nabla\cdot\mathbf{S}_{\mathrm{g}}$ is positive if $\overline{q}_y$ is negative. As already mentioned in §4.4.2, Rossby-wave breaking can also lead to positive $\nabla\cdot\mathbf{S}_{\mathrm{g}}$ in the region of negative $\overline{q}_y$.

Now let us consider the case in which $\alpha_{\mathrm{N}} \neq \alpha_{\mathrm{R}}$. In order to avoid the complications of the critical surface, i.e. the surface on which $\overline{U}(y,z) = 0$, in what follows we assume that $\overline{U} \neq 0$. For simplicity, we further assume that $\overline{U} = O(1)$. Thus, for weakly dissipated Rossby waves ($\mu \ll 1$), the disturbance QGPV can be manipulated as follows,

$$\begin{aligned}
q'_{\mathrm{g}} &= \mathrm{Re}\left\{\left[\left(\frac{\partial^2\widehat{\psi}}{\partial y^2} - k^2\widehat{\psi}\right) + \frac{1}{\rho_{\mathrm{B}}}\frac{\partial}{\partial z}\left(\epsilon\rho_{\mathrm{B}}\frac{\partial\widehat{\psi}}{\partial z}\right)\right]\exp\left(ikx\right)\right\}\\
&= \mathrm{Re}\left\{\left[\left(\frac{\partial^2\widehat{\psi}}{\partial y^2} - k^2\widehat{\psi}\right) - \frac{\left(ik\overline{U}+\mu\alpha_{\mathrm{R}}\right)}{\left(ik\overline{U}+\mu\alpha_{\mathrm{N}}\right)}\left(\frac{\partial^2\widehat{\psi}}{\partial y^2} - k^2\widehat{\psi}\right) - \frac{ik\overline{q}_y}{\left(ik\overline{U}+\mu\alpha_{\mathrm{N}}\right)}\widehat{\psi}\right]\exp(ikx)\right\}\\
&= \mathrm{Re}\left\{\left[\frac{i\mu(\alpha_{\mathrm{R}}-\alpha_{\mathrm{N}})}{k\overline{U}}\left(\frac{\partial^2\widehat{\psi}}{\partial y^2} - k^2\widehat{\psi}\right) - \frac{\overline{q}_y}{\overline{U}}\left(1+\frac{i\mu\alpha_{\mathrm{N}}}{k\overline{U}}\right)\widehat{\psi}\right]\exp\left(ikx\right)\right\} + O(\mu^2 a).
\end{aligned} \tag{4.4.15}$$

The first expression comes from substituting solution (4.3.14) into Eq.(4.3.2). The second expression results from substituting Eq.(4.3.15) into the first expression. The last expression can be obtained by expanding the second expression in perturbation series in powers of μ.

In addition, v'_g is given by

$$v'_g = \frac{\partial \psi'_g}{\partial x} = \mathrm{Re}\left\{ik\widehat{\psi}\exp(ikx)\right\}. \tag{4.4.16}$$

Substituting Eqs.(4.4.15) and (4.4.16) into Eq.(4.4.6), we obtain

$$\nabla\cdot\mathbf{S}_g = \frac{\mu(\alpha_R - \alpha_N)\rho_B}{2\overline{U}}\mathrm{Re}\left\{\widetilde{\widehat{\psi}}\left(\frac{\partial^2\widehat{\psi}}{\partial y^2} - k^2\widehat{\psi}\right)\right\} - \frac{\mu\alpha_N\rho_B\overline{q}_y}{2\overline{U}^2}|\widehat{\psi}|^2 + O(\mu^2 a^2), \tag{4.4.17}$$

where $\widetilde{\widehat{\psi}}$ is the complex-conjugate of $\widehat{\psi}$. When $\alpha_N = \alpha_R = \alpha \neq 0$, the first term on the right-hand side of Eq.(4.4.17) vanishes, so that $\nabla\cdot\mathbf{S}_g < 0$ if $\overline{q}_y > 0$ (and vice versa). This is in agreement with the results mentioned above. When $\alpha_N \neq \alpha_R$, Eq.(4.4.17) indicates that the distribution of $\nabla\cdot\mathbf{S}_g$ depends strongly on the meridional structure of wave amplitude $\widehat{\psi}$, which is governed by Eq.(4.3.15). In particular, if $\overline{U}$ depends on z alone and $\widehat{\psi}$ is given by Eq.(4.3.16), then Eq.(4.4.17) reduces to

$$\nabla\cdot\mathbf{S}_g = -\frac{\mu(k^2+\ell^2)\left[\alpha_N(u_c - \overline{U}) + \alpha_R\overline{U}\right]}{2\overline{U}^2}\rho_B|\widehat{\Psi}|^2\cos^2(\ell y). \tag{4.4.18}$$

Eq.(4.4.18) shows that $\nabla\cdot\mathbf{S}_g$ is positive whenever

$$\alpha_N\left[u_c(z) - \overline{U}(z)\right] + \alpha_R\overline{U}(z) < 0. \tag{4.4.19}$$

In particular, if the waves are dissipated by Newtonian cooling alone (i.e., $\alpha_R = 0$ but $\alpha_N \neq 0$), $\nabla\cdot\mathbf{S}_g$ is positive when $\overline{U}(z) > u_c(z)$. Note that for constant $\overline{U}$ and u_c, Eq.(4.4.19) reduces to Eq.(4.3.36), which is the condition for $m_\dagger$ being negative.

Fig.4.10 shows the vertical distributions of $\rho_B{}^{-1}\nabla\cdot\mathbf{S}_g$ at $y = 0$. We see that $\nabla\cdot\mathbf{S}_g$ is always negative for $\alpha_N = \alpha_R = 0.5$ (Fig.4.10a). For $\alpha_N = 1$, $\alpha_R = 0$, on the other hand, significant positive $\nabla\cdot\mathbf{S}_g$ can be found in the lower stratosphere (near $z = 20$ km) for all wave components considered (Fig.4.10b). Positive values of $\nabla\cdot\mathbf{S}_g$ are also found in the mesosphere. Comparing with Fig.4.8b we can see that these positive values are associated with the condition $\overline{U} > u_c$.

In the above discussion, we assume that Rossby waves are dissipated by Rayleigh friction and Newtonian cooling. In fact, this assumption is not necessary. To carry out a more general analysis, it will prove instructive to split q'_g into a relative vorticity (RV)

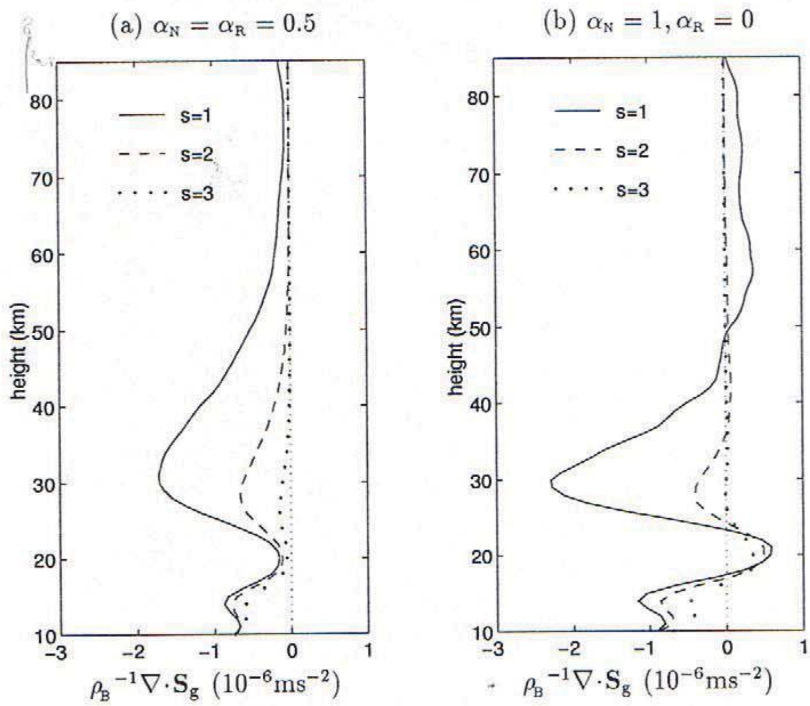


Figure 4.10: Vertical structures of $\rho_B{}^{-1}\nabla\cdot\mathbf{S}_g$ at $y = 0$, induced respectively by waves with s =1–3.

term and a static stability (SS) term, and then examine the contribution of each term to q_g' under various conditions. Note that by the RV term we refer to the sum of the first and second terms on the right-hand side of Eq.(4.3.2), and the SS term to the third term in the same equation. To keep the algebra as simple as possible we still assume that $\overline{U}$ depends on z alone and $\widehat{\psi}$ is given by Eq.(4.3.16). Thus, for conservative waves we have

$$\begin{aligned} q_g' &= \underbrace{\frac{\partial^2\psi_g'}{\partial x^2} + \frac{\partial^2\psi_g'}{\partial y^2}}_{\text{RV}} + \underbrace{\frac{1}{\rho_B}\frac{\partial}{\partial z}\left(\epsilon\rho_B\frac{\partial\psi_g'}{\partial z}\right)}_{\text{SS}} \\ &= \underbrace{-(k^2+\ell^2)\psi_g'}_{\text{RV}} + \underbrace{\frac{(\overline{U}-u_c)(k^2+\ell^2)}{\overline{U}}\psi_g'}_{\text{SS}} \\ &= -\frac{u_c(k^2+\ell^2)}{\overline{U}}\psi_g'. \end{aligned} \tag{4.4.20}$$

Using the last expression, we can rewrite the second expression in Eq.(4.4.20) as

$$q_g' = \underbrace{\frac{\overline{U}}{u_c}q_g'}_{\text{RV}} + \underbrace{\left(1-\frac{\overline{U}}{u_c}\right)q_g'}_{\text{SS}}. \tag{4.4.21}$$

When $\overline{U}/u_c > 0$, which is assumed to be true in the following discussion, Eq.(4.4.21) shows that the sign of the RV term is always identical to the sign of q'_g, while the sign of the SS term is the same as that of q'_g only if $\overline{U} < u_c$. When $\overline{U} > u_c$, the sign of the SS term is opposite to that of q'_g.

We next include the effects of mechanical and thermal dissipation on q'_g. As mentioned in §4.3.1, the mechanical dissipation of waves refers to those physical mechanisms represented by X'_g and Y'_g in the linearised momentum equations, which correlate negatively with, and therefore act to damp, the disturbance velocities u'_g and v'_g. It is not difficult to show that these mechanisms also act to damp the RV term in Eq.(4.4.21). Similarly, the thermal dissipation refers to those mechanisms represented by J'_g in the linearised thermodynamic equation, which act to damp the temperature disturbance $\partial\Phi'_g/\partial z$ and the SS term in Eq.(4.4.21) as well. Since the RV term in Eq.(4.4.21) has the same sign as q'_g, the effect of the mechanical dissipation alone is to decrease the magnitude of q'_g. In other words, the mechanical dissipation term correlates negatively with q'_g, and then always acts to damp q'_g. The negative correlation implies that the EP flux divergence is negative and the wave-induced QGPV transport is downgradient in the absence of thermal dissipation. This situation remains unchanged in the presence of thermal dissipation, provided that $\overline{U} < u_c$. If $\overline{U} > u_c$, however, the effect of thermal dissipation is to increase rather than decrease the magnitude of q'_g, since the SS term in Eq.(4.4.21) has an opposite sign to q'_g under this condition. This implies a positive correlation of the thermal dissipation term with q'_g. Therefore, if the thermal dissipation is present as the only source of dissipation for the waves, the EP flux divergence is negative when $\overline{U} < u_c$ and positive otherwise. When $\overline{U} > u_c$, positive EP flux divergences might still occur in the presence of both thermal and mechanical dissipations, depending on the relative strength of the thermal dissipation as compared with that of the mechanical dissipation. This result does not depend crucially on the details of dissipation assumptions (such as Rayleigh friction, Newtonian cooling, etc.). It is a straightforward consequence of the opposite-signed relations of the RV term and the SS term to q'_g in Eq.(4.4.21) when $\overline{U} > u_c$.

The results described above are interesting and important. They remind us that "dissipative" could, in principle, mean different things for different purposes. For example, when $\overline{U} > u_c$, "dissipative" in the sense of Newtonian cooling in our model is not dissipative from a TEM wave-activity viewpoint, because thermal dissipation actually tries to enhance QGPV anomalies.

4.5 The difference between GLM and TEM meridional circulations

Planetary-scale Rossby waves in the stratosphere play an important role in meridional and vertical transport of ozone and other chemical tracers. In the absence of eddy motions the zonal mean motion of the middle atmosphere would relax to a radiatively determined state, in which the temperature $\overline{T}$ would correspond to a radiative equilibrium, say $\overline{T}_{\mathrm{R}}$, and the circulation would consist only of a zonal-mean zonal flow in thermal wind balance with meridional gradients of $\overline{T}_{\mathrm{R}}$ (Andrews *et al.*, 1987). However, comparisons of the hypothetical radiatively-determined temperature field with the observed temperature field have revealed some striking differences in certain parts of the middle atmosphere (e.g., Fels, 1985). These differences, therefore, must be maintained by eddy transports. Since the large departure of $\overline{T}$ from $\overline{T}_{\mathrm{R}}$ implies a large net radiative heating rate $\overline{J}_{\mathrm{g}}$, it can be expected from Eqs.(4.4.3) and (4.4.4) that a TEM meridional circulation is being driven.

It is well-known that the TEM meridional circulation is generally not the same as the generalized Lagrangian mean (GLM) meridional circulation, which is more closely related to the meridional mass circulation. The difference between the TEM and GLM circulations depends on wave transience, nonlinearity, and nonconservative effects (Andrews and McIntyre, 1976a; Andrews *et al.*, 1987). Dunkerton (1978) showed that the Stokes correction to the diabatic heating was relatively small. Thus he suggested that the GLM circulation might be inferred, as a first approximation, from the TEM circulation. Since direct application of the GLM theory to atmospheric data encounters serious practical difficulties, in practice the TEM meridional circulation has been frequently used for describing and interpreting the tracer transport in the atmosphere.

The hypothesis posed by Dunkerton (*op. cit.*) is re-examined in this section. The result obtained here raises again the question of how well and under what circumstances the TEM circulation can approximate the meridional mass transport in the atmosphere.

We begin with brief discussions of the particle displacement and Stokes correction for quasi-geostrophic motion.

4.5.1 The particle displacement and Stokes correction for quasi-geostrophic motions

The particle displacement vector $\boldsymbol{\xi}'$ was formally introduced in Section 3.2 and defined by Eq.(3.2.1). For the quasi-geostrophic motion, $\boldsymbol{\xi}'$ can be consistently expanded as follows,

$$\boldsymbol{\xi}' = \boldsymbol{\xi}'_{\mathrm{g}} + \mathcal{R}\boldsymbol{\xi}'_{\mathrm{a}} + \cdots = (\xi'_{\mathrm{g}}, \eta'_{\mathrm{g}}, \zeta'_{\mathrm{g}}) + \mathcal{R}(\xi'_{\mathrm{a}}, \eta'_{\mathrm{a}}, \zeta'_{\mathrm{a}}) + \cdots, \tag{4.5.1}$$

where $\boldsymbol{\xi}'_{\mathrm{g}} = (\xi'_{\mathrm{g}}, \eta'_{\mathrm{g}}, \zeta'_{\mathrm{g}})$ may be conveniently referred to as the *geostrophic particle displacement vector*. Substituting Eq.(4.5.1) into Eq.(3.2.6) and equating coefficients of like powers of $\mathcal{R}$, with the use of Eq.(3.2.11) we obtain

$$\frac{\overline{D}\xi'_{\mathrm{g}}}{Dt} = u^l_{\mathrm{g}} = u'_{\mathrm{g}} + \boldsymbol{\xi}'_{\mathrm{g}} \cdot \nabla\overline{U} + O(a^2), \tag{4.5.2}$$

$$\frac{\overline{D}\eta'_{\mathrm{g}}}{Dt} = v^l_{\mathrm{g}} = v'_{\mathrm{g}} + O(a^2), \tag{4.5.3}$$

$$\frac{\overline{D}\zeta'_{\mathrm{g}}}{Dt} = w^l_{\mathrm{g}} = w'_{\mathrm{g}} + O(a^2), \tag{4.5.4}$$

$$\frac{\overline{D}\xi'_{\mathrm{a}}}{Dt} = u^l_{\mathrm{a}} = u'_{\mathrm{a}} + \boldsymbol{\xi}'_{\mathrm{a}} \cdot \nabla\overline{U} + O(a^2), \tag{4.5.5}$$

$$\frac{\overline{D}\eta'_{\mathrm{a}}}{Dt} = v^l_{\mathrm{a}} = v'_{\mathrm{a}} + O(a^2), \tag{4.5.6}$$

$$\frac{\overline{D}\zeta'_{\mathrm{a}}}{Dt} = w^l_{\mathrm{a}} = w'_{\mathrm{a}} + O(a^2), \tag{4.5.7}$$

where it is assumed that $\overline{u}_{\mathrm{g}}$, $\overline{v}_{\mathrm{g}}$, $\overline{w}_{\mathrm{g}}$, $\overline{u}_{\mathrm{a}}$, $\overline{v}_{\mathrm{a}}$, and $\overline{w}_{\mathrm{g}}$ are all wave-induced quantities, and therefore are all $O(a^2)$ or smaller. From the linearised version of Eq.(4.2.26) together with Eqs.(4.5.2)–(4.5.4), we have

$$\frac{\partial\xi'_{\mathrm{g}}}{\partial x} + \frac{\partial\eta'_{\mathrm{g}}}{\partial y} = O(a^2), \qquad \zeta'_{\mathrm{g}} = O(a^2), \tag{4.5.8}$$

when suitable initial conditions are given.

For a wave-induced quantity $\overline{\varphi}$, which should be $O(a^2)$ or smaller, the quasi-geostrophic version of Eq.(3.2.10) can be written as

$$\overline{\varphi}^{\mathrm{S}} = \overline{(\boldsymbol{\xi}'_{\mathrm{g}} \cdot \nabla)\varphi'} + O(\mathcal{R}a\varphi') + O(a^2\varphi') = \frac{\partial}{\partial y}\left(\overline{\eta'_{\mathrm{g}}\varphi'}\right) + O(\mathcal{R}a\varphi') + O(a^2\varphi'), \tag{4.5.9}$$

where $w'_{\mathrm{g}} = 0$ has been used. $\overline{\varphi}^{\mathrm{S}}$ is the Stokes correction to $\overline{\varphi}$. It represents the difference between the Lagrangian mean and the Eulerian mean of φ, i.e.

$$\overline{\varphi}^{\mathrm{L}} = \overline{\varphi} + \overline{\varphi}^{\mathrm{S}}. \tag{4.5.10}$$

4.5.2 Comparisons of the TEM and GLM circulations

To illustrate the wave-driven mean meridional circulation in a simple context, we now assume that the wave–mean system has settled down as a whole to an exactly steady state in which not only is the wave amplitude steady but also the mean circulation. We further assume that dissipation processes are weak in the sense that

$$X'_{\mathrm{g}} \sim O(\mu a), \quad J'_{\mathrm{g}} \sim O(\mu a), \quad \overline{X}_{\mathrm{g}} \sim O(\mu a^2), \quad \overline{J}_{\mathrm{g}} \sim O(\mu a^2). \tag{4.5.11}$$

In addition, from Eqs.(4.5.9) and (4.5.11) it follows that $\overline{X}^{\mathrm{S}}_{\mathrm{g}} \sim O(\mu a^2)$, $\overline{J}^{\mathrm{S}}_{\mathrm{g}} \sim O(\mu a^2)$, and then $\overline{X}^{\mathrm{L}}_{\mathrm{g}} \sim O(\mu a^2)$, $\overline{J}^{\mathrm{L}}_{\mathrm{g}} \sim O(\mu a^2)$.

The steady versions of Eqs.(4.4.2), (4.4.3) and (4.4.10) are as follows:

$$-f_0 \overline{v}^*_{\mathrm{a}} = \frac{1}{\rho_{\mathrm{B}}} \nabla \cdot \mathbf{S}_{\mathrm{g}} + \overline{X}_{\mathrm{g}}, \tag{4.5.12}$$

$$\mathcal{B} N^2 \overline{w}^*_{\mathrm{a}} = \overline{J}_{\mathrm{g}}, \tag{4.5.13}$$

$$\nabla \cdot \mathbf{S}_{\mathrm{g}} = -\rho_{\mathrm{B}} \overline{\eta'_{\mathrm{g}} Z'_{\mathrm{g}}}. \tag{4.5.14}$$

From Eqs.(4.5.11)–(4.5.14) we can see that $\overline{v}^*_{\mathrm{a}} = O(\mu a^2)$ and $\overline{w}^*_{\mathrm{a}} = O(\mu a^2)$. Using Eqs.(4.2.36), (4.5.9) and (4.5.10), we can manipulate Eq.(4.5.13) as follows,

$$\overline{w}^*_{\mathrm{a}} = \frac{\epsilon}{f_0^2} \overline{J}_{\mathrm{g}} = \frac{\epsilon}{f_0^2} \left(\overline{J}^{\mathrm{L}}_{\mathrm{g}} - \overline{J}^{\mathrm{S}}_{\mathrm{g}} \right) = \overline{w}^{\mathrm{L}}_{\mathrm{a}} - \frac{\epsilon}{f_0^2} \frac{\partial}{\partial y} \left(\overline{\eta'_{\mathrm{g}} J'_{\mathrm{g}}} \right) + O(\mathcal{R} \mu a^2) + O(\mu a^3). \tag{4.5.15}$$

Rewriting the above equation as

$$\overline{w}^{\mathrm{L}}_{\mathrm{a}} - \overline{w}^*_{\mathrm{a}} = \frac{\epsilon}{f_0^2} \overline{J}^{\mathrm{S}}_{\mathrm{g}} = \underbrace{\frac{\epsilon}{f_0^2} \frac{\partial}{\partial y} \overline{\left(\eta'_{\mathrm{g}} J'_{\mathrm{g}} \right)}}_{O(\mu a^2)} + O(\mathcal{R} \mu a^2) + O(\mu a^3), \tag{4.5.16}$$

we see that the difference between $\overline{w}^{\mathrm{L}}_{\mathrm{a}}$ and $\overline{w}^*_{\mathrm{a}}$ is related to $\overline{J}^{\mathrm{S}}_{\mathrm{g}}$, the Stokes correction to the diabatic heating. For adiabatic motions, $\overline{J}^{\mathrm{S}}_{\mathrm{g}} = 0$, so that

$$\overline{w}^{\mathrm{L}}_{\mathrm{a}} = \overline{w}^*_{\mathrm{a}}, \tag{4.5.17}$$

which was first pointed out by Dunkerton (1978). Eq.(4.5.17) shows that, within the framework of quasi-geostrophic theory, the TEM meridional circulation is equivalent to its GLM counterpart in the absence of thermal dissipation. Dunkerton (*op. cit.*) further argued that, even in the presence of thermal dissipation, the leading contribution to $\overline{J}^{\mathrm{S}}_{\mathrm{g}}$, i.e. the $\overline{\eta'_{\mathrm{g}} J'_{\mathrm{g}}}$ term, was small and then could be neglected. However, Eq.(4.5.16) shows that $\overline{J}^{\mathrm{S}}_{\mathrm{g}}$ is an $O(\mu a^2)$ quantity in the presence of thermal dissipation. This implies that

$\overline{J}_{\rm g}^{\rm S}$ is by no means negligible compared to $\overline{w}_{\rm a}^{*}$, which is also an $O(\mu a^2)$ quantity. In other words, the difference between the GLM and TEM meridional circulation is unlikely to be negligible when diabatic heating plays an important role.

To make the above argument more explicit, we now further assume that the mechanical dissipation due to gravity waves or other small-scale disturbances is absent, so that the dissipation of the whole system is purely thermal, namely, $X'_{\rm g} = 0$ and $\overline{X}_{\rm g} = 0$. Under such circumstances, Eqs.(4.5.12) and (4.5.14) reduce to

$$\overline{v}_{\rm a}^{*} = -\frac{1}{f_0 \rho_{\rm B}} \nabla \cdot \mathbf{S}_{\rm g}, \tag{4.5.18}$$

$$\nabla \cdot \mathbf{S}_{\rm g} = -\overline{\frac{\eta'_{\rm g}}{f_0} \frac{\partial}{\partial z} \left(\epsilon \rho_{\rm B} J'_{\rm g} \right)}, \tag{4.5.19}$$

respectively. Eq.(4.5.13) remains unchanged. From Eq.(4.5.18) we see that, in the absence of mechanical dissipation, the meridional component of the TEM circulation, $\overline{v}_{\rm a}^{*}$, is related, and has an opposite sign, to the EP flux divergence $\nabla \cdot \mathbf{S}_{\rm g}$. Substituting Eq.(4.5.19) into Eq.(4.5.18) yields

$$\overline{v}_{\rm a}^{*} = \overline{\frac{\eta'_{\rm g}}{f_0{}^2 \rho_{\rm B}} \frac{\partial}{\partial z} (\epsilon \rho_{\rm B} J'_{\rm g})}. \tag{4.5.20}$$

Substituting Eq.(4.5.20) into Eq.(4.4.4) and integrating the resulting equation with respect to z, we obtain

$$\begin{aligned} \overline{w}_{\rm a}^{*} &= \frac{1}{f_0{}^2 \rho_{\rm B}} \int_z^{\infty} \frac{\partial}{\partial y} \left[\overline{\eta'_{\rm g} \frac{\partial}{\partial z} (\epsilon \rho_{\rm B} J'_{\rm g})} \right] dz \\ &= -\frac{\epsilon}{f_0{}^2} \frac{\partial}{\partial y} \left(\overline{\eta'_{\rm g} J'_{\rm g}} \right) - \frac{1}{f_0{}^2 \rho_{\rm B}} \int_z^{\infty} \epsilon \rho_{\rm B} \frac{\partial}{\partial y} \left(\overline{J'_{\rm g} \frac{\partial \eta'_{\rm g}}{\partial z}} \right) dz, \end{aligned} \tag{4.5.21}$$

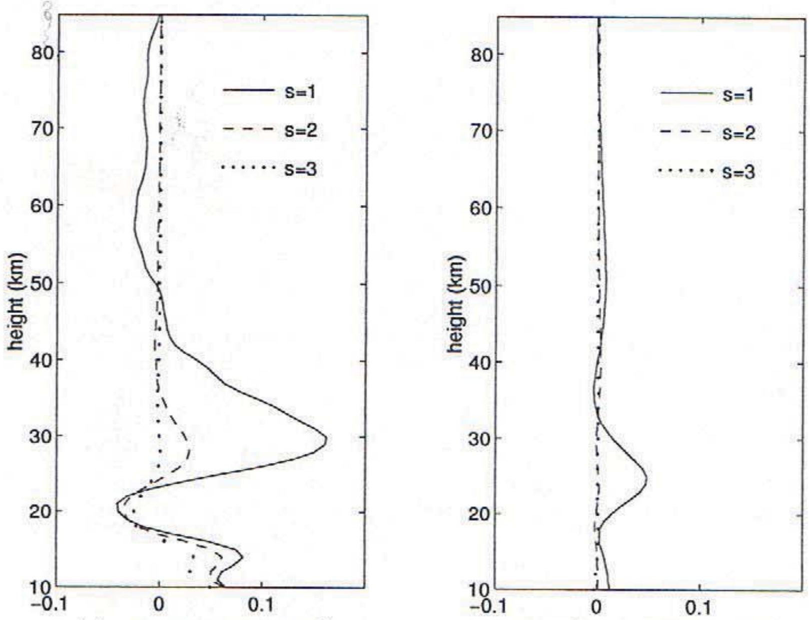
where the boundary condition $\rho_{\rm B} \overline{w}_{\rm a}^{*} \to 0$ as $z \to \infty$ has been used to fix the constant of integration. The procedure leading to Eq.(4.5.21) is referred to as the downward control principle, which states that $\overline{w}_{\rm a}^{*}$ at any altitude is directly related to the forcing above that level (McIntyre, 1987; Haynes and McIntyre, 1987; Haynes *et al.*, 1991).

Substituting Eq.(4.5.21) into Eq.(4.5.16), we see that

$$\overline{w}_{\rm a}^{\rm L} = \underbrace{-\frac{1}{f_0{}^2 \rho_{\rm B}} \int_z^{\infty} \epsilon \rho_{\rm B} \frac{\partial}{\partial y} \left(\overline{J'_{\rm g} \frac{\partial \eta'_{\rm g}}{\partial z}} \right) dz}_{\geq O(\mu a^2)} + O(\mathcal{R} \mu a^2) + O(\mu a^3). \tag{4.5.22}$$

On the other hand, the GLM continuity equation can be written as (see Andrews and McIntyre, 1978b; McIntyre, 1980a),

$$\frac{\partial \overline{v}_{\rm a}^{\rm L}}{\partial y} + \frac{1}{\rho_{\rm B}} \frac{\partial}{\partial z} \left(\rho_{\rm B} \overline{w}_{\rm a}^{\rm L} \right) = 0. \tag{4.5.23}$$



(a) $\overline{v}_{\mathrm{a}}^{*}$ at 45°N ($\mathrm{m\,s^{-1}}$) (b) $\overline{v}_{\mathrm{a}}^{\mathrm{L}}$ at 45°N ($\mathrm{m\,s^{-1}}$)

Figure 4.11: (a) Vertical distributions of the TEM meridional velocity $\overline{v}_{\mathrm{a}}^{*}$ at 45°N, driven by steady, thermally dissipative Rossby waves (i.e., $\alpha_{\mathrm{N}} = 1, \alpha_{\mathrm{R}} = 0$) with s = 1–3, respectively. (b) Same as (a), but for the GLM meridional velocity $\overline{v}_{\mathrm{a}}^{\mathrm{L}}$.

Substituting Eq.(4.5.22) into Eq.(4.5.23) and integrating the resulting equation with respect to y, we obtain

$$\overline{v}_{\mathrm{a}}^{\mathrm{L}} = -\frac{\epsilon}{f_0^2}\overline{\left(J'_{\mathrm{g}}\frac{\partial \eta'_{\mathrm{g}}}{\partial z}\right)} + O(\mathcal{R}\mu a^2) + O(\mu a^3). \tag{4.5.24}$$

An independent check of Eqs.(4.5.22) and (4.5.24) is given in Appendix 4B.

As already mentioned, the sign of $\overline{v}_{\mathrm{a}}^{*}$ in this case is opposite to the sign of the EP flux divergence $\nabla\cdot\mathbf{S}_{\mathrm{g}}$. It is shown in Section 4.4 that $\nabla\cdot\mathbf{S}_{\mathrm{g}}$ induced by thermally dissipating Rossby waves is positive where $\overline{U} > u_{\mathrm{c}}$, and *vice versa*. Therefore $\overline{v}_{\mathrm{a}}^{*}$ is negative (equatorward) where $\overline{U} > u_{\mathrm{c}}$, and *vice versa*. These results, however, are unlikely to be applicable to $\overline{v}_{\mathrm{a}}^{\mathrm{L}}$. For example, if the waves are assumed to be dissipated by Newtonian cooling alone, substituting Eqs.(4.3.7) and (4.3.14) into Eq.(4.5.24) and using Eqs.(4.5.3) and (4.2.26), we see that

$$\overline{v}_{\mathrm{a}}^{\mathrm{L}} = \frac{\mu\epsilon\alpha_{\mathrm{N}}}{f_0}\overline{\frac{\partial \psi'_{\mathrm{g}}}{\partial z}\frac{\partial}{\partial z}\left(\frac{\psi'_{\mathrm{g}}}{\overline{U}}\right)} + O(\mathcal{R}\mu a^2) + O(\mu a^3)$$

$$= \frac{\mu\epsilon\alpha_{\mathrm{N}}}{f_0\overline{U}}\left[\overline{\left(\frac{\partial\psi'_{\mathrm{g}}}{\partial z}\right)^2} - \frac{1}{\overline{U}}\frac{\partial\overline{U}}{\partial z}\overline{\left(\psi'_{\mathrm{g}}\frac{\partial\psi'_{\mathrm{g}}}{\partial z}\right)}\right] + O(\mathcal{R}\mu a^2) + O(\mu a^3). \qquad (4.5.25)$$

For constant $\overline{U}$, Eq.(4.5.25) indicates that $\overline{v}^{\mathrm{L}}_{\mathrm{a}}$ is positive whenever $\overline{U} > 0$, in contrast to the negative $\overline{v}^{*}_{\mathrm{a}}$ where $\nabla\cdot\mathbf{S}_{\mathrm{g}} > 0$ (i.e., $\overline{U} > u_{\mathrm{c}} > 0$).

In summary, in the presence of thermal dissipation, the difference between the GLM and TEM meridional circulations has the same order of magnitude as the GLM and TEM circulations themselves, so that it is unlikely to be negligible. In fact, the GLM circulation can even be opposite to its TEM counterpart under certain circumstances.

McIntyre and Norton (1990) showed that in the steady state, and with the customary neglect of the PVS transport directly contributed by the diabatic heating and mechanical forcing, the quasi-material contours defined by the displacement field do not systematically drift northwards or southwards. In other words, the GLM meridional velocity must vanish in their model. Therefore, if $\overline{v}^{\mathrm{L}}_{\mathrm{a}}$ given by Eq.(4.5.24) is non-zero, it must be related to the diffusive PVS transport directly contributed by the diabatic heating, which is one of the factors neglected in McIntyre and Norton's argument (other factors neglected by them include wave breaking and the diffusive PVS transport directly contributed by mechanical forcing, which are also neglected in the present argument). Note that the leading order term on the right-hand side of Eq.(4.5.24) depends on the vertical derivative of the northward parcel displacement η'_{g}. Therefore, if η'_{g} is independent of z, $\overline{v}^{\mathrm{L}}_{\mathrm{a}}$ is negligible to the order $O(\mu a^2)$. In this special case, the conclusion of McIntyre and Norton (*op. cit.*) is still applicable. In general situations, however, η'_{g} may vary with z, and then $\overline{v}^{\mathrm{L}}_{\mathrm{a}}$ can be of $O(\mu a^2)$.

The vertical structures of $\overline{v}^{*}_{\mathrm{a}}$ and $\overline{v}^{\mathrm{L}}_{\mathrm{a}}$ at $y = 0$, calculated respectively from Eqs.(4.5.20) and (4.5.24) with the Newtonian cooling assumption, are plotted respectively in Fig.4.11a and Fig.4.11b. The stream functions $\overline{\chi}^{*}$ and $\overline{\chi}^{\mathrm{L}}$ for $s = 1$, where $\overline{\chi}^{*}$ and $\overline{\chi}^{\mathrm{L}}$ are defined respectively by

$$\rho_{\mathrm{B}}\overline{v}^{*}_{\mathrm{a}} = -\frac{\partial\overline{\chi}^{*}}{\partial z}, \qquad \rho_{\mathrm{B}}\overline{w}^{*}_{\mathrm{a}} = \frac{\partial\overline{\chi}^{*}}{\partial y}, \qquad (4.5.26)$$

$$\rho_{\mathrm{B}}\overline{v}^{\mathrm{L}}_{\mathrm{a}} = -\frac{\partial\overline{\chi}^{\mathrm{L}}}{\partial z}, \qquad \rho_{\mathrm{B}}\overline{w}^{\mathrm{L}}_{\mathrm{a}} = \frac{\partial\overline{\chi}^{\mathrm{L}}}{\partial y}, \qquad (4.5.27)$$

are shown in Fig.4.12a and Fig.4.12b, respectively. The first feature to note is that the TEM circulation, driven by each wave component (i.e., $s = 1, 2, 3$), is opposite to its GLM counterpart in the mesosphere and the upper stratosphere. The equatorward TEM

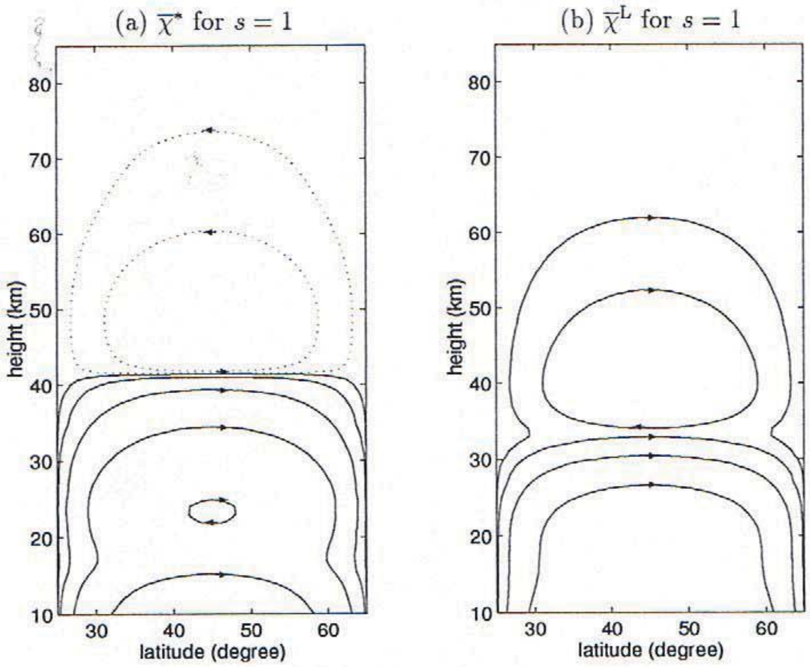


Figure 4.12: (a) Contours of the stream functions of the TEM meridional circulation, driven by steady, thermally dissipating, non-breaking Rossby waves with $s = 1$. (b) Same as (a), but for the GLM meridional circulation. In both (a) and (b), the values of contour lines are $\pm 10^0$, $\pm 10^1$, $\pm 10^2$, $\pm 10^3$, $\pm 10^4$, $\pm 10^5$, respectively, with unit $10^{-2}\,\mathrm{kg\,m^{-1}s^{-1}}$. Negative contours are represented by dotted lines, and positive contours by solid lines. The zero contour is omitted. Note that velocity is parallel to the stream function contours, but its magnitude is not inversely proportional to the spacing of the contours because of the unequal contour intervals, and the multiplication factor ρ_{B} in Eqs.(4.5.26) and (4.5.27).

transport in these regions is associated with the positive EP flux divergence shown in Fig.4.10b. In these regions, the GLM circulation shows consistent upward motion at low latitudes and downward motion at high latitudes, resembling to the mass-weighted zonal-mean meridional circulation in isentropic coordinates shown by Ko *et al.* (1985), and the Lagrangian-mean circulation calculated by Lyjak and Smith (1987). The second feature to note is the equatorward TEM transport in the lower stratosphere, where the EP flux divergence induced by the thermally dissipating Rossby waves is positive (see Fig.4.10b). In comparison, the GLM transport in this region is much weaker, and usually poleward. The third feature is that the TEM circulation is much stronger (over two times) than the corresponding GLM circulation. Incidentally, the weak equatorward GLM transport near $z = 35$ km (see Fig.4.11b and Fig.4.12b) is likely to be associated with the large vertical shear of basic zonal flow $\overline{U}$ there.

4.6 Discussion

In this chapter, we have focused our attention on the structure and mean effects of dissipative Rossby waves in a Charney–Drazin model on the mid-latitude β-plane. In order to clarify the fundamental characteristics as far as possible, we have assumed that the waves are dissipated by Newtonian cooling and/or Rayleigh friction, and, in most cases, neglected the meridional variation of the mean zonal flow. In the case of constant mean zonal flow $\overline{U}$ and constant buoyancy frequency N, our results show that the vertical wave-energy flux is always positive (upward) in the presence of Newtonian cooling and/or Rayleigh friction. Because of wave dissipation, the wave-energy density always decays with altitude, though the amplitude of the disturbance geopotential (or disturbance velocities, temperature, etc.) may, or may not, grow with height. The extra damping of waves due to weak dissipation is more important in the CD window defined by Eq.(4.3.34), than out of the window. By contrast, the modification of wave phase by dissipation is less significant in the CD window than out of the window. If the waves are dissipated by Newtonian cooling alone, their phase lines tilt westward in the GG window defined by Eq.(4.3.35), and eastward otherwise. These results remain qualitatively applicable when $\overline{U}$ and N vary with height.

Our analyses show that the existence of divergent EP flux in the stratosphere and mesosphere is physically possible. In regions of negative (southward) gradient of mean QGPV, it is pointed out that transient, conservative breaking Rossby waves are likely to result in positive $\nabla\cdot\mathbf{S}_g$, and, in addition, steady, nonbreaking Rossby waves also lead to positive $\nabla\cdot\mathbf{S}_g$ if the waves are dissipated by Rayleigh friction and Newtonian cooling with equal constant relaxation coefficients. In regions of positive (northward) gradient of mean QGPV, it is found that the sign of EP flux divergence associated with steady, nonbreaking Rossby waves depends on the strength of the thermal dissipation as compared with that of the mechanical dissipation. For example, if the Rossby waves are dissipated by Rayleigh friction and Newtonian cooling with equal constant relaxation coefficients, the associated EP flux divergence is always negative, as pointed out by Holton (1983a), Robinson (1986), and Andrews (1987). On the other hand, if the waves are dissipated by Newtonian cooling alone, positive EP flux divergence will occur whenever the basic zonal wind $\overline{U}$ exceeds the critical Rossby velocity u_c defined by Eq.(4.3.20).

The effects of steady, thermally dissipative Rossby waves on the mass transport were also investigated in this chapter. The problem was formulated in both the GLM and TEM frameworks. Results indicate that, in the presence of thermal dissipation, the difference between the GLM and TEM meridional circulations is unlikely to be negligible. This is in sharp contrast to the argument of Dunkerton (1978). In particular, the TEM meridional circulation can be opposite to its GLM counterpart in some cases. This fact reminds us that the TEM and GLM descriptions of the same thing could look strikingly different under certain circumstances.

The model used in this chapter is highly simplified. In order to represent more faithfully the wave–mean interactions in the real atmosphere, the restriction to linear, steady Rossby waves must be relaxed to include the Rossby- and gravity-wave breaking events, and not only vertical shear, but also horizontal shear of the basic zonal wind should be taken into account. In addition, further observational and theoretical work needs to be done to obtain more accurate parameterization schemes for the dissipative processes.

Appendix 4A: Demonstration of the upward wave-energy flux due to Rossby waves in the case of constant $\overline{U}$ and N^2

In this appendix, we wish to show whether the wave solution expressed by Eq.(4.3.25) with (4.3.28) and (4.3.29) in the case of constant mean zonal flow $\overline{U}$ and constant buoyancy frequency N is consistent with the assumption that the waves are forced from below. A wave that is forced from below should obey the *radiation condition*: that is, it should not transfer energy downward at great heights (Andrews *et al.*, 1987). Multiplying the linearised versions of Eqs.(4.2.27), (4.2.28), (4.2.29) by $\rho_{\mathrm{B}} u'_{\mathrm{g}}$, $\rho_{\mathrm{B}} v'_{\mathrm{g}}$, $\rho_{\mathrm{B}} \epsilon f_0^{-2} \partial \Phi'_{\mathrm{g}} / \partial z$, respectively, adding the resulting equations and averaging zonally, with a little manipulation we obtain the wave-energy equation

$$\frac{\partial}{\partial t}\left\{\tfrac{1}{2}\rho_{\mathrm{B}}\left[\overline{u_{\mathrm{g}}'^{2}}+\overline{v_{\mathrm{g}}'^{2}}+\frac{\epsilon}{f_0^2}\overline{\left(\frac{\partial \Phi'_{\mathrm{g}}}{\partial z}\right)^2}\right]\right\}+\frac{\partial}{\partial y}\left[\rho_{\mathrm{B}}\left(\overline{v'_{\mathrm{g}}\Phi'_{\mathrm{a}}}+\overline{v'_{\mathrm{a}}\Phi'_{\mathrm{g}}}+\frac{\beta^{\dagger}y}{f_0}\overline{v'_{\mathrm{g}}\Phi'_{\mathrm{g}}}\right)\right]$$
$$+\frac{\partial}{\partial z}\left(\rho_{\mathrm{B}}\overline{w'_{\mathrm{a}}\Phi'_{\mathrm{g}}}\right)=\rho_{\mathrm{B}}\left(\overline{X'_{\mathrm{g}}u'_{\mathrm{g}}}+\overline{Y'_{\mathrm{g}}v'_{\mathrm{g}}}+\overline{\frac{\epsilon J'_{\mathrm{g}}}{f_0^2}\frac{\partial \Phi'_{\mathrm{g}}}{\partial z}}\right). \qquad (4A.1)$$

Eq.(4A.1) shows that the vertical wave-energy flux for the case of constant $\overline{U}$ is represented by $\rho_{\mathrm{B}}\overline{w'_{\mathrm{a}}\Phi'_{\mathrm{g}}}$. Note that w'_{a} can be obtained by substituting the wave solution

(4.3.14) with Eqs.(4.3.16) and (4.3.25) into the linearised version of Eq.(4.2.29), namely

$$\begin{aligned} w'_{\mathrm{a}} &= -\frac{\epsilon}{f_0}\left[\frac{\overline{D}}{Dt}\left(\frac{\partial \psi'_{\mathrm{g}}}{\partial z}\right) + \mu\alpha_{\mathrm{N}}\frac{\partial \psi'_{\mathrm{g}}}{\partial z}\right] \\ &= -\frac{a\epsilon}{f_0}\mathrm{Re}\left\{\left(ik\overline{U} + \mu\alpha_{\mathrm{N}}\right)(im_\dagger - n_\dagger)\cos(\ell y)\exp\left[ikx + (im_\dagger - n_\dagger)z\right]\right\}. \end{aligned} \tag{4A.2}$$

Therefore the vertical wave-energy flux is given by

$$\rho_{\mathrm{B}}\overline{w'_{\mathrm{a}}\Phi'_{\mathrm{g}}} = \tfrac{1}{2}\epsilon\left(k\overline{U}m_\dagger + \mu\alpha_{\mathrm{N}}n_\dagger\right)a^2\cos^2(\ell y)\exp\left[-2(n_\dagger + n_H)z\right]. \tag{4A.3}$$

For conservative Rossby waves, Eq.(4A.3) reduces directly to

$$\rho_{\mathrm{B}}\overline{w'_{\mathrm{a}}\Phi'_{\mathrm{g}}} = \tfrac{1}{2}\epsilon k\overline{U}m_\dagger a^2\cos^2(\ell y)\exp\left[-2(n_\dagger + n_H)z\right], \tag{4A.4}$$

which shows that $m_\dagger$ must be chosen to be positive within the CD window to guarantee an upward wave-energy flux. It can be verified that a positive $m_\dagger$ corresponds to a positive vertical group velocity (see Andrews *et al.*, 1987). Outside the CD window, the vertical wave-energy flux associated with conservative Rossby waves is zero.

For dissipating Rossby waves, Eq.(4.3.29) shows that $(n_\dagger + n_H) > 0$. Thus it can be shown that

$$\begin{aligned} \left(k\overline{U}m_\dagger + \mu\alpha_{\mathrm{N}}n_\dagger\right) &= \frac{k\overline{U}C_{\mathrm{i}}}{2(n_\dagger + n_H)} + \mu\alpha_{\mathrm{N}}n_\dagger \\ &= \frac{\mu k^2\overline{U}(k^2+\ell^2)\left[\alpha_{\mathrm{N}}\left(u_{\mathrm{c}} - \overline{U}\right) + \alpha_{\mathrm{R}}\overline{U}\right]}{2\epsilon(n_\dagger + n_H)\left(k^2\overline{U}^2 + \mu^2\alpha_{\mathrm{N}}{}^2\right)} + \mu\alpha_{\mathrm{N}}n_\dagger \\ &= \frac{\mu\left[\alpha_{\mathrm{N}}\epsilon\left(C_{\mathrm{r}} + n_H{}^2\right) + \alpha_{\mathrm{R}}(k^2+\ell^2)\right]}{2\epsilon(n_\dagger + n_H)} + \mu\alpha_{\mathrm{N}}n_\dagger \\ &= \frac{\mu\left[\alpha_{\mathrm{N}}\epsilon\left(m_\dagger{}^2 + n_\dagger{}^2\right) + \alpha_{\mathrm{R}}(k^2+\ell^2)\right]}{2\epsilon(n_\dagger + n_H)} \\ &> 0 \qquad [\text{because } (n_\dagger + n_H) > 0]. \end{aligned} \tag{4A.5}$$

The first expression results from using Eq.(4.3.27). The second expression results from substituting Eq.(4.3.22) into the first expression. The third expression comes from using Eq.(4.3.21) with a little manipulation. The fourth expression comes from using Eq.(4.3.26). Substituting Eq.(4A.5) into Eq.(4A.3) we see that the vertical wave-energy flux is always positive for dissipative Rossby waves, in accordance with the assumption that the waves are forced from below.

Appendix 4B: An independent check of Eqs.(4.5.22) and (4.5.24)

Assume that the wave-mean system has settled down as a whole to an exactly steady state and the dissipation of the whole system is purely thermal. Under such circumstances, the GLM equation for the zonal momentum can be written as (see Andrews and McIntyre, 1978b; McIntyre, 1980a),

$$f_0 \overline{v}_{\mathrm{a}}^{\mathrm{L}} = -\overline{j_{\mathrm{g}}' \frac{\partial \zeta_{\mathrm{a}}'}{\partial x}}, \tag{4B.1}$$

where j_{g}' is defined by

$$\frac{\overline{D} j_{\mathrm{g}}'}{Dt} \equiv J_{\mathrm{g}}^{l} = J_{\mathrm{g}}' + O(\mu a^2). \tag{4B.2}$$

For stationary Rossby waves, $\overline{D}/Dt = \overline{U}\partial/\partial x$, and then Eq.(4B.1) reduces to

$$f_0 \overline{v}_{\mathrm{a}}^{\mathrm{L}} = -\overline{j_{\mathrm{g}}' \frac{\partial \zeta_{\mathrm{a}}'}{\partial x}} = \overline{\zeta_{\mathrm{a}}' \frac{\partial j_{\mathrm{g}}'}{\partial x}} = \frac{\overline{\zeta_{\mathrm{a}}' J_{\mathrm{g}}'}}{\overline{U}} + O(\mu a^3), \tag{4B.3}$$

and from Eqs.(4.5.3) and (4.2.26) we have

$$\psi_{\mathrm{g}}' = \overline{U} \eta_{\mathrm{g}}' + O(a^2). \tag{4B.4}$$

In addition, the linearised version of Eq.(4.2.29) can be written as

$$\begin{aligned} w_{\mathrm{a}}' &= -\frac{\epsilon}{f_0{}^2} \left[\frac{\overline{D}}{Dt} \left(\frac{\partial \Phi_{\mathrm{g}}'}{\partial z} \right) - f_0 v_{\mathrm{g}}' \frac{d\overline{U}}{dz} \right] + O(\mu a) \\ &= -\frac{\epsilon}{f_0} \frac{\overline{D}}{Dt} \left(\frac{\partial \psi_{\mathrm{g}}'}{\partial z} - \eta_{\mathrm{g}}' \frac{d\overline{U}}{dz} \right) + O(\mu a). \end{aligned} \tag{4B.5}$$

Combination of Eq.(4.5.7) with Eq.(4B.5) gives

$$\zeta_{\mathrm{a}}' = -\frac{\epsilon}{f_0} \left(\frac{\partial \psi_{\mathrm{g}}'}{\partial z} - \eta_{\mathrm{g}}' \frac{d\overline{U}}{dz} \right) + O(\mu a) + O(a^2). \tag{4B.6}$$

Substituting Eq.(4B.6) into Eq.(4B.3) and using Eq.(4B.4), we obtain

$$\overline{v}_{\mathrm{a}}^{\mathrm{L}} = -\frac{\epsilon}{f_0{}^2} \overline{\left(J_{\mathrm{g}}' \frac{\partial \eta_{\mathrm{g}}'}{\partial z} \right)} + O(\mu^2 a^2) + O(\mu a^3). \tag{4B.7}$$

To $O(\mu a^2)$, Eq.(4B.7) is identical to Eq.(4.5.24).

Substituting Eq.(4B.7) into Eq.(4.5.23) gives

$$\frac{\partial}{\partial z} \left(\rho_{\mathrm{B}} \overline{w}_{\mathrm{a}}^{\mathrm{L}} \right) = -\rho_{\mathrm{B}} \frac{\partial \overline{v}_{\mathrm{a}}^{\mathrm{L}}}{\partial y} = \frac{\epsilon \rho_{\mathrm{B}}}{f_0{}^2} \frac{\partial}{\partial y} \overline{\left(J_{\mathrm{g}}' \frac{\partial \eta_{\mathrm{g}}'}{\partial z} \right)} + O(\mu^2 a^2) + O(\mu a^3). \tag{4B.8}$$

Integrating Eq.(4B.8) with respect to z, we obtain

$$\overline{w}_{\mathrm{a}}^{\mathrm{L}} = \underbrace{-\frac{1}{{f_0}^2\rho_{\mathrm{B}}}\int_z^\infty \epsilon\rho_{\mathrm{B}}\frac{\partial}{\partial y}\overline{\left(J'_{\mathrm{g}}\frac{\partial\eta'_{\mathrm{g}}}{\partial z}\right)}dz}_{\geq O(\mu a^2)} + O(\mu^2 a^2) + O(\mu a^3), \tag{4B.9}$$

where the boundary condition $\rho_{\mathrm{B}}\overline{w}_{\mathrm{a}}^{\mathrm{L}} \to 0$ as $z \to \infty$ has been used to fix the constant of integration. To $O(\mu a^2)$, Eq.(4B.9) is identical to Eq.(4.5.22).

CHAPTER 5

STRATOSPHERIC ROSSBY WAVES ON AN IDEALIZED POLAR VORTEX

5.1 Introduction

In this chapter the discussion of extratropical Rossby waves and their mean effects will be continued. We shall concentrate exclusively upon the Rossby waves associated with the winter-time stratospheric polar vortex. The primary motivation is the considerable importance of such waves for understanding the formation of the Antarctic ozone hole and the recently observed depletion in middle- and high-latitude total ozone of both hemispheres (Stolarski *et al.*, 1991, 1992). Our analysis may also provide important insights into the mechanism of the so-called *4-day waves* observed in the winter stratosphere (e.g., Randel and Lait, 1991).

The idea that small disturbances at a vortex edge can propagate as wave motions has long been established. An early discussion of the theory was given by Lord Kelvin (1880), who considered wave disturbances on an initially undisturbed cylindrical vortex consisting of a core of uniform vorticity, surrounded by irrotational fluid. In the atmosphere, observations have showed that the winter-time stratospheric circulation is dominated by a cold, cyclonic polar vortex and planetary-scale disturbances on that vortex. The centre of the polar vortex is surrounded by a region of unusually strong radial gradients of potential vorticity (PV) on isentropic surfaces (McIntyre and Palmer, 1983, 1984; Newman, 1986; Mahlman and Umscheid, 1987; Tuck, 1989; Haynes, 1990). Since the dynamical restoring mechanism of Rossby waves depends on the existence of the PV gradients on isentropic surfaces (Hoskins *et al.* 1985), the intense PV gradients at the edge of the polar vortex are expected to be well able to support Rossby waves that would have a horizontal structure decaying exponentially away from the vortex edge (McIntyre and Palmer, 1983, 1984; Tao, 1991; Dritschel and Saravanan, 1994). To develop a suitable and simple model for the dynamical analysis of these waves, in Section 5.2 we shall first introduce the so-called γ*-plane approximation*, in which a quadratic variation of the Coriolis parameter f with latitude (the γ term) is taken into account. The approximation is motivated by the fact that the familiar linear meridional gradient of f (i.e., the β term) vanishes at the pole, so that the quadratic variation is the dominant gradient in the polar region. In Section 5.3, analytical solutions analogous to Rossby waves associated with the polar vortex are discussed in a sharp-edge model with all gradient of the PV being concentrated at the edge of the polar vortex. Both free travelling and forced Rossby waves are examined. Aspects of the roles of the friction and diabatic heating in dissipating these waves are investigated in Section 5.4, using the Rayleigh friction and the Newtonian cooling idealizations.

In Section 5.5 the eddy forcing of the mean flow induced by dissipating, non-breaking Rossby waves is discussed. Section 5.6 is concerned with the meridional circulation driven by dissipating Rossby waves. We shall also investigate if such a circulation can result in a persistent mean outflow from the polar vortex.

5.2 The γ-plane approximation in the polar region

Our starting point is the set of primitive equations given in Section 4.2. For the sake of simplicity, the buoyancy frequency N, defined by Eq.(4.2.10), is assumed to be constant in this chapter. In addition, the lower boundary is now a material surface. Thus the kinematic lower boundary condition is

$$\frac{D}{Dt}(z^* - h) = 0 \quad \text{at} \quad z^* = h(\lambda, \phi, t), \tag{5.2.1}$$

where z^* is the geometric height and $h(\lambda, \phi, t)$ the shape of the topography (Andrews *et al.*, 1987). In terms of Φ, Eq.(5.2.1) can be written as

$$\frac{D}{Dt}(\Phi - gh) + w\frac{d\Phi_{\rm B}}{dz} = 0, \quad \text{on} \quad z^* = h(\lambda, \phi, t); \tag{5.2.2}$$

where z is the 'pressure height' defined by Eq.(4.2.1). As far as the upper boundary condition is concerned, we require as in Chapter 4 that the density of wave-energy stays finite at great heights.

For the purpose of understanding the dynamical properties of the polar vortex, which have attracted the persistent attention of scientists since the discovery of the Antarctic 'ozone hole' by Farman *et al.* (1985), it is convenient to make some geometrical simplifications to the set of primitive equations. In Section 4.3 the usual β-plane approximation was introduced to simplify mid-latitude Rossby wave motions. This approximation, however, is not suitable for the analysis of the polar vortex. For one thing, the β-plane cannot

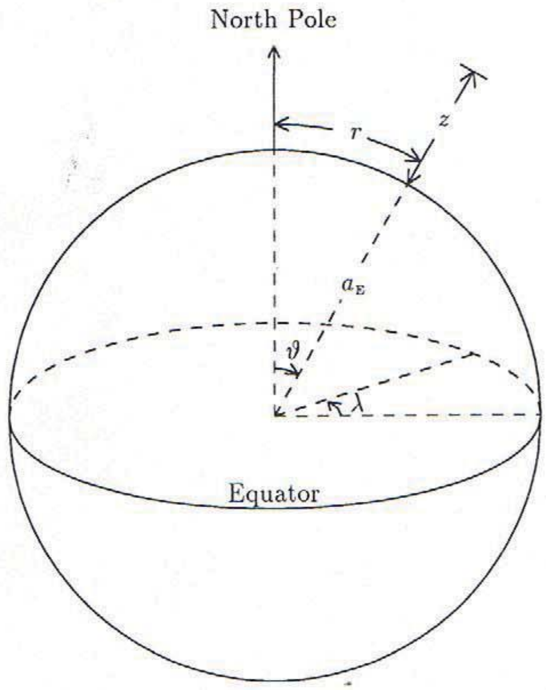


Figure 5.1: The polar-cap geometry.

accommodate the convergence of meridians toward the poles (Holton, 1975); and for another, if one tries to make the β-plane approximation using a plane tangent to the sphere at the pole, no term that corresponds to the usual β term will appear, and then the result will be merely an f-plane approximation in which the Coriolis parameter f is constant (e.g., Phillips, 1963). Though an f-plane model can provide substantial simplification in the analysis and can capture some important characteristics of the polar vortex (Polvani and Plumb, 1992), it is essential, as noted by Rossby (1949), to retain the significant characteristics of the spherical earth in the vicinity of the pole when we are concerned with the motions of the polar vortex as a whole. To obtain a simplified model suitable for polar atmosphere, Simmons (1974) introduced a polar-cap geometry in which the region around the North Pole was mapped into cylindrical polar coordinates (Fig.5.1) and a quadratic variation of the Coriolis parameter f with latitude was included as an approximation. Similar approximations were introduced by LeBlond (1964) in a polar oceanic basin. In contrast with the β-plane, such a geometrical simplification may be referred to as the 'γ-plane', referring to the coefficient γ describing the quadratic variation of f with latitude (Nof, 1990). Note that the γ term sometimes is called the δ

term in the literature (e.g., Verkley, 1990; Yang, 1991).

In the polar-cap geometry shown in Fig.5.1, the region around the North Pole is mapped into cylindrical polar coordinates λ, r and z, where r is defined by

$$r = a_E\left(\frac{\pi}{2} - \phi\right) = a_E\vartheta, \tag{5.2.3}$$

with ϕ being the latitude as before and $\vartheta = (\frac{1}{2}\pi - \phi)$ being the co-latitude. For convenience we also define a southward velocity $\mathfrak{v}$ as

$$\mathfrak{v} = -v = a_E\frac{D\vartheta}{Dt}. \tag{5.2.4}$$

For small ϑ, the trigonometric functions in the primitive equations can be expanded as power series about the co-latitude $\vartheta = 0$, i.e.,

$$\cos\phi = \sin\vartheta = \vartheta - \frac{\vartheta^3}{6} + \cdots = \frac{r}{a_E} - \frac{r^3}{6a_E{}^3} + \cdots, \tag{5.2.5}$$

$$\sin\phi = \cos\vartheta = 1 - \frac{\vartheta^2}{2} + \cdots = 1 - \frac{r^2}{2a_E{}^2} + \cdots, \tag{5.2.6}$$

$$\frac{1}{\cos\phi} = \frac{1}{\sin\vartheta} = \frac{1}{\vartheta} + \frac{\vartheta}{6} + \cdots = \frac{a_E}{r} + \frac{r}{6a_E} + \cdots. \tag{5.2.7}$$

When Eqs.(5.2.3)–(5.2.7) are substituted into Eqs.(4.2.2), (4.2.3), (4.2.5), and (4.2.9), the primitive equations can be formally rewritten as

$$\frac{Du}{Dt} + \left[2\Omega + u\left(\frac{1}{r} + \cdots\right)\right]\left(1 - \frac{r^2}{2a_E{}^2} + \cdots\right)\mathfrak{v} + \left(\frac{1}{r} + \frac{r}{6a_E{}^2} + \cdots\right)\frac{\partial\Phi}{\partial\lambda} = X, \tag{5.2.8}$$

$$\frac{D\mathfrak{v}}{Dt} - \left[2\Omega + u\left(\frac{1}{r} + \cdots\right)\right]\left(1 - \frac{r^2}{2a_E{}^2} + \cdots\right)u + \frac{\partial\Phi}{\partial r} = Y, \tag{5.2.9}$$

$$\left(\frac{1}{r} + \frac{r}{6a_E{}^2} + \cdots\right)\left\{\frac{\partial}{\partial r}\left[\mathfrak{v}\left(r - \frac{r^3}{6a_E{}^2} + \cdots\right)\right] + \frac{\partial u}{\partial\lambda}\right\} + \frac{1}{\rho_B}\frac{\partial}{\partial z}(\rho_B w) = 0, \tag{5.2.10}$$

$$\frac{D}{Dt}\left(\frac{\partial\Phi}{\partial z}\right) + N^2 w = \frac{\kappa J}{H}, \tag{5.2.11}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\left(\frac{1}{r} + \frac{r}{6a_E{}^2} + \cdots\right)\frac{\partial}{\partial\lambda} + \mathfrak{v}\frac{\partial}{\partial r} + w\frac{\partial}{\partial z}. \tag{5.2.12}$$

Combining Eq.(5.2.2) with Eq.(5.2.11) and using Eqs.(4.2.4) and (2A.3), we have

$$\frac{D}{Dt}\left(\frac{\partial\Phi}{\partial z} - \frac{N^2}{g}\Phi + N^2 h\right) = \frac{\kappa J}{H}, \quad \text{at} \quad z = h(r,\lambda,t). \tag{5.2.13}$$

We now let V, L, D and M represent the scales of horizontal velocity, horizontal length, vertical length, and height of bottom topography, respectively. We further assume, as in Chapter 4, that geostrophic balance is the leading approximation of the motion, so that the typical magnitude of geopotential Φ is $2\Omega VL$. Then we can define the following non-dimensional variables (denoted by an over-check):

$$\left.\begin{array}{llll}\check{r} = r/L, & \check{z} = z/D, & \check{h} = h/M, & \check{t} = t/(L/V), \\ (\check{u}, \check{v}) = (u, v)/V, & & \check{w} = w/(VD/L), & \check{\Phi} = \Phi/(\cancel{f_0}VL), \\ (\check{X}, \check{Y}) = (X, Y)/(V^2/L), & & \check{J} = J/[\underbrace{\cancel{f_0}}_{2\Omega} HV^2/(\kappa D)], & \check{\rho}_{\mathrm{B}} = \rho_{\mathrm{B}}/\rho_{\mathrm{s}}. \end{array}\right\} \qquad (5.2.14)$$

2Ω

Substituting the above scalings into (5.2.8)–(5.2.13), and dropping the over-check from the non-dimensional variables, we obtain the following non-dimensional equations

$$\mathcal{R}\frac{Du}{Dt} + \left\{1 - \mathcal{R}\left[\gamma r^2 + \cdots - u\left(\frac{1}{r} + \cdots\right)\right]\right\} v + \left(\frac{1}{r} + \frac{1}{3}\mathcal{R}\gamma r + \cdots\right)\frac{\partial \Phi}{\partial \lambda} = \mathcal{R}X, \qquad (5.2.15)$$

$$\mathcal{R}\frac{Dv}{Dt} - \left\{1 - \mathcal{R}\left[\gamma r^2 + \cdots - u\left(\frac{1}{r} + \cdots\right)\right]\right\} u + \frac{\partial \Phi}{\partial r} = \mathcal{R}Y, \qquad (5.2.16)$$

$$\left(\frac{1}{r} + \frac{1}{3}\mathcal{R}\gamma r + \cdots\right)\left\{\frac{\partial}{\partial r}\left[v\left(r - \frac{1}{3}\mathcal{R}\gamma r^3 + \cdots\right)\right] + \frac{\partial u}{\partial \lambda}\right\} + \frac{1}{\rho_{\mathrm{B}}}\frac{\partial}{\partial z}(\rho_{\mathrm{B}} w) = 0, \qquad (5.2.17)$$

$$\mathcal{R}\frac{D}{Dt}\left(\frac{\partial \Phi}{\partial z}\right) + \mathcal{B}w = \mathcal{R}J, \qquad (5.2.18)$$

and the lower boundary condition

$$\frac{D}{Dt}\left(\frac{\partial \Phi}{\partial z} - n_\kappa \Phi + ah\right) = J, \quad \text{at } z = \mathcal{R}ah(r, \lambda, t)/\mathcal{B}, \qquad (5.2.19)$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + u\left(\frac{1}{r} + \frac{1}{3}\mathcal{R}\gamma r + \cdots\right)\frac{\partial}{\partial \lambda} + v\frac{\partial}{\partial r} + w\frac{\partial}{\partial z}, \qquad (5.2.20)$$

and the non-dimensional parameters $\mathcal{R}$ (Rossby number), $\mathcal{B}$ (Burger number), γ, a and n_κ are defined respectively as follows,

$$\mathcal{R} = \frac{V}{2\Omega L}, \quad \mathcal{B} = \left(\frac{ND}{2\Omega L}\right)^2, \quad \gamma = \frac{\Omega L^3}{V a_{\mathrm{E}}{}^2}, \quad a = \frac{N^2 DM}{2\Omega VL}, \quad n_\kappa = \frac{N^2 D}{g}. \qquad (5.2.21)$$

In the polar stratosphere, the horizontal velocity scale can be taken to be $30\,\mathrm{ms}^{-1}$. The appropriate horizontal length scale L for our model is the radius of the polar vortex, which is about 3×10^6 m. This specific magnitude also corresponds to the wavelength of zonal wave number 1 around the 60°N divided by 2π. Other parameter values are assumed as $H = 7 \times 10^3$ m, $D = 10^4$ m, $M = 500$ m, and $N^2 = 5 \times 10^{-4}\mathrm{s}^{-2}$. This choice of scales, together with the facts that $a_{\mathrm{E}} = 6.371 \times 10^6\,\mathrm{m}$ and $2\Omega = 1.4584 \times 10^{-4}\mathrm{s}^{-1}$, leads to

$$\mathcal{R} = 0.0686, \quad \mathcal{B} = 0.2612, \quad \gamma = 1.6169, \quad a = 0.1905, \quad n_\kappa = 0.5102. \qquad (5.2.22)$$

Eq.(5.2.22) shows that the Rossby number $\mathcal{R}$ may be regarded as a small parameter. Therefore all variable fields can be expanded in asymptotic series in $\mathcal{R}$, namely,

$$\left.\begin{array}{l} u(r,\lambda,z,t,\mathcal{R}) = \mathcal{R}^0 u_{\mathrm{g}}(r,\lambda,z,t) + \mathcal{R}^1 u_{\mathrm{a}}(r,\lambda,z,t) + \cdots \\ v(r,\lambda,z,t,\mathcal{R}) = \mathcal{R}^0 v_{\mathrm{g}}(r,\lambda,z,t) + \mathcal{R}^1 v_{\mathrm{a}}(r,\lambda,z,t) + \cdots \\ \vdots \end{array}\right\} \tag{5.2.23}$$

where v_{g}, v_{a}, etc., are independent of $\mathcal{R}$. Substituting these expansions into Eqs.(5.2.15)–(5.2.20) and equating coefficients of like powers of $\mathcal{R}$, we obtain the leading order equations

$$v_{\mathrm{g}} = -\frac{1}{r}\frac{\partial \Phi_{\mathrm{g}}}{\partial \lambda}, \quad u_{\mathrm{g}} = \frac{\partial \Phi_{\mathrm{g}}}{\partial r}, \quad w_{\mathrm{g}} = 0; \tag{5.2.24}$$

and the first order equations

$$\frac{D_{\mathrm{g}} u_{\mathrm{g}}}{Dt} - \left(\gamma r^2 - \frac{u_{\mathrm{g}}}{r}\right) v_{\mathrm{g}} + \frac{\gamma r}{3}\frac{\partial \Phi_{\mathrm{g}}}{\partial \lambda} - X_{\mathrm{g}} = -\frac{1}{r}\frac{\partial \Phi_{\mathrm{a}}}{\partial \lambda} - v_{\mathrm{a}}, \tag{5.2.25}$$

$$\frac{D_{\mathrm{g}} v_{\mathrm{g}}}{Dt} + \left(\gamma r^2 - \frac{u_{\mathrm{g}}}{r}\right) u_{\mathrm{g}} - Y_{\mathrm{g}} = -\frac{\partial \Phi_{\mathrm{a}}}{\partial r} + u_{\mathrm{a}}, \tag{5.2.26}$$

$$\frac{\gamma}{3r}\frac{\partial}{\partial r}(v_{\mathrm{g}} r^3) = \frac{1}{r}\left[\frac{\partial}{\partial r}(r v_{\mathrm{a}}) + \frac{\partial u_{\mathrm{a}}}{\partial \lambda}\right] + \frac{1}{\rho_{\mathrm{B}}}\frac{\partial}{\partial z}(\rho_{\mathrm{B}} w_{\mathrm{a}}), \tag{5.2.27}$$

$$\left[\frac{D_{\mathrm{g}}}{Dt}\left(\frac{\partial \Phi_{\mathrm{g}}}{\partial z}\right) - J_{\mathrm{g}}\right] = -\mathcal{B} w_{\mathrm{a}}, \tag{5.2.28}$$

where

$$\frac{D_{\mathrm{g}}}{Dt} = \frac{\partial}{\partial t} + \frac{u_{\mathrm{g}}}{r}\frac{\partial}{\partial \lambda} + v_{\mathrm{g}}\frac{\partial}{\partial r} \tag{5.2.29}$$

is the convective derivative in the quasi-geostrophic flow.

The corresponding lower boundary condition is

$$\frac{D_{\mathrm{g}}}{Dt}\left(\frac{\partial \Phi_{\mathrm{g}}}{\partial z} - n_{\kappa}\Phi_{\mathrm{g}} + ah\right) = J_{\mathrm{g}}, \quad \text{at } z = 0. \tag{5.2.30}$$

Taking $r^{-1}\{\partial\,[r(5.2.25)]\,\partial r - \partial\,[(5.2.26)]\,\partial\lambda\}$, after some manipulation, yields

$$\begin{aligned} \frac{D_{\mathrm{g}}}{Dt}\left\{-\gamma r^2 + \frac{1}{r}\left[\frac{\partial}{\partial r}(r u_{\mathrm{g}}) - \frac{\partial v_{\mathrm{g}}}{\partial \lambda}\right]\right\} - \frac{\gamma}{3r}\frac{\partial}{\partial r}(r^3 v_{\mathrm{g}}) \\ -\frac{1}{r}\left[\frac{\partial}{\partial r}(r X_{\mathrm{g}}) - \frac{\partial Y_{\mathrm{g}}}{\partial \lambda}\right] = -\frac{1}{r}\left[\frac{\partial}{\partial r}(r v_{\mathrm{a}}) + \frac{\partial u_{\mathrm{a}}}{\partial \lambda}\right]. \end{aligned} \tag{5.2.31}$$

Combining Eq.(5.2.31) with Eqs.(5.2.27) and (5.2.28), we obtain the QGPV equation as follows,

$$\frac{D_{\mathrm{g}} q_{\mathrm{g}}}{Dt} = Z_{\mathrm{g}}, \tag{5.2.32}$$

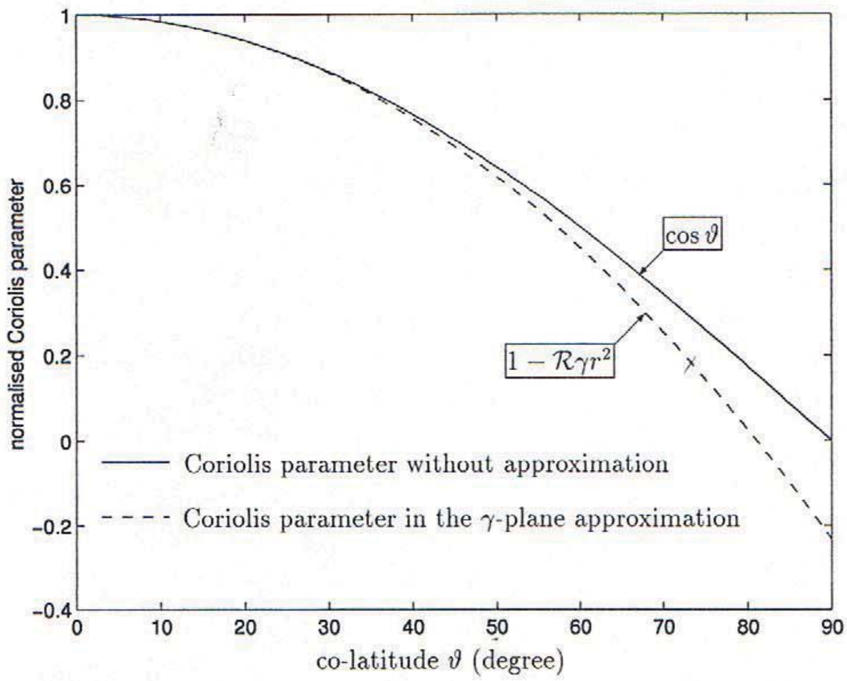


Figure 5.2: The Coriolis parameter (solid line) and its approximation on the γ-plane (dashed line), both are normalised by 2Ω.

where

$$
\begin{aligned}
q_{\mathrm{g}} &= -\gamma r^2 + \frac{1}{r}\left[\frac{\partial}{\partial r}(r u_{\mathrm{g}}) - \frac{\partial v_{\mathrm{g}}}{\partial \lambda}\right] + \frac{1}{\mathcal{B}\rho_{\mathrm{B}}}\frac{\partial}{\partial z}\left(\rho_{\mathrm{B}}\frac{\partial \Phi_{\mathrm{g}}}{\partial z}\right) \\
&= -\gamma r^2 + \frac{1}{r^2}\left[r\frac{\partial}{\partial r}\left(r\frac{\partial \Phi_{\mathrm{g}}}{\partial r}\right) + \frac{\partial^2 \Phi_{\mathrm{g}}}{\partial \lambda^2} + \frac{r^2}{\mathcal{B}\rho_{\mathrm{B}}}\frac{\partial}{\partial z}\left(\rho_{\mathrm{B}}\frac{\partial \Phi_{\mathrm{g}}}{\partial z}\right)\right],
\end{aligned}
\tag{5.2.33}
$$

$$
Z_{\mathrm{g}} = \frac{1}{r}\left[\frac{\partial}{\partial r}(r X_{\mathrm{g}}) - \frac{\partial Y_{\mathrm{g}}}{\partial \lambda}\right] + \frac{1}{\mathcal{B}\rho_{\mathrm{B}}}\frac{\partial}{\partial z}(\rho_{\mathrm{B}} J_{\mathrm{g}}). \tag{5.2.34}
$$

It can be seen that the effect of the spherical shape of the earth in the QGPV equation is represented by the term $-\gamma r^2$ (γ term) alone. Thus the γ term in the γ-plane approximation plays the same role as the β term plays in the mid-latitude β-plane approximation. Its presence is due to the variation of the Coriolis parameter f with latitude. In contrast with the fact that f linearly varies with latitude on the mid-latitude or equatorial β-planes, f varies quadratically with latitude on a polar γ-plane. Fig.5.2 shows that the variation of f in high latitudes is simulated quite well on the γ-plane. The importance of the γ effect for the dynamics of stratospheric polar vortex was first recognised by Rossby (1949) and considered in detail by Simmons (1974). In the analysis given above, the

γ-plane approximation is justified for the QGPV equation in the polar region.

5.3 Conservative interfacial Rossby waves associated with the polar vortex

Since Rossby waves owe their existence to isentropic gradients of PV, they are expected to be significant around the edge of the polar vortex that is characterized by unusually large isentropic gradients of PV (McIntyre and Palmer, 1983, 1984). This section uses an idealized, analytically tractable model in the γ-plane to investigate the characteristics of Rossby waves associated with a polar vortex. Both free travelling and forced waves are analyzed. The effects of friction and diabatic heating on these waves will be discussed in the Section 5.4.

5.3.1 A sharp-edge model for the polar vortex

To get the simplest relevant model, we consider a steady, zonally symmetric basic zonal flow $\overline{U}(r)$, which is independent of height. From Eq.(5.2.33), $\overline{U}(r)$ can be related to a zonal mean QGPV by

$$\overline{q}_g = -\gamma r^2 + \frac{1}{r}\frac{\partial}{\partial r}(r\overline{U}). \tag{5.3.1}$$

To elucidate and emphasize the dynamical effect of the steep PV gradient on the edge of the polar vortex, we shall consider first an undisturbed polar vortex of uniform QGPV, say $\overline{q}_1$, surrounded by an infinite region of uniform and lower QGPV, say $\overline{q}_2$, as shown schematically in Fig.5.3. Following Dritschel and Saravanan (1994), we shall refer to this idealized model as the *'sharp-edge model'*.

Assume that the vortex has a radius r_e. The QGPV gradients can be expressed by

$$\overline{q}_r \equiv \frac{\partial \overline{q}_g}{\partial r} = -(\overline{q}_1 - \overline{q}_2)\,\delta_{r_e}(r) = -\Delta\overline{q}\,\delta_{r_e}(r), \tag{5.3.2}$$

where $\Delta\overline{q} \equiv (\overline{q}_1 - \overline{q}_2) > 0$ and $\delta_{r_e}(r)$ is the Dirac δ-function defined by $\delta_{r_e}(r) = \delta(r - r_e) = 0$, $r \neq r_e$, and

$$\int_{r_1}^{r_2} \delta_{r_e}(r)F(r)dr = F(r_e) \qquad (r_1 < r_e < r_2), \tag{5.3.3}$$

where $F(r)$ is an arbitrary function that is continuous at $r = r_e$.

Integrating Eq.(5.3.1) with respect to r, yields

$$\overline{U}(r) = \frac{\overline{q}_1 r}{2} + \frac{\gamma r^3}{4} - \mathcal{H}(r - r_e)\frac{\Delta\overline{q}(r^2 - r_e^2)}{2r}, \tag{5.3.4}$$

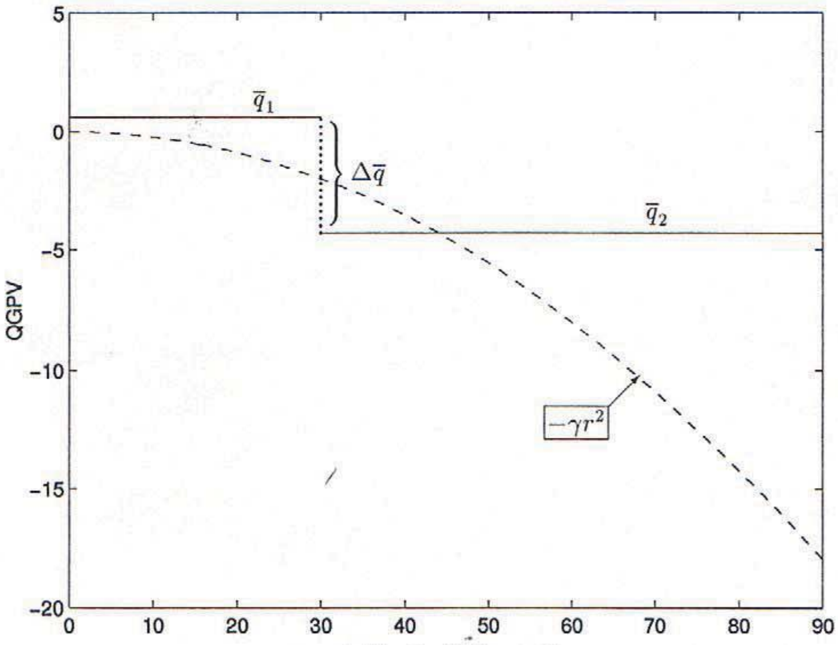


Figure 5.3: Schematic diagram of the basic QGPV (solid lines) in the sharp-edge model. The dashed line represents the variation of the Coriolis parameter, i.e., the $-\gamma r^2$ term in Eq.(5.3.1).

where $\mathcal{H}$ is the Heaviside step function defined by

$$\mathcal{H}(x) = \tfrac{1}{2}\left(1 + \frac{x}{|x|}\right) = \begin{cases} 0 & \text{for } x < 0, \\ 1 & \text{for } x > 0. \end{cases} \tag{5.3.5}$$

In the present model, we choose $r_e = 1.1119$ (i.e., $\vartheta = 30°$), $\overline{q}_1 = 0.6$, $\overline{q}_2 = -4.3$. With these parameters, variation of $\overline{U}(r)$ with the co-latitude is shown in Fig.5.4. The unrealistically strong winds near equator result from the unrealistic assumption that there is an infinite region of uniform QGPV surrounding the polar vortex. Nevertheless, their effect on the waves in question is likely to be negligible in our model because the amplitudes of the waves are expected to decay quickly away from the vortex edge, as will be shown later on.

We now consider small-amplitude disturbances, which are all taken to be $O(a)$, to the basic zonal flow mentioned above. The presence of these small disturbances can only lead to $O(a^2)$ departures of zonal mean motion from the basic state (Andrews *et al.*,

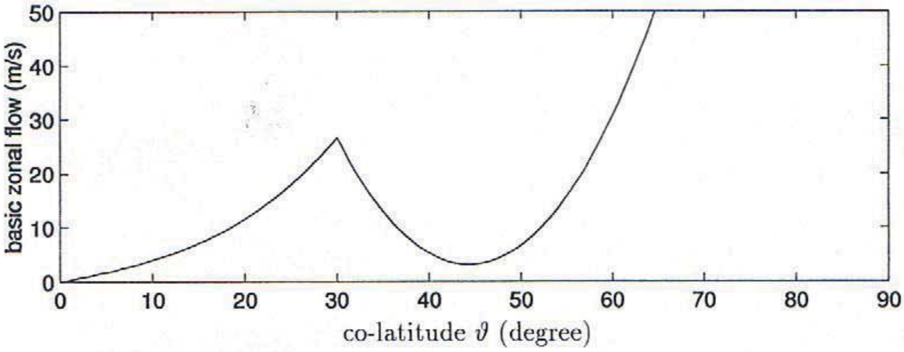


Figure 5.4: Basic zonal flow (dimensional $\overline{U}$) in the sharp-edge model on the γ-plane.

1987), namely

$$u_g = \overline{U} + u'_g + O(a^2), \quad v_g = v'_g + O(a^2), \quad \text{etc.} \tag{5.3.6}$$

Substituting Eq.(5.3.6) into Eq.(5.2.32) and other relevant equations, we obtain the disturbance QGPV equation as

$$\frac{\overline{D}q'_g}{Dt} + v'_g\overline{q}_r = Z'_g + O(a^2), \tag{5.3.7}$$

where the $O(a^2)$ term in Eq.(5.3.7) represents the error incurred by linearization, and

$$\frac{\overline{D}}{Dt} \equiv \frac{\partial}{\partial t} + \frac{\overline{U}}{r}\frac{\partial}{\partial \lambda}, \tag{5.3.8}$$

$$q'_g = \frac{1}{r^2}\left[r\frac{\partial}{\partial r}\left(r\frac{\partial \Phi'_g}{\partial r}\right) + \frac{\partial^2 \Phi'_g}{\partial \lambda^2} + \frac{r^2}{B\rho_B}\frac{\partial}{\partial z}\left(\rho_B\frac{\partial \Phi'_g}{\partial z}\right)\right], \tag{5.3.9}$$

$$Z'_g = \frac{1}{r}\left[\frac{\partial}{\partial r}(rX'_g) - \frac{\partial Y'_g}{\partial \lambda}\right] + \frac{1}{B\rho_B}\frac{\partial}{\partial z}(\rho_B J'_g). \tag{5.3.10}$$

In what follows we shall consider the linear, conservative waves, so that $Z'_g = 0$ and the effects of nonlinearity will be neglected.

5.3.2 Free travelling interfacial Rossby waves

Rossby waves in the atmosphere can be separated into *forced* and *free travelling* waves. Forced waves are continually maintained by excitation mechanisms, while free travelling waves may be initially excited by random disturbances, barotropic or baroclinic instabilities, etc., but are not necessarily so maintained as the forced waves (Andrews *et al.*, 1987). It was generally believed that forced Rossby waves are more important than free travelling Rossby waves for the stratospheric atmosphere. A recent study of Elbern and Speth (1993), however, showed that free travelling Rossby waves could account for a considerable part of global atmospheric oscillations. In this subsection, the characteristics of free travelling Rossby waves associated with a polar vortex are investigated using the disturbance QGPV equation on the γ-plane. Topographically forced Rossby waves will be discussed in §5.3.3.

To investigate free travelling Rossby waves, we assume that the lower boundary is flat, i.e., $h \equiv 0$, and look for linear solutions to Eq.(5.3.7) of the form

$$\Phi'_g = \mathrm{Re}\left\{\widehat{\Phi}(r,z)\exp[i(s\lambda-\sigma t)+n_H z]\right\} = \mathrm{Re}\left\{\widehat{\Theta}(r)\widehat{\Psi}(z)\exp[i(s\lambda-\sigma t)+n_H z]\right\}, \qquad (5.3.11)$$

where n_H is defined by Eq.(4.3.18), s is the specified zonal wavenumber, σ is the undetermined frequency, and $\widehat{\Phi}(r,z) = \widehat{\Theta}(r)\widehat{\Psi}(z)$. Substituting Eq.(5.3.11) into (5.3.7), yields

$$\frac{d}{dr}\left(r\frac{d\widehat{\Theta}}{dr}\right) - \left[\frac{sr\Delta\bar{q}\,\delta_{r_e}(r)}{\sigma r - s\overline{U}(r)} + C\,r + \frac{s^2}{r}\right]\widehat{\Theta} = 0, \qquad (5.3.12)$$

$$\frac{d^2\widehat{\Psi}}{dz^2} + \left(\mathcal{B}\,C - {n_H}^2\right)\widehat{\Psi} = 0, \qquad (5.3.13)$$

where C is a separation constant. Note that, for historical reasons, a so-called *equivalent depth* is often introduced as the separation constant in atmospheric tidal theory. Its dimensionless form, $\hbar$ say, can be related to C in Eqs.(5.3.12) and (5.3.13) by

$$\hbar = \frac{n_\kappa}{\mathcal{B}\,C}. \qquad (5.3.14)$$

It is required that wave energy per unit volume is finite everywhere. Then the boundary conditions are

$$|\widehat{\Theta}| = \text{bounded}, \quad \text{at } r = 0 \text{ and as } r \to \infty; \qquad (5.3.15)$$

$$\frac{d\widehat{\Psi}}{dz} + (n_H - n_\kappa)\widehat{\Psi} = 0, \quad \text{at } z = 0, \qquad (5.3.16)$$

$$|\widehat{\Psi}|^2 = \text{bounded}, \quad \text{as } z \to \infty. \qquad (5.3.17)$$

Now Eqs.(5.3.13), (5.3.16) and (5.3.17) compose an eigenvalue problem for the separation constant C, and Eqs.(5.3.12) and (5.3.15) compose another eigenvalue problem for the frequency σ. The solution of Eq.(5.3.13) satisfying the boundary conditions (5.3.16) and (5.3.17) can be written as

$$\widehat{\Psi}(z) = \exp\left[(n_\kappa - n_H)z\right], \qquad (5.3.18)$$

which implies that

$$C = n_\kappa(2n_H - n_\kappa)/\mathcal{B}, \tag{5.3.19}$$

$$\Phi'_g \propto \exp(n_\kappa z). \tag{5.3.20}$$

From Eqs.(5.2.24) it follows that u'_g and v'_g are also proportional to $\exp(n_\kappa z)$. Waves of this kind, which have no phase tilt with height and whose velocity and geopotential disturbances grow with height, are sometimes referred to as "Lamb waves".

Substituting Eq.(5.3.19) into Eq.(5.3.14) yields

$$\hbar = 1/(2n_H - n_\kappa). \tag{5.3.21}$$

Before proceeding to solve Eq.(5.3.12) with boundary conditions (5.3.15), it is helpful to make some general remarks on this eigenvalue problem and its solutions. Firstly, note that although Eq.(5.3.12) is written in the usual Sturm–Liouville form, it is obviously not a classical Sturm–Liouville equation on the domain $[0,\infty)$. For one thing, the eigenvalue is σ rather than C, and for another the equation is singular at $r = r_e$ because of the δ-function behaviour of $\overline{q}_r$. Secondly, the equation itself implies that, in order to balance the δ-function term in the equation, its solution must be continuous at $r = r_e$ ("continuity condition", say), and that the derivative of its solution must have a jump at $r = r_e$ ("jump condition", say). The strength of the jump is given by

$$\lim_{\epsilon\to 0}\left(r\frac{d\widehat{\Theta}}{dr}\right)\Bigg|_{r_e-\epsilon}^{r_e+\epsilon} = \frac{sr_e\Delta\overline{q}\,\widehat{\Theta}(r_e)}{\sigma r_e - s\overline{U}(r_e)}. \tag{5.3.22}$$

A good way to see this is to integrate Eq.(5.3.12) term by term from $r = r_e-$ to $r = r_e+$, namely

$$\lim_{\epsilon\to 0}\int_{r_e-\epsilon}^{r_e+\epsilon}\frac{d}{dr}\left(r\frac{d\widehat{\Theta}}{dr}\right)dr - \lim_{\epsilon\to 0}\int_{r_e-\epsilon}^{r_e+\epsilon}\frac{sr\Delta\overline{q}\,\widehat{\Theta}(r)\delta_{r_e}(r)}{\sigma r - s\overline{U}(r)}dr$$
$$-\lim_{\epsilon\to 0}\int_{r_e-\epsilon}^{r_e+\epsilon}\left(Cr + \frac{s^2}{r}\right)\widehat{\Theta}(r)dr = 0. \tag{5.3.23}$$

The first and the second terms in Eq.(5.3.23) are given by

$$\lim_{\epsilon\to 0}\int_{r_e-\epsilon}^{r_e+\epsilon}\frac{d}{dr}\left(r\frac{d\widehat{\Theta}}{dr}\right)dr = \lim_{\epsilon\to 0}\left(r\frac{d\widehat{\Theta}}{dr}\right)\Bigg|_{r_e-\epsilon}^{r_e+\epsilon}, \tag{5.3.24}$$

$$\lim_{\epsilon\to 0}\int_{r_e-\epsilon}^{r_e+\epsilon}\frac{sr\Delta\overline{q}\,\widehat{\Theta}(r)\delta_{r_e}(r)}{\sigma r - s\overline{U}(r)}dr = \frac{sr_e\Delta\overline{q}\,\widehat{\Theta}(r_e)}{\sigma r_e - s\overline{U}(r_e)}, \tag{5.3.25}$$

respectively. The third term in Eq.(5.3.23) reduces to zero in the limit, since $\widehat{\Theta}(r)$ is finite at $r = r_e$:

$$\lim_{\epsilon\to 0}\int_{r_e-\epsilon}^{r_e+\epsilon}\left(Cr+\frac{s^2}{r}\right)\widehat{\Theta}(r)dr = 0. \tag{5.3.26}$$

Then Eq.(5.3.22) follows from substituting Eqs.(5.3.24)-(5.3.26) into Eq.(5.3.23).

An immediate consequence of the jump condition is that the solution of Eq.(5.3.12) must be non-zero at $r = r_e$ ("non-zero condition", say). This is because the derivative of the solution must have a jump at $r = r_e$, so that $\widehat{\Theta}(r_e)$ appearing on the right-hand side of Eq.(5.3.22) should be different from zero.

Under the typical parameter conditions of the winter stratosphere, i.e., with use of parameters shown in Eq.(5.2.22), C given by Eq.(5.3.19) is positive (C=1.7939). Thus, away from $r = r_e$, Eq.(5.3.12) reduces to a Bessel equation whose solutions are $I_s(\sqrt{C}r)$ and $K_s(\sqrt{C}r)$, where I_s and K_s are the modified Bessel functions of the first and the second kind, respectively. With the requirements represented by lateral boundary conditions (5.3.15) and the continuity condition at $r = r_e$, we obtain the solution of Eq.(5.3.12) on the domain $[0,\infty)$ in the form

$$\widehat{\Theta}(r) = \widehat{\Theta}_s(r) = \begin{cases} B_sK_s(\sqrt{C}r_e)I_s(\sqrt{C}r), & \text{if } r \le r_e; \\ B_sI_s(\sqrt{C}r_e)K_s(\sqrt{C}r), & \text{if } r > r_e; \end{cases} \tag{5.3.27}$$

where B_s is an arbitrary constant. Following standard practice, we now choose B_s to make the solutions $\widehat{\Theta}_s(r)$ normalised with respect to the weighting function r, namely,

$$\int_0^\infty r\widehat{\Theta}_s^2(r)dr = 1. \tag{5.3.28}$$

This can be done by substituting Eq.(5.3.27) into Eq.(5.3.28) and using the following integral and recurrence formulae

$$\int_0^r \tilde{r}I_s^2(b\tilde{r})d\tilde{r} = -\frac{r^2}{2b}\left\{\frac{d}{dr}\left[I_s(br)\right]\right\}^2 + \frac{1}{2}\left(r^2+\frac{s^2}{b^2}\right)I_s^2(br), \tag{5.3.29}$$

$$\int_r^\infty \tilde{r}K_s^2(b\tilde{r})d\tilde{r} = \frac{r^2}{2b}\left\{\frac{d}{dr}\left[K_s(br)\right]\right\}^2 - \frac{1}{2}\left(r^2+\frac{s^2}{b^2}\right)K_s^2(br), \tag{5.3.30}$$

$$\frac{d}{dr}\left[I_s(r)\right] = \frac{1}{2}\left[I_{s-1}(r)+I_{s+1}(r)\right], \tag{5.3.31}$$

$$\frac{d}{dr}\left[K_s(r)\right] = -\frac{1}{2}\left[K_{s-1}(r)+K_{s+1}(r)\right], \tag{5.3.32}$$

$$I_s(r)K_{s+1}(r) + I_{s+1}(r)K_s(r) = 1/r\ , \tag{5.3.33}$$

(see Prudnikov *et al.*, , 1986, p47–p48 and Watson, 1944, p79–80). After some manipu-

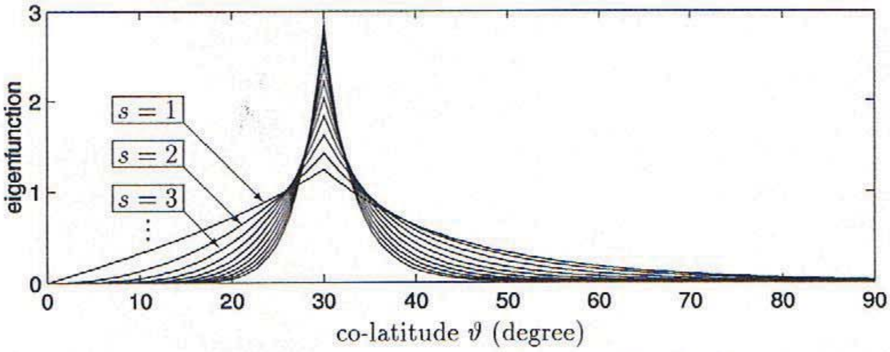


Figure 5.5: Horizontal structure of the free travelling Rossby waves associated with the polar vortex in the sharp-edge model.

lation we obtain

$$B_s = \frac{2}{r_e}\sqrt{\frac{2}{I_s^2(\sqrt{C}r_e)\left[K_{s-1}(\sqrt{C}r_e)+K_{s+1}(\sqrt{C}r_e)\right]^2 - K_s^2(\sqrt{C}r_e)\left[I_{s-1}(\sqrt{C}r_e)+I_{s+1}(\sqrt{C}r_e)\right]^2}}\,. \tag{5.3.34}$$

Eq.(5.3.27) shows that the free travelling Rossby waves discussed here are trapped on the vortex edge, as shown in Fig.5.5. These waves owe their existence to the sharp change of the PV gradient on the vortex edge. They will be referred to as the *'interfacial Rossby waves'*, in partial analogy with the *interfacial gravity waves* in the textbooks.

Note that the solution given in Eq.(5.3.27) should not be referred to as the eigenfunction of the eigenvalue problem, because it is independent of the eigenvalue σ, which has yet to be found. The eigenvalue σ can be determined by substituting the solution (5.3.27) into Eq.(5.3.22). After some manipulation we obtain

$$\sigma = s\left[\frac{\overline{U}(r_e)}{r_e} - \Delta\overline{q} I_s(\sqrt{C}r_e)K_s(\sqrt{C}r_e)\right]. \tag{5.3.35}$$

Eq.(5.3.35) shows that, in general, the free travelling wave can travel either westward or eastward, depending on whether σ is negative or positive. The case of a strong polar jet seems particularly interesting and important. When $\overline{U}(r_e) \gg r_e\Delta\overline{q}I_s(\sqrt{C}r_e)K_s(\sqrt{C}r_e)$, from Eq.(5.3.35) it follows that

$$\frac{\sigma}{s} \approx \frac{\overline{U}(r_e)}{r_e}. \tag{5.3.36}$$

In other words, waves with various zonal wavenumbers travel eastward with approximately the same zonal angular phase velocity. This means that the Rossby elasticity has negligible effect on the angular phase speed, and the wave passes the observer like a passive tracer. The existence of waves with this character in polar latitudes of the winter stratosphere has been reported by several observational studies (Venne and Stanford, 1979, 1982; Prata, 1984; Lait and Stanford, 1988, Randel and Lait, 1991). These eastward-propagating waves are substantially stronger in the Southern Hemisphere, where the polar jet is stronger, than in the Northern Hemisphere, where the polar jet is weaker. They have peak amplitudes between 60° and 75° latitudes, little or no vertical phase tilt, and largest amplitudes in the upper stratosphere. In addition, higher zonal-wavenumber components move with successively faster frequencies (wave 1 with a period near 4 days, wave 2 near 2 days, etc.), but all components travel with approximately the same zonal angular phase velocity. All these features are remarkably in accordance with the above theory.

5.3.3 Topographically forced interfacial Rossby waves

To consider topographically forced Rossby waves, we assume that the waves are forced from below by small-amplitude undulation of topography of the form

$$h = \mathrm{Re}\left\{\widehat{h}(r)\exp(is\lambda)\right\}. \tag{5.3.37}$$

For consistency with this boundary undulation, we now seek steady, conservative, and stationary solutions to Eq.(5.3.7) of the form

$$\Phi'_{\mathrm{g}} = \mathrm{Re}\left\{\widehat{\Theta}(r)\widehat{\Psi}(z)\exp(is\lambda + n_H z)\right\}. \tag{5.3.38}$$

Substituting Eq.(5.3.38) into (5.3.7) we obtain the equation for the horizontal structure function $\widehat{\Theta}(r)$ as follows,

$$\frac{d}{dr}\left(r\frac{d\widehat{\Theta}}{dr}\right) + \left[\frac{r\Delta\overline{q}\,\delta_{r_c}(r)}{\overline{U}(r)} - C\,r - \frac{s^2}{r}\right]\widehat{\Theta} = 0, \tag{5.3.39}$$

where C is the separation constant again. The equation for the vertical structure function, $\widehat{\Psi}(z)$, and the lateral and upper boundary conditions remain the same as those represented respectively by Eqs.(5.3.13), (5.3.15) and (5.3.17). The lower boundary condition from Eq.(5.2.30) is, after substitution of Eqs.(5.3.37) and (5.3.38),

$$\left(\frac{d\widehat{\Psi}}{dz} + n_H\widehat{\Psi}\right)\widehat{\Theta}(r) = -a\widehat{h}(r), \quad \text{at } z = 0. \tag{5.3.40}$$

Eqs.(5.3.39) and (5.3.15) consist of an eigenvalue problem for C. Applying the same argument to Eq.(5.3.39) as was done in §5.3.2 to Eq.(5.3.12), we see that the eigenfunction of Eq.(5.3.39) must satisfy the continuity condition, the jump condition, and the non-zero condition at $r = r_e$ (see pages 133 and 134 for more details). To find the eigenfunction we note that three possibilities arise, according as the eigenvalue C is negative, zero, or positive:

1. $C = -b_s{}^2 < 0$ *with* b_s *being a positive number*

In this case, the vertical structure equation (5.3.13) indicates that the waves always decay with height. The solutions to the horizontal structure equation (5.3.39) with boundary conditions (5.3.15), denoted by $\widehat{\Theta}_{b_s}(r)$, can be written as

$$\widehat{\Theta}(r) = \widehat{\Theta}_{b_s}(r) = \begin{cases} B_1 J_s(b_s r), & \text{if } r \leq r_e \\ B_2 J_s(b_s r) + B_3 Y_s(b_s r), & \text{if } r > r_e, \end{cases} \tag{5.3.41}$$

where B_1, B_2 and B_3 are arbitrary constants, and $J_s(b_s r)$ and $Y_s(b_s r)$ are Bessel functions of the first kind and the second kind, respectively.

Applying the continuity condition at $r = r_e$ to Eq.(5.3.41), yields

$$B_1 J_s(b_s r_e) = B_2 J_s(b_s r_e) + B_3 Y_s(b_s r_e). \tag{5.3.42}$$

The non-zero condition at $r = r_e$ implies that

$$B_1 \neq 0, \quad \text{and} \quad J_s(b_s r_e) \neq 0. \tag{5.3.43}$$

Thus, from Eq.(5.3.42) it follows that

$$B_2 = B_1 - B_3 Y_s(b_s r_e)/J_s(b_s r_e), \tag{5.3.44}$$

Substituting solution (5.3.41) and $\sigma = 0$ into the jump condition (5.3.22), with some manipulation, we obtain

$$J_s^2(b_s r_e) = -\frac{2B_3\overline{U}(r_e)}{B_1 \pi r_e \Delta\overline{q}}, \tag{5.3.45}$$

where the following recurrence formulae for $J_s(x)$ and $Y_s(x)$ have been used,

$$\frac{d}{dr}\left[J_s(r)\right] = \frac{1}{2}\left[J_{s-1}(r) - J_{s+1}(r)\right], \tag{5.3.46}$$

$$\frac{d}{dr}\left[Y_s(r)\right] = \frac{1}{2}\left[Y_{s-1}(r) - Y_{s+1}(r)\right], \tag{5.3.47}$$

$$J_s(r)Y_{s+1}(r) - J_{s+1}(r)Y_s(r) = -2/(\pi r), \tag{5.3.48}$$

(again, see Watson, 1944, p45 & p66).

From Eqs.(5.3.43) and (5.3.45) we see that that B_1 and B_3 must have opposite signs and neither of them can be zero, i.e. $B_1B_3 < 0$. In addition, any negative number $C = -b_s{}^2$ that is not a root of the equation $J_s(b_s r_e) = 0$ can be an eigenvalue of Eq.(5.3.39) with boundary conditions (5.3.15), because, apart from the requirement that $B_1B_3 < 0$, B_1 and B_3 can be arbitrarily chosen to satisfy Eq.(5.3.45). In other words, the negative eigenvalues of Eq.(5.3.39) with boundary conditions (5.3.15) consist of a piecewise continuous spectrum. Note that the associated solutions, given by Eq.(5.3.41), are not square-integrable in the sense that $\int_0^\infty r\widehat{\Theta}_{b_s}^2(r)dr$ is not convergent. These solutions sometimes are referred to as "pseudo-eigenfunctions" (Pryce, 1993). In addition, the solutions are not necessarily confined at the edge of the polar vortex. In fact, the solution given by Eq.(5.3.41) is of oscillating shape and its maximum value can approach the pole ($r = 0$) when b_s is sufficiently large. Waves represented by these solutions are not specially relevant to the problem concerned in this chapter, and, therefore, will not be discussed further.

2. $C = 0$

For $C = 0$, the solutions of Eq.(5.3.39) with boundary conditions (5.3.15) can be written as

$$\widehat{\Theta}(r) = \begin{cases} B_1(r/r_e)^s, & \text{if } r \le r_e \\ B_1(r_e/r)^s,, & \text{if } r > r_e, \end{cases} \tag{5.3.49}$$

where B_1 is an arbitrary constant. When Eq.(5.3.49) with $\sigma = 0$ is substituted into the jump condition (5.3.22), we obtain that

$$s = \frac{r_e\Delta\overline{q}}{2\overline{U}(r_e)}. \tag{5.3.50}$$

Eq.(5.3.50) shows that $C = 0$ could be an eigenvalue for the problem only when there exists an integer s that satisfies Eq.(5.3.50). This is improbable. Obviously, a wave depending on such a coincidence is of little practical interest. So this case will not be further discussed, either.

3. $C = \ell_s{}^2 > 0$ *with* ℓ_s *being a positive number*

In this case, the eigenfunctions of Eq.(5.3.39) with boundary conditions (5.3.15), denoted by $\widehat{\Theta}_{\ell_s}(r)$, can be written as

$$\widehat{\Theta}_{\ell_s}(r) = \begin{cases} B_{\ell_s}K_s(\ell_s r_e)I_s(\ell_s r), & \text{if } r \le r_e, \\ B_{\ell_s}I_s(\ell_s r_e)K_s(\ell_s r), & \text{if } r > r_e, \end{cases} \tag{5.3.51}$$

where B_{ℓ_s} is an arbitrary constant. We shall choose B_{ℓ_s} as

$$B_{\ell_s} = \frac{2}{r_e}\sqrt{\frac{2}{I_s^2(\ell_s r_e)[K_{s-1}(\ell_s r_e)+K_{s+1}(\ell_s r_e)]^2 - K_s^2(\ell_s r_e)[I_{s-1}(\ell_s r_e)+I_{s+1}(\ell_s r_e)]^2}} \tag{5.3.52}$$

to make the eigenfunction normalized in the sense that $\int_0^\infty r\widehat{\Theta}^2_{\ell_s}(r)dr = 1$ (see the relevant argument in §5.3.2).

Substituting Eq.(5.3.51) and $\sigma = 0$ into the jump condition (5.3.22) leads to the following transcendental equation

$$I_s(\ell_s r_e)K_s(\ell_s r_e) = \frac{\overline{U}(r_e)}{r_e\Delta\overline{q}}, \tag{5.3.53}$$

from which ℓ_s and then the eigenvalue $C = \ell_s^2$ can be determined. These waves, which are trapped on the vortex edge ($r = r_e$), may be called *forced interfacial Rossby waves.*

To determine ℓ_s from Eq.(5.3.53), we first note that the asymptotic formulae for the associated Bessel functions $I_s(r)$ and $K_s(r)$ for large and positive argument r may be written as (see Watson, 1944, p202–203)

$$I_s(r) \sim \sqrt{\frac{e^{2r}}{2\pi r}}\left[1 + \frac{1-4s^2}{1!\,8r} + \frac{(1-4s^2)(9-4s^2)}{2!\,(8r)^2} + \cdots\right], \tag{5.3.54}$$

$$K_s(r) \sim \sqrt{\frac{2}{\pi r e^{2r}}}\left[1 - \frac{1-4s^2}{1!\,8r} + \frac{(1-4s^2)(9-4s^2)}{2!\,(8r)^2} - \cdots\right], \tag{5.3.55}$$

respectively. Eqs.(5.3.54) and (5.3.55) suggest that $I_s(r)K_s(r) \sim (\pi r)^{-1}$ as r approaches infinity. In addition, $I_s(0)K_s(0) = (2s)^{-1}$. In fact, $I_s(x)K_s(x)$ is a monotonically decaying function of r with a finite maximum at $r = 0$, as is evident from Fig.5.6. Thus, there is at most one real solution ℓ_s of Eq.(5.3.53). In other words, if $\overline{U}(r_e)$ satisfies the following criterion

$$0 < \overline{U}(r_e) < r_e\Delta\overline{q}\, I_s(0)K_s(0) = \frac{r_e\Delta\overline{q}}{2s}, \tag{5.3.56}$$

we can expect one positive eigenvalue $C = \ell_s^2$ from Eq.(5.3.53); otherwise there will be no positive eigenvalue for the problem.

For the parameters chosen in the present model (i.e., $r_e = 1.1119$, $\overline{q}_1 = 0.6$, $\overline{q}_2 = -4.3$), it can be shown that $\overline{U}(r_e)/(r_e\Delta\overline{q}) = 0.1632$. Thus, from Fig.5.6 we see that no positive eigenvalue $C = \ell_s^2$ for the Rossby wave with $s > 3$. The positive eigenvalues for $s \leq 3$ are given in Table 5.1. The corresponding eigenfunctions are shown in Fig.5.7. We see that all waves are trapped on the vortex edge.

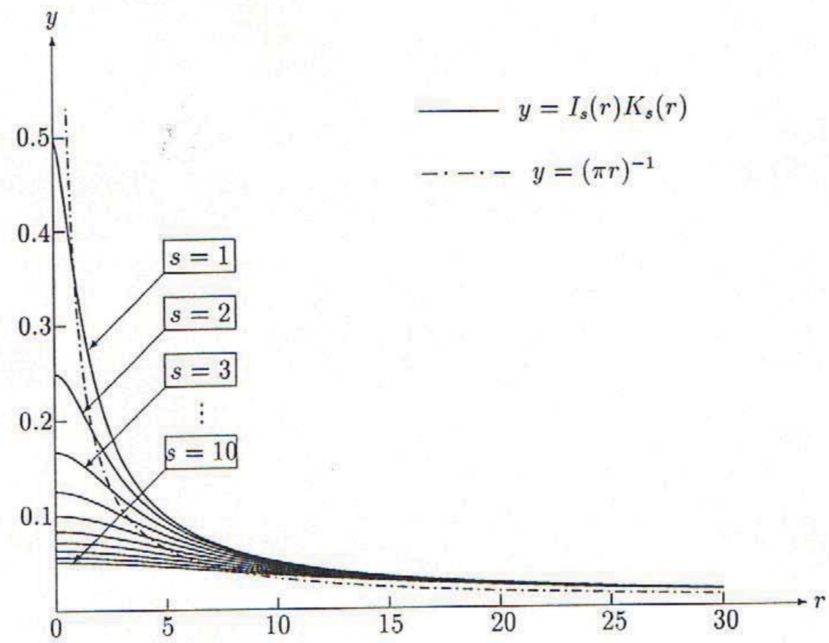


Figure 5.6: Some curves of $I_s(r)K_s(r)$ for various values of zonal wavenumber s.

When $C = \ell_s^2$ exists and is known, the vertical structure of the waves can be determined by solving Eq.(5.3.13). A solution satisfying the boundary condition (5.3.17) has the form

$$\hat{\Psi}_{\ell_s}(z) = A_{\ell_s} \exp[(im - n)z], \tag{5.3.57}$$

where A_{ℓ_s} is a constant to be determined and

$$m = \pm\mathcal{H}(\varpi)\sqrt{\varpi}, \quad \text{where } \varpi = (\mathcal{B}\ell_s^2 - n_H{}^2), \tag{5.3.58}$$

$$n = \mathcal{H}(\varpi')\sqrt{\varpi'}, \quad \text{where } \varpi' = -(\mathcal{B}\ell_s^2 - n_H{}^2); \tag{5.3.59}$$

with the step function $\mathcal{H}$ defined by Eq.(5.3.5). Two possibilities can be seen from Eqs.(5.3.58) and (5.3.59), according as the term $(\mathcal{B}\ell_s^2 - n_H{}^2)$ is positive or negative. For $(\mathcal{B}\ell_s^2 - n_H{}^2) < 0$, no oscillation in z is possible. This is a kind of vertically diffracting (or trapped) wave without phase tilt with height. In the real atmosphere, the waves of this kind cannot penetrate into the middle and upper stratosphere, but may be important in the troposphere and the lower stratosphere. On the other hand, for $(\mathcal{B}\ell_s^2 - n_H{}^2) > 0$, waves can propagate vertically and their phase lines tilt with height. These vertically

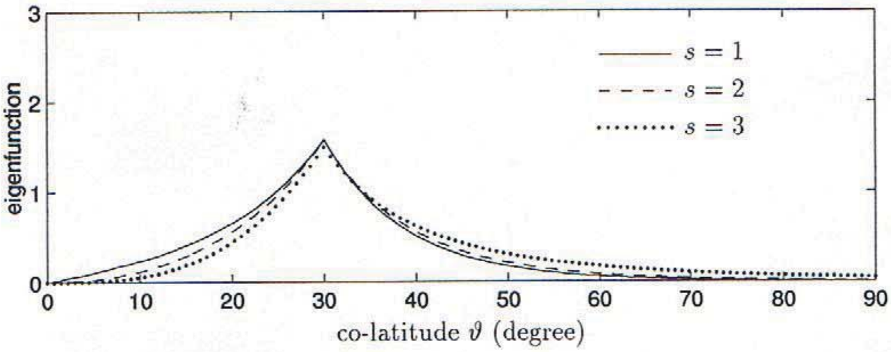


Figure 5.7: Horizontal structure of the topographically forced Rossby waves associated with the polar vortex in the sharp-edge model.

Table 5.1: *Eigenvalues and some associated parameters for forced Rossby waves in the sharp-edge model*

s	C	ℓ_s	$\mathcal{B}\ell_s^2 - {n_H}^2$	m	n
1	6.8418	2.6157	0.6257	0.7910	0
2	4.2228	2.0549	0.2905	0.5390	0
3	0.2820	0.5310	-0.2139	0	0.4625

propagating Rossby waves may play a significant role in the dynamics of the stratosphere. It is generally believed that planetary-scale Rossby waves normally propagate upward from the troposphere into the stratosphere (Andrews *et al.*, 1987). This requires that the waves transfer energy upward. In other words, waves must satisfy the radiation condition. It can be shown that an upward propagating Rossby wave requires a positive m, so the upper sign must be chosen in Eq.(5.3.58). This implies that the phase lines of vertically propagating Rossby waves tilt westward with height.

With the parameters given in Eq.(5.2.22), m and n for waves with $s = 1, 2, 3$ are tabulated in Table 5.1. We see that wave 1 and wave 2 propagate vertically, while wave 3 decays with height. This is in broad agreement with observations, which show that only planetary-scale Rossby waves survive in the winter stratosphere (Andrews *et al.*, 1987).

It remains to determine the constant A_{ℓ_s} in Eq.(5.3.57). Now the lower boundary condition (5.3.40) must be used. For simplicity we consider a special topography

$$\widehat{h}(r) = \widehat{\Theta}_{\ell_s}(r). \tag{5.3.60}$$

Substituting Eqs.(5.3.57) and (5.3.60) into Eq.(5.3.40), yields

$$A_{\ell_s} = -\frac{a}{im - (n + n_\kappa - n_H)}. \tag{5.3.61}$$

5.4 The effect of weak dissipation on the forced interfacial Rossby waves

In this section the effect of weak dissipation on forced interfacial Rossby waves is examined. For simplicity we assume that the waves in question are weakly dissipated by Rayleigh friction and Newtonian cooling, namely

$$X'_{\mathrm{g}} = -\mu\alpha_{\mathrm{R}} u'_{\mathrm{g}}, \quad Y'_{\mathrm{g}} = -\mu\alpha_{\mathrm{R}} v'_{\mathrm{g}}, \quad J'_{\mathrm{g}} = -\mu\alpha_{\mathrm{N}}\frac{\partial\Phi'_{\mathrm{g}}}{\partial z}, \quad \text{with } \mu \ll 1. \tag{5.4.1}$$

Substituting Eq.(5.4.1) into Eq.(5.3.7) and seeking steady, stationary solutions of the form

$$\Phi'_{\mathrm{g}} = \mathrm{Re}\left\{\widehat{\Theta}_\dagger(r)\widehat{\Psi}_\dagger(z)\exp(is\lambda + n_H z)\right\} \tag{5.4.2}$$

we obtain the equation for horizontal structure of the topographically forced, dissipating Rossby wave as

$$\left(1 - \frac{i\mu\alpha_{\mathrm{R}} r}{s\overline{U}(r)}\right)\left[\frac{d}{dr}\left(r\frac{d\widehat{\Theta}_\dagger}{dr}\right) - \frac{s^2}{r}\widehat{\Theta}_\dagger\right] + \left[\frac{r\Delta\overline{q}\,\delta_{r_e}(r)}{\overline{U}(r)} - C_\dagger r\left(1 - \frac{i\mu\alpha_{\mathrm{N}} r}{s\overline{U}(r)}\right)\right]\widehat{\Theta}_\dagger = 0, \tag{5.4.3}$$

where $C_\dagger$ is the separation constant. Comparing this equation with (5.3.39), we see that the effects of dissipation are represented by those terms which are proportional to $i\mu$ in Eq.(5.4.3). The vertical structure equation and the boundary conditions remain the same as those represented by Eqs.(5.3.13), (5.3.15), (5.3.17) and (5.3.40), respectively, with C being replaced by $C_\dagger$. For simplicity we have neglected the diabatic effect on the lower boundary, arising from the term on the right-hand side of Eq.(5.2.30).

If $\mu = 0$, Eq.(5.4.3) reduces simply to Eq.(5.3.39), and the eigenvalue problem has been solved in §5.3.3. For a small but nonzero μ, we can seek approximate solutions to $\widehat{\Theta}_\dagger$ and $C_\dagger$ by letting

$$C_\dagger = C_0 + i\mu C_1 + \mu^2 C_2 + \cdots, \tag{5.4.4}$$

$$\widehat{\Theta}_\dagger(r) = \widehat{\Theta}_0(r) + i\mu\widehat{\Theta}_1(r) + \mu^2\widehat{\Theta}_2(r) + \cdots, \tag{5.4.5}$$

where $C_0, C_1, C_2, \cdots$; $\widehat{\Theta}_0, \widehat{\Theta}_1, \widehat{\Theta}_2, \cdots$ are independent of the small parameter μ. Substituting Eqs.(5.4.4) and (5.4.5) into Eq.(5.4.3) and equating coefficients of like powers of μ, we obtain

$$\text{at } \mu^0: \quad \frac{d}{dr}\left(r\frac{d\widehat{\Theta}_0}{dr}\right) + \left(\frac{r\Delta\overline{q}\,\delta_{r_e}(r)}{\overline{U}(r)} - C_0\,r - \frac{s^2}{r}\right)\widehat{\Theta}_0 = 0; \tag{5.4.6}$$

$$\begin{aligned} \text{at } \mu^1: \quad & \frac{d}{dr}\left(r\frac{d\widehat{\Theta}_1}{dr}\right) + \left(\frac{r\Delta\overline{q}\,\delta_{r_e}(r)}{\overline{U}(r)} - C_0\,r - \frac{s^2}{r}\right)\widehat{\Theta}_1 \\ & = r\left\{\frac{\alpha_{\rm R}}{s\overline{U}(r)}\left[\frac{d}{dr}\left(r\frac{d\widehat{\Theta}_0}{dr}\right) - \frac{s^2}{r}\widehat{\Theta}_0\right] + \left[C_1 - \frac{\alpha_{\rm N}C_0 r}{s\overline{U}(r)}\right]\widehat{\Theta}_0\right\}. \end{aligned} \tag{5.4.7}$$

Eq.(5.4.6) is identical to Eq.(5.3.39). This means that the leading order eigenvalue and eigenfunction for an interfacial Rossby wave associated with the polar vortex are, respectively,

$$C_0 = \ell_s^2, \qquad \widehat{\Theta}_0(r) = \widehat{\Theta}_{\ell_s}(r), \tag{5.4.8}$$

where ℓ_s can be determined from Eq.(5.3.53) and $\widehat{\Theta}_{\ell_s}$ is given by Eq.(5.3.51). Substituting Eq.(5.4.8) into Eq.(5.4.7), we have

$$\begin{aligned} & \frac{d}{dr}\left(r\frac{d\widehat{\Theta}_1}{dr}\right) + \left(\frac{r\Delta\overline{q}\,\delta_{r_e}(r)}{\overline{U}(r)} - \ell_s^2 r - \frac{s^2}{r}\right)\widehat{\Theta}_1 \\ & = r\left\{\frac{\alpha_{\rm R}}{s\overline{U}(r)}\left[\frac{d}{dr}\left(r\frac{d\widehat{\Theta}_{\ell_s}}{dr}\right) - \frac{s^2}{r}\widehat{\Theta}_{\ell_s}\right] + \left[C_1 - \frac{\alpha_{\rm N}\ell_s^2 r}{s\overline{U}(r)}\right]\widehat{\Theta}_{\ell_s}\right\} \\ & = r\left(C_1 - \frac{D_1 r}{\overline{U}(r)} - \frac{\alpha_{\rm R} r\Delta\overline{q}\,\delta_{r_e}(r)}{s\overline{U}^2(r)}\right)\widehat{\Theta}_{\ell_s}(r), \end{aligned} \tag{5.4.9}$$

where

$$D_1 = \ell_s^2(\alpha_{\rm N} - \alpha_{\rm R})/s. \tag{5.4.10}$$

Away from $r = r_e$, (5.4.9) is a non-homogeneous Bessel equation. Using the method of variation of parameters, we obtain its solution in the form

$$\widehat{\Theta}_1(r) = \begin{cases} B_{\ell_s}K_s(\ell_s r_e)\Big\{\big[e_1 + \mathcal{N}_{11}(r)\big]I_s(\ell_s r) - \mathcal{N}_{12}(r)K_s(\ell_s r)\Big\}, & \text{if } r \le r_e \\ B_{\ell_s}I_s(\ell_s r_e)\Big\{\big[e_2 - \mathcal{N}_{21}(r)\big]K_s(\ell_s r) - \mathcal{N}_{22}(r)I_s(\ell_s r)\Big\}, & \text{if } r > r_e, \end{cases} \tag{5.4.11}$$

where e_1, e_2 are integral constants and

$$\mathcal{N}_{11}(r) = C_1\int_0^r \tilde{r}I_s(\ell_s\tilde{r})K_s(\ell_s\tilde{r})d\tilde{r} - D_1\int_0^r \frac{\tilde{r}^2 I_s(\ell_s\tilde{r})K_s(\ell_s\tilde{r})}{\overline{U}(\tilde{r})}d\tilde{r}, \tag{5.4.12}$$

$$\mathcal{N}_{12}(r) = C_1\int_0^r \tilde{r}I_s^2(\ell_s\tilde{r})d\tilde{r} - D_1\int_0^r \frac{\tilde{r}^2 I_s^2(\ell_s\tilde{r})}{\overline{U}(\tilde{r})}d\tilde{r}, \tag{5.4.13}$$

$$\mathcal{N}_{21}(r) = C_1\int_{r_e}^r \tilde{r}I_s(\ell_s\tilde{r})K_s(\ell_s\tilde{r})d\tilde{r} - D_1\int_{r_e}^r \frac{\tilde{r}^2 I_s(\ell_s\tilde{r})K_s(\ell_s\tilde{r})}{\overline{U}(\tilde{r})}d\tilde{r}, \tag{5.4.14}$$

$$\mathcal{N}_{22}(r) = C_1\int_r^\infty \tilde{r}K_s^2(\ell_s\tilde{r})dr - D_1\int_r^\infty \frac{\tilde{r}^2 K_s^2(\ell_s\tilde{r})}{\overline{U}(\tilde{r})}d\tilde{r}. \tag{5.4.15}$$

Table 5.2: C_1 *defined in Eq.(5.4.19) for* $s = 1, 2, 3$ *and various combinations of* α_N *and* α_R

	$\alpha_N = \alpha_R = 0.5$	$\alpha_N = 1, \alpha_R = 0$	$\alpha_N = 0, \alpha_R = 1$
$s = 1$	9.6602	18.8033	0.5171
$s = 2$	4.8025	6.3707	3.2342
$s = 3$	2.8956	0.3358	5.4554

Since the solution (5.4.11) should be continuous at $r = r_e$, as implied by Eq.(5.4.9) itself, e_1 and e_2 must satisfy the following relation,

$$e_2 = e_1 + \mathcal{N}_{11}(r_e) + \frac{\mathcal{N}_{22}(r_e)I_s^2(\ell_s r_e) - \mathcal{N}_{12}(r_e)K_s^2(\ell_s r_e)}{I_s(\ell_s r_e)K_s(\ell_s r_e)}, \tag{5.4.16}$$

with e_1 being arbitrary. Note that the solution given by Eq.(5.4.11) does satisfy the boundary conditions (5.3.15). This is guaranteed by the following asymptotic representations

$$\mathcal{N}_{12}(r)K_s(\ell_s r) \sim \left(C_1 - \frac{\overline{q}_1 D_1}{2}\right) I_s^2(\ell_s r)K_s(\ell_s r), \qquad \text{for small } r, \tag{5.4.17}$$

$$\mathcal{N}_{22}(r)I_s(\ell_s r) \sim \frac{C_1 r^2 I_s(\ell_s r)}{2\ell_s}\left\{\left(\frac{d}{dr}\left[K_s(\ell_s r)\right]\right)^2 - \ell_s K_s^2(\ell_s r)\right\} \quad \text{for large } r, \tag{5.4.18}$$

which tend to zero as r tends to zero and infinity, respectively. Eqs.(5.4.17) and (5.4.18) can be easily verified from substitution of Eq.(5.3.4) and use of Eqs.(5.3.29) and (5.3.30).

Now C_1 can be determined by substituting Eqs.(5.4.11) and (5.3.51) into Eq.(5.4.9) and then integrating term by term from $r = r_e-$ to $r = r_e+$. After some manipulation we obtain

$$C_1 = \frac{\alpha_R + {\ell_s}^2\Delta\overline{q}(\alpha_N - \alpha_R)\left[e_3 K_s^2(\ell_s r_e) + e_4 I_s^2(\ell_s r_e)\right]}{s\Delta\overline{q}\left[e_5 K_s^2(\ell_s r_e) + e_6 I_s^2(\ell_s r_e)\right]}, \tag{5.4.19}$$

where

$$e_3 = \int_0^{r_e} \frac{\tilde{r}^2 I_s^2(\ell_s\tilde{r})}{\overline{U}(\tilde{r})}d\tilde{r}, \qquad e_4 = \int_{r_e}^{\infty} \frac{\tilde{r}^2 K_s^2(\ell_s\tilde{r})}{\overline{U}(\tilde{r})}d\tilde{r}, \tag{5.4.20}$$

$$e_5 = \int_0^{r_e} \tilde{r} I_s^2(\ell_s\tilde{r})d\tilde{r}, \qquad e_6 = \int_{r_e}^{\infty} \tilde{r} K_s^2(\ell_s\tilde{r})d\tilde{r}. \tag{5.4.21}$$

Table 5.2 lists some values of the eigenvalues C_1 for $s = 1, 2, 3$ and for various combinations of α_N and α_R. The corresponding eigenfunctions $\widehat{\Theta}_1(r)$ are plotted in Fig.5.8. We see that $\widehat{\Theta}_1(r)$ also decays away from the vortex edge.

To the first order in μ, therefore, the eigenvalue and eigenfunction for the problem can be written as

$$C_\dagger = \ell_s^2 + i\mu C_1 + O(\mu^2), \tag{5.4.22}$$

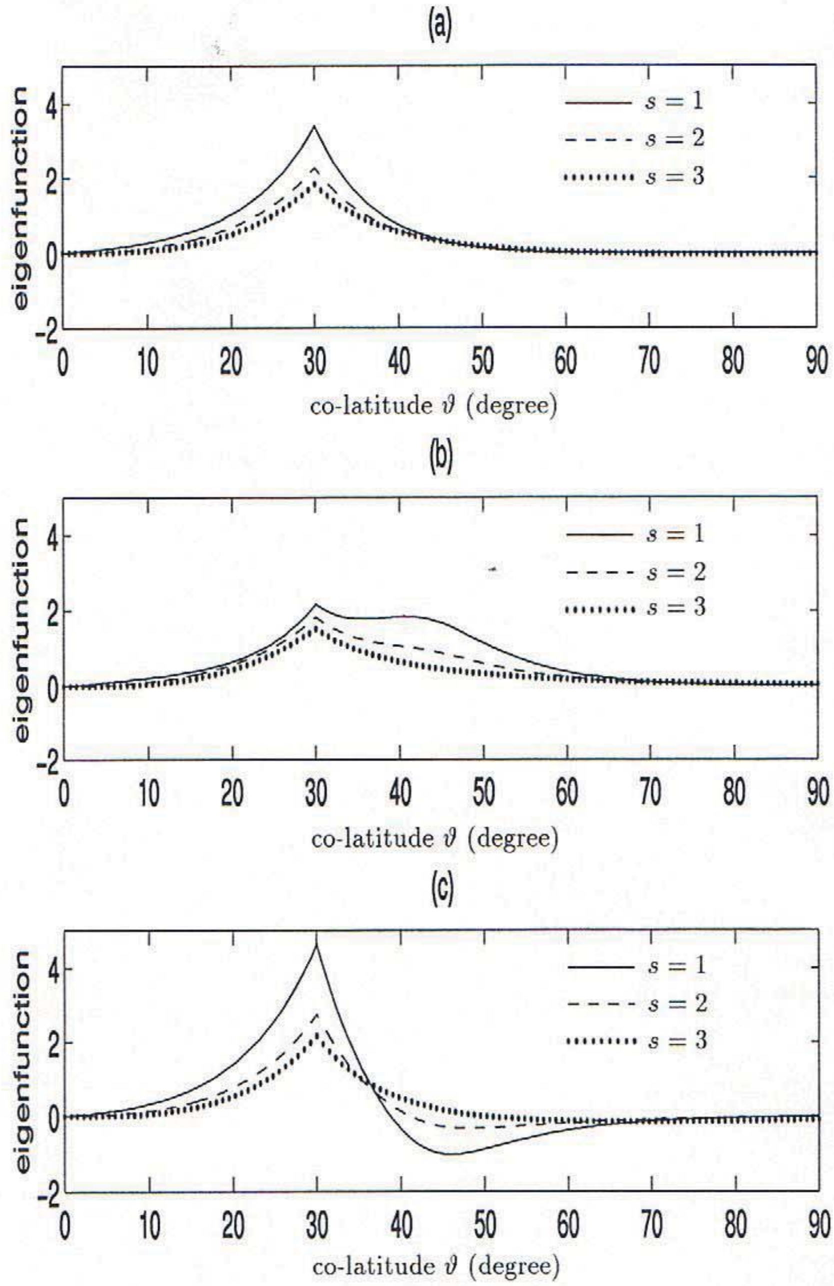


Figure 5.8: Eigenfunctions $\widehat{\Theta}_1(r)$ for various combinations of $\alpha_{\rm N}$ and $\alpha_{\rm R}$. (a) $\alpha_{\rm N} = \alpha_{\rm R} = 0.5$; (b) $\alpha_{\rm N} = 1,\ \alpha_{\rm R} = 0$; (c) $\alpha_{\rm N} = 0,\ \alpha_{\rm R} = 1$.

$$\widehat{\Theta}_\dagger(r) = \widehat{\Theta}_{\ell_s}(r) + i\mu\widehat{\Theta}_1(r) + O(\mu^2), \tag{5.4.23}$$

where C_1 is defined by Eq.(5.4.19) and $\widehat{\Theta}_1(r)$ by Eq.(5.4.11), respectively.

Substituting Eq.(5.4.22) into Eq.(5.3.13) and solving the resulting equation, we obtain the corresponding vertical structure function $\widehat{\Psi}_\dagger(z)$ as

$$\widehat{\Psi}_\dagger(z) = A_\dagger \exp[(im_\dagger - n_\dagger)z], \tag{5.4.24}$$

where $m_\dagger$ and $n_\dagger$ are real numbers determined by the following equations

$$m_\dagger{}^2 - n_\dagger{}^2 = \mathcal{B}\ell_s^2 - n_H{}^2, \tag{5.4.25}$$

$$2m_\dagger n_\dagger = \mu\mathcal{B}C_1, \tag{5.4.26}$$

and, for the special topography represented by Eq.(5.3.60), $A_\dagger$ is given by

$$A_\dagger = -\frac{a}{im_\dagger - (n_\dagger + n_\kappa - n_H)}. \tag{5.4.27}$$

The upper boundary condition (5.3.17) requires $n_\dagger \geq 0$. With this constraint $n_\dagger$ and $m_\dagger$ can be solved from Eqs.(5.4.25) and (5.4.26) as

$$n_\dagger = \sqrt{\frac{1}{2}\left[\sqrt{(\mathcal{B}\ell_s^2 - n_H{}^2)^2 + (\mu\mathcal{B}C_1)^2} - (\mathcal{B}\ell_s^2 - n_H{}^2)\right]}, \tag{5.4.28}$$

$$m_\dagger = \operatorname{sgn}(\mu C_1)\sqrt{\frac{1}{2}\left[\sqrt{(\mathcal{B}\ell_s^2 - n_H{}^2)^2 + (\mu\mathcal{B}C_1)^2} + (\mathcal{B}\ell_s^2 - n_H{}^2)\right]}. \tag{5.4.29}$$

Note that, for $\mu \to 0$, $m_\dagger \to m$ and $n_\dagger \to n$, where m and n are defined respectively by Eqs.(5.3.58) and (5.3.59), with m being positive for wave energy propagating upward.

For $\mu \ll 1$ and $(\mathcal{B}\ell_s^2 - n_H{}^2) \neq 0$, Eqs.(5.4.28) and (5.4.29) can be expanded in Taylor series as,

$$n_\dagger = n + \mu n_1 + O(\mu^2), \tag{5.4.30}$$

$$m_\dagger = \operatorname{sgn}(\mu C_1)\left[m + \mu m_1 + O(\mu^2)\right], \tag{5.4.31}$$

where

$$n_1 = \mathcal{H}(\varpi)\frac{\mathcal{B}|C_1|}{2m}, \quad \text{where } \varpi = (\mathcal{B}\ell_s^2 - n_H{}^2), \tag{5.4.32}$$

$$m_1 = \mathcal{H}(\varpi')\frac{\mathcal{B}|C_1|}{2n}, \quad \text{where } \varpi' = -(\mathcal{B}\ell_s^2 - n_H{}^2). \tag{5.4.33}$$

From Eqs.(5.3.45), (5.3.46), (5.4.30)–(5.4.33) it is obvious that $n_\dagger = O(\mu)$ and $m_\dagger = O(1)$ if $(\mathcal{B}\ell_s^2 - n_H{}^2) > 0$, otherwise $n_\dagger = O(1)$ and $m_\dagger = O(\mu)$.

Substituting Eqs.(5.4.30) and (5.4.31) into (5.4.27), after some manipulation we obtain

$$A_\dagger = A_{\ell_s}\left(1 - \mu\frac{im_1 - n_1}{im - n + n_H} + \cdots\right), \tag{5.4.34}$$

where A_{ℓ_s} is defined by Eq.(5.3.61).

5.5 The quasi-geostrophic EP flux on the γ-plane

The TEM equations and EP flux divergences on the mid-latitude β-plane were discussed in Section 4.4. Their γ-plane versions are outlined in this section.

5.5.1 The TEM equations and the generalized Eliassen–Palm theorem

First, a residual mean meridional circulation is defined by

$$\overline{v}_{\rm a}^{*} = \overline{v}_{\rm a} - \frac{1}{\mathcal{B}\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\overline{v'_{\rm g}\frac{\partial \Phi'_{\rm g}}{\partial z}}\right), \qquad \overline{w}_{\rm a}^{*} = \overline{w}_{\rm a} + \frac{1}{\mathcal{B}\,r}\frac{\partial}{\partial r}\left(\overline{r v'_{\rm g}\frac{\partial \Phi'_{\rm g}}{\partial z}}\right). \tag{5.5.1}$$

Then the TEM equations take the following forms

$$\frac{\partial \overline{u}_{\rm g}}{\partial t} + \overline{v}_{\rm a}^{*} = \frac{1}{\rho_{\rm B} r}\nabla\cdot\mathbf{S}_{\rm g} + \overline{X}_{\rm g}, \tag{5.5.2}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \overline{\Phi}_{\rm g}}{\partial z}\right) + \mathcal{B}\overline{w}_{\rm a}^{*} = \overline{J}_{\rm g}, \tag{5.5.3}$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\overline{v}_{\rm a}^{*}\right) + \frac{1}{\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\overline{w}_{\rm a}^{*}\right) = 0, \tag{5.5.4}$$

where the quasi-geostrophic EP flux $\mathbf{S}_{\rm g}$ is defined by,

$$\mathbf{S}_{\rm g} = \rho_{\rm B}\left(0,\ -r\overline{v'_{\rm g}u'_{\rm g}},\ -\mathcal{B}^{-1}r\overline{v'_{\rm g}\partial\Phi'_{\rm g}/\partial z}\right). \tag{5.5.5}$$

The term $(\rho_{\rm B} r)^{-1}\,\nabla\cdot\mathbf{S}_{\rm g}$ in Eq.(5.5.2) represents the net driving of the mean zonal flow by Rossby waves. It combines the eddy momentum flux $\overline{v'_{\rm g}u'_{\rm g}}$ and eddy heat flux $\overline{v'_{\rm g}\partial\Phi'_{\rm g}/\partial z}$ in the form

$$\frac{1}{\rho_{\rm B} r}\nabla\cdot\mathbf{S}_{\rm g} \equiv -\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\overline{v'_{\rm g}u'_{\rm g}}\right) - \frac{1}{\mathcal{B}\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\overline{v'_{\rm g}\frac{\partial \Phi'_{\rm g}}{\partial z}}\right). \tag{5.5.6}$$

$\overline{X}_{\rm g}$ represents all further contributions to the mean zonal force per unit mass associated with gravity waves and other small-scale disturbances, and $\overline{J}_{\rm g}$ is the zonal-mean net radiative heating rate.

From Eqs.(5.2.24) and (5.5.6) it follows that

$$\begin{aligned}
\overline{v'_{\rm g}q'_{\rm g}} &= \overline{v'_{\rm g}\left\{\frac{1}{r}\left[\frac{\partial}{\partial r}\left(r u'_{\rm g}\right) + \frac{\partial v'_{\rm g}}{\partial \lambda}\right] + \frac{1}{\mathcal{B}\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\frac{\partial \Phi'_{\rm g}}{\partial z}\right)\right\}} \\
&= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\overline{u'_{\rm g}v'_{\rm g}}\right) + \overline{\frac{u'_{\rm g}}{r}\frac{\partial}{\partial r}\left(r v'_{\rm g}\right)} + \frac{1}{\mathcal{B}\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\overline{v'_{\rm g}\frac{\partial \Phi'_{\rm g}}{\partial z}}\right) - \frac{1}{\mathcal{B}}\overline{\frac{\partial \Phi'_{\rm g}}{\partial z}\frac{\partial v'_{\rm g}}{\partial z}} \\
&= -\frac{1}{\rho_{\rm B} r}\nabla\cdot\mathbf{S}_{\rm g} - \frac{1}{2r}\frac{\partial}{\partial \lambda}\overline{\left[u_{\rm g}'^{\,2} - \frac{1}{\mathcal{B}}\left(\frac{\partial \Phi'_{\rm g}}{\partial z}\right)^2\right]} = -\frac{1}{\rho_{\rm B} r}\nabla\cdot\mathbf{S}_{\rm g}.
\end{aligned} \tag{5.5.7}$$

Note that the minus sign in the last expression occurs because a positive v'_g corresponds to a southward motion.

When the basic QGPV gradient $\overline{q}_r$ is smooth and non-zero everywhere, the quasi-geostrophic version of the generalized EP theorem on the γ-plane can be written as,

$$\overline{v'_g q'_g} = -\frac{1}{\rho_B r}\nabla\cdot\mathbf{S}_g = -\frac{\partial}{\partial t}\left(\tfrac{1}{2}\,\overline{q'^2_g}\big/\,\overline{q}_r\right) + \overline{Z'_g q'_g}\big/\,\overline{q}_r. \tag{5.5.8}$$

Alternatively, it can be shown that

$$\overline{v'_g q'_g} = -\frac{1}{\rho_B r}\nabla\cdot\mathbf{S}_g = \frac{\partial}{\partial t}\left(\overline{\eta'_g q'_g} + \tfrac{1}{2}\overline{\eta'^2_g}\,\overline{q}_r\right) - \overline{\eta'_g Z'_g}, \tag{5.5.9}$$

where η'_g is the southward parcel displacement, defined by

$$\frac{\overline{D}\eta'_g}{Dt} = v'_g. \tag{5.5.10}$$

Note that Eq.(5.5.9) holds even when $\overline{q}_r = 0$. Thus, it is more general than Eq.(5.5.8) and specially useful in the sharp-edge model described in Section 5.3.

5.5.2 EP flux divergences induced by dissipating Rossby waves

In a Charney–Drazin model on the mid-latitude β-plane, it has been demonstrated (§4.4.3) that when the meridional gradient of mean QGPV is northward everywhere, the quasi-geostrophic EP flux divergence is always negative if the steady Rossby waves are dissipated by Rayleigh friction and Newtonian cooling with equal relaxation coefficients, and it may be positive in situations in which either the meridional gradient of mean QGPV is southward or other types of dissipation are present. In particular, when the Rossby waves are dissipated purely thermally, the quasi-geostrophic EP flux divergence is positive if $\overline{U} > u_c$, where $\overline{U}$ is the mean zonal flow and u_c is the Rossby critical velocity defined by Eq.(4.3.20).

In this section, we examine the EP flux divergence on the γ-plane. Let us start with the simple case in which the Rossby waves are dissipated by Rayleigh friction and Newtonian cooling with equal relaxation coefficients, namely $\alpha_R = \alpha_N = \alpha \neq 0$. Under such circumstance, it can be shown that

$$Z'_g = -\mu\alpha q'_g. \tag{5.5.11}$$

This is because in the quasi-geostrophic equations, $\overline{D}/Dt$ and $\mu\alpha$ always occur in combination $(\overline{D}/Dt + \mu\alpha)$ when $\alpha_{\rm R} = \alpha_{\rm N} = \alpha$; therefore

$$\left(\frac{\overline{D}}{Dt} + \mu\alpha\right) q'_{\rm g} + v'_{\rm g}\overline{q}_r = 0 \tag{5.5.12}$$

follows from the conservative case

$$\frac{\overline{D}q'_{\rm g}}{Dt} + v'_{\rm g}\overline{q}_r = 0. \tag{5.5.13}$$

When waves are steady and $\overline{q}_r \neq 0$ everywhere, from Eqs.(5.5.8) and (5.5.11) it follows that

$$\overline{v'_{\rm g}q'_{\rm g}} = -\frac{1}{\rho_{\rm B}r}\nabla\cdot\mathbf{S}_{\rm g} = -\mu\alpha\,\overline{q'^2_{\rm g}}\big/\,\overline{q}_r. \tag{5.5.14}$$

It shows that any basic state with smooth, negative (northward) $\overline{q}_r$ must have $\nabla\cdot\mathbf{S}_{\rm g} < 0$ when Rossby waves are assumed steady and dissipated by Rayleigh friction and Newtonian cooling with $\alpha_{\rm R} = \alpha_{\rm N} = \alpha \neq 0$. This result has been pointed out by Holton (1983a), Robinson (1986), Andrews (1987), and in §4.4.3.

We now turn our attention to the sharp-edge model, in which $\overline{q}_r$ is assumed zero everywhere except at the edge of the polar vortex ($r = r_{\rm e}$). Note first that, away from $r = r_{\rm e}$, Eq.(5.5.13) reduces to a hyperbolic equation because $\overline{q}_r = 0$. Thus, for conservative waves, $q'_{\rm g}$ is identically zero if it is so at some initial instant. This can be shown explicitly by substituting the conservative wave solution obtained in Section 5.3 into Eq.(5.3.9), namely,

$$\begin{aligned} q'_{\rm g} &= \underbrace{\frac{1}{r}\left[\frac{\partial}{\partial r}\left(ru'_{\rm g}\right) - \frac{\partial v'_{\rm g}}{\partial\lambda}\right]}_{\rm RV} + \underbrace{\frac{1}{\mathcal{B}\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\frac{\partial\Phi'_{\rm g}}{\partial z}\right)}_{\rm SS} \\ &= \underbrace{\ell_s^2\Phi'_{\rm g}}_{\rm RV} - \underbrace{\ell_s^2\Phi'_{\rm g}}_{\rm SS} = 0, \qquad (r \neq r_{\rm e}). \end{aligned} \tag{5.5.15}$$

In Eq.(5.5.15) we see that, away from the edge of polar vortex, the relative vorticity (RV) term takes an opposite sign to the static stability (SS) term. For conservative wave motion, these two terms cancel each other exactly, leading to $q'_{\rm g} \equiv 0$.

For the steady, dissipative Rossby waves, Eq.(5.5.8) shows that

$$\overline{Z'_{\rm g}q'_{\rm g}} = 0, \qquad (r \neq r_{\rm e}), \tag{5.5.16}$$

because $\overline{q}_r = 0$ away from the vortex edge. If the waves are assumed to be dissipated by Rayleigh friction and Newtonian cooling with $\alpha_{\rm R} = \alpha_{\rm N} = \alpha \neq 0$, we can substitute Eq.(5.5.11) into Eq.(5.5.16) to show that

$$-\mu\alpha\overline{q_{\rm g}'^{\,2}} = 0, \qquad (r \neq r_{\rm e}). \tag{5.5.17}$$

Eq.(5.5.17) implies that $q_{\rm g}'$ is again zero away from $r = r_{\rm e}$. Therefore the argument illustrated by Eq.(5.5.15) must remain valid for this case. Furthermore, when $q_{\rm g}' = 0$ is substituted into the left-hand side of Eq.(5.5.8), we see that $\nabla\cdot\mathbf{S}_{\rm g} = 0$. This result is in sharp contrast to the analogous problem in the Charney–Drazin model with smooth, northward meridional gradient of mean QGPV; in that case the EP flux divergence is always negative for steady, dissipating Rossby waves with $\alpha_{\rm N} = \alpha_{\rm R} = \alpha \neq 0$ (see the discussion in §4.4.3 and Holton, 1983a; Robinson, 1986; Andrews, 1987).

To see what would happen when $\alpha_{\rm N} \neq \alpha_{\rm R}$, we now calculate $\nabla\cdot\mathbf{S}_{\rm g}$ from Eq.(5.5.9) using the wave solution obtained in Section 5.4. First, from Eqs.(5.5.10) and (5.2.24) it follows that

$$\eta_{\rm g}' = -\Phi_{\rm g}'/\overline{U} + O(a^2). \tag{5.5.18}$$

In addition, $Z_{\rm g}'$ is given by

$$Z_{\rm g}' = -\mu\left\{\frac{\alpha_{\rm R}}{r^2}\left[\frac{\partial}{\partial r}\left(r\frac{\partial\Phi_{\rm g}'}{\partial r}\right) + \frac{\partial^2\Phi_{\rm g}'}{\partial\lambda^2}\right] + \frac{\alpha_{\rm N}}{\mathcal{B}\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\frac{\partial\Phi_{\rm g}'}{\partial z}\right)\right\}. \tag{5.5.19}$$

Substituting solution (5.4.2) into Eq.(5.5.19) gives

$$\begin{aligned} Z_{\rm g}' = -\mu{\rm Re}\Bigg\{\Bigg[\alpha_{\rm R}\left(\ell_s^2 - \frac{\Delta\overline{q}\,\delta_{r_{\rm e}}(r)}{\overline{U}(r)}\right) \\ +\alpha_{\rm N}\frac{(im_\dagger - n_\dagger)^2 - n_H{}^2}{\mathcal{B}}\Bigg]\widehat{\Theta}_{\ell_s}(r)\widehat{\Psi}_{\ell_s}(z)\exp\left[is\lambda + n_H z + \mu(im_1 - n_1)z\right]\Bigg\}, \end{aligned} \tag{5.5.20}$$

correct to $O(\mu a)$. Now, with the aid of Eqs.(5.4.25) and (5.4.26), it can be shown that

$$(im_\dagger - n_\dagger)^2 - n_H{}^2 = n_\dagger{}^2 - m_\dagger{}^2 - n_H{}^2 - i(2m_\dagger n_\dagger) = -\mathcal{B}\ell_s^2 + O(\mu). \tag{5.5.21}$$

Thus, away from the edge of the polar vortex, we see that

$$Z_{\rm g}' = -\mu\ell_s^2(\alpha_{\rm R} - \alpha_{\rm N})\Phi_{\rm g}' + O(\mu^2 a), \qquad (r \neq r_{\rm e}). \tag{5.5.22}$$

Substituting Eqs.(5.5.18) and (5.5.22) into (5.5.9), for steady waves we obtain

$$\nabla\cdot\mathbf{S}_{\rm g} = -\frac{\mu\,(\alpha_{\rm N} - \alpha_{\rm R})\,\ell_s^2 r}{2\overline{U}(r)}|A_{\ell_s}|^2\widehat{\Theta}_{\ell_s}^2(r)\exp\left[-2(n + \mu n_1)z\right] + O(\mu^2 a^2), \qquad (r \neq r_{\rm e}). \tag{5.5.23}$$

Eq.(5.5.23) shows that, away from $r = r_{\rm e}$, $\nabla\cdot\mathbf{S}_{\rm g}$ is zero to the first order in μ if $\alpha_{\rm R} = \alpha_{\rm N}$. In fact, the general argument given above has shown that $\nabla\cdot\mathbf{S}_{\rm g}$ should be zero to any order in μ if $\alpha_{\rm R} = \alpha_{\rm N}$. Since both $\overline{U}(r)$ and $\widehat{\Theta}_{\ell,}(r)$ are continuous at $r = r_{\rm e}$, Eq.(5.5.23) indicates that $\nabla\cdot\mathbf{S}_{\rm g}$ is also continuous there. In addition, $\overline{U}(r)$ is assumed positive in our model. Therefore, we see from Eq.(5.5.23) that $\nabla\cdot\mathbf{S}_{\rm g} < 0$ when $\alpha_{\rm R} < \alpha_{\rm N}$, and *vice versa*. Note that these results are also in sharp contrast to the analogous problem in the Charney–Drazin model with smooth, northward meridional gradients of mean QGPV (see the discussion in §4.4.3).

In the winter stratosphere, sizable regions of divergent EP flux have been documented in a recent paper by Rosenlof and Holton (1993). Maximum positive values can be seen within the polar vortex both in Northern and Southern Hemispheres. The physical reality and robustness of these phenomena are likely to be trustworthy because the EP flux divergences were calculated from a 10-year data set that is based on linear balance wind estimates rather than geostrophic wind estimates. We have shown that the Rossby waves associated with the polar vortex can cause positive EP flux divergences when $\alpha_{\rm R} > \alpha_{\rm N}$. This suggests a possible mechanism for explaining the observed phenomena mentioned above. Observational and model studies have shown that gravity wave drag is important in the lower stratosphere where tilting and overturning of isentropic surfaces can lead to gravity-wave breaking over mountainous regions (e.g., Palmer *et al.*, 1986, & refs. therein). This will strengthen the mechanical dissipation of Rossby waves. According to the above discussion, when the mechanical dissipation is strong enough to dominate the thermal dissipation, interfacial Rossby waves will lead to positive $\nabla\cdot\mathbf{S}_{\rm g}$ both inside and outside the polar vortex. On the other hand, in the wintertime lower stratosphere Rossby wave breaking is a frequent occurrence outside the polar vortex (McIntyre and Palmer, 1983,1984; Juckes and McIntyre, 1987; Juckes, 1989; Waugh *et al.*, 1994). Since breaking Rossby waves induce a downgradient eddy flux of PV, they are likely to contribute to systematically large negative values of $\nabla\cdot\mathbf{S}_{\rm g}$ in the midlatitude surf zone. Outside the polar vortex, therefore, the negative $\nabla\cdot\mathbf{S}_{\rm g}$ due to breaking Rossby waves will submerge the positive $\nabla\cdot\mathbf{S}_{\rm g}$ due to mechanically dissipating, non-breaking Rossby waves associated with the polar vortex. For this reason, significant positive $\nabla\cdot\mathbf{S}_{\rm g}$ may only be observed within the polar vortex in the lower stratosphere. We shall return to this phenomenon in the following section.

5.6 Wave-induced mass transport across the edge of the polar vortex

Planetary-scale Rossby waves in the stratosphere play an important role in the meridional and vertical transport of ozone and other chemical tracers. Noticing the characteristics of Rossby-wave breaking in the winter stratosphere, McIntyre (1989) pointed out that the steep PV gradients on isentropic surfaces on the edge of the polar vortex could act as a flexible PV barrier against poleward eddy transport of mass and chemical tracers, and then the polar vortex might behave as an isolated material entity, or air mass, if the vortex is not disturbed too violently. This vortex isolation hypothesis is believed to be essentially important in the formation of the Antarctic ozone hole (McIntyre and Norton, 1990; McIntyre, 1992).

However, the recently observed depletion in middle- and high-latitude total ozone of both hemispheres (Stolarski *et al.*, 1991, 1992) has raised a controversial question, namely how effective such PV barriers might be in the real stratosphere (e.g., Anderson *et al.*, 1989; Proffitt *et al.*, 1989, 1990; Pierce and Fairlie, 1993). Tuck *et al.*, (1992, 1993) argued that there might be a significant rate of transport of processed air through the polar vortex, taking chlorine-activated air out of the lower-stratospheric polar vortex to lower latitudes where sunlight enables ozone destruction to occur (the so-called 'flowing processor' hypothesis). McIntyre and Pyle (1993) pointed out that this hypothesis required an implausibly high rate of air transport out of the polar vortex, which could not be provided by vortex erosion alone. It might be argued that, since the PV barrier at the edge of the polar vortex inhibits only eddy transport but not mean transport, there might be a mean outflow from the vortex in the lower stratosphere. From a TEM perspective, such an mean outflow might be driven by, for instance, the anomalously-signed (positive) EP flux divergences in the high-latitude lower stratosphere, as documented in Fig.3 in the paper by Rosenlof and Holton (1993).

It should be remembered, however, that besides the quasi-geostrophic EP flux divergence, there is another term $\overline{X}_g$, representing contributions to the mean zonal force from gravity waves and other small-scale disturbances, on the right-hand side of Eq.(5.5.2). Therefore, the TEM circulation is not determined by the structure of the EP flux alone. Furthermore, the actual mass transport is closely related to the GLM circulation, which, in general, is not the same as its TEM counterpart. As pointed out in Section 4.5, the TEM and GLM meridional circulation in the Charney–Drazin model may be oppositelysigned in certain situations. This problem is further investigated in a sharp-edge model in this section.

For simplicity, let us first assume that dissipation of the wave–mean system is purely thermal and the system has settled down as a whole to an exactly steady state in which not only is the wave amplitude steady but also the mean circulation. Under such circumstances, Eqs.(5.5.2), (5.5.3) and (5.5.9) reduce to

$$\overline{v}_{\rm a}^{*} = \frac{1}{\rho_{\rm B} r}\nabla\cdot\mathbf{S}_{\rm g}, \tag{5.6.1}$$

$$\overline{w}_{\rm a}^{*} = \overline{J}_{\rm g}/\mathcal{B}, \tag{5.6.2}$$

$$\nabla\cdot\mathbf{S}_{\rm g} = \rho_{\rm B} r\overline{\eta'_{\rm g}Z'_{\rm g}} = \overline{\frac{r\eta'_{\rm g}}{\mathcal{B}}\frac{\partial}{\partial z}(\rho_{\rm B}J'_{\rm g})}. \tag{5.6.3}$$

Substituting Eq.(5.6.3) into Eq.(5.6.1) yields

$$\overline{v}_{\rm a}^{*} = \overline{\frac{\eta'_{\rm g}}{\mathcal{B}\rho_{\rm B}}\frac{\partial}{\partial z}(\rho_{\rm B}J'_{\rm g})}. \tag{5.6.4}$$

Now Eq.(5.5.4) can be rearranged as

$$\frac{\partial}{\partial z}\left(\rho_{\rm B}\overline{w}_{\rm a}^{*}\right) = -\frac{\rho_{\rm B}}{r}\frac{\partial}{\partial r}\left(r\overline{v}_{\rm a}^{*}\right) = -\frac{1}{\mathcal{B}\,r}\frac{\partial}{\partial r}\left[\overline{r\eta'_{\rm g}\frac{\partial}{\partial z}(\rho_{\rm B}J'_{\rm g})}\right]. \tag{5.6.5}$$

Integrating Eq.(5.6.5) with respect to z gives

$$\begin{aligned}\overline{w}_{\rm a}^{*} &= \frac{1}{\mathcal{B}\rho_{\rm B}\,r}\int_z^{\infty}\frac{\partial}{\partial r}\left[\overline{r\eta'_{\rm g}\frac{\partial}{\partial z}(\rho_{\rm B}J'_{\rm g})}\right]dz \\ &= -\frac{1}{\mathcal{B}\,r}\left[\frac{\partial}{\partial r}\left(r\overline{\eta'_{\rm g}J'_{\rm g}}\right) - \frac{1}{\rho_{\rm B}}\int_z^{\infty}\rho_{\rm B}\frac{\partial}{\partial r}\left(\overline{rJ'_{\rm g}\frac{\partial\eta'_{\rm g}}{\partial z}}\right)dz\right],\end{aligned} \tag{5.6.6}$$

where the boundary condition $\rho_{\rm B}\overline{w}_{\rm a}^{*}\to 0$ as $z\to\infty$ has been used to fix the constant of integration.

It is well-known that the residual mean circulation only provides an estimate of the actual mass circulation, which is a Lagrangian quantity. To examine the GLM circulation, we first note that the γ-plane version of Eq.(4.5.9) can be written as

$$\overline{\varphi}^{\rm S} = \overline{(\boldsymbol{\xi}'_{\rm g}\cdot\nabla)\varphi'} + O(a^2\varphi') = \frac{1}{r}\frac{\partial}{\partial r}\left(r\overline{\eta'_{\rm g}\varphi'}\right) + O(\mathcal{R}a\varphi') + O(a^2\varphi'). \tag{5.6.7}$$

In addition, the GLM continuity equation and steady thermodynamic equation can be written as, respectively:

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\overline{v}_{\rm a}^{\rm L}\right) + \frac{1}{\rho_{\rm B}}\frac{\partial}{\partial z}\left(\rho_{\rm B}\overline{w}_{\rm a}^{\rm L}\right) = 0, \tag{5.6.8}$$

$$\overline{w}_{\rm a}^{\rm L} = \overline{J}_{\rm g}^{\rm L}/\mathcal{B}. \tag{5.6.9}$$

On the other hand, Eq.(5.6.2) can be manipulated as follows,

$$\overline{w}_{\rm a}^{*} = \overline{J}_{\rm g}/\mathcal{B} = \left(\overline{J}_{\rm g}^{\rm L} - \overline{J}_{\rm g}^{\rm S}\right)/\mathcal{B} = \overline{w}_{\rm a}^{\rm L} - \frac{1}{\mathcal{B}\,r}\frac{\partial}{\partial r}\left(r\overline{\eta_{\rm g}' J_{\rm g}'}\right) + O(\mathcal{R}\mu a^2) + O(\mu a^3), \quad (5.6.10)$$

where Eq.(5.6.9) has been used and $J_{\rm g}' = O(\mu a)$ has been assumed. Substituting Eq.(5.6.6) into Eq.(5.6.10), together with use of Eq.(5.6.8), we obtain the GLM vertical and meridional velocities, respectively,

$$\begin{aligned} \overline{w}_{\rm a}^{\rm L} &= \overline{w}_{\rm a}^{*} + \frac{1}{\mathcal{B}\,r}\frac{\partial}{\partial r}\left(r\overline{\eta_{\rm g}' J_{\rm g}'}\right) + O(\mathcal{R}\mu a^2) + O(\mu a^3) \\ &= -\frac{1}{\mathcal{B}\rho_{\rm B} r}\int_z^{\infty} \rho_{\rm B}\frac{\partial}{\partial r}\left(r\overline{J_{\rm g}'\frac{\partial \eta_{\rm g}'}{\partial z}}\right) dz + O(\mathcal{R}\mu a^2) + O(\mu a^3), \quad (5.6.11) \\ \overline{v}_{\rm a}^{\rm L} &= -\frac{\overline{J_{\rm g}'\,\partial\eta_{\rm g}'}}{\mathcal{B}\;\partial z} + O(\mathcal{R}\mu a^2) + O(\mu a^3). \quad (5.6.12) \end{aligned}$$

Eqs.(5.6.4), (5.6.6), (5.6.11), and (5.6.12) show that both the TEM and GLM meridional circulations depend explicitly on the diabatic heating. Thus, for steady and conservative waves,

$$\overline{v}_{\rm a}^{\rm L} = \overline{v}_{\rm a}^{*}, \qquad \overline{w}_{\rm a}^{\rm L} = \overline{w}_{\rm a}^{*}, \quad (5.6.13)$$

to $O(a^2)$ (see Dunkerton, 1978). For dissipative waves, the above relationship will not be exactly true. The difference between $\overline{w}_{\rm a}^{\rm L}$ and $\overline{w}_{\rm a}^{*}$ is $\overline{J}_{\rm g}^{\rm S}/\mathcal{B}$, where $\overline{J}_{\rm g}^{\rm S}$ is the Stokes correction to the diabatic heating. Substituting the Newtonian cooling approximation and the wave solutions obtained in Section 5.4 into Eqs.(5.6.4), (5.6.6), (5.6.11), and (5.6.12), after some manipulation we obtain

$$\overline{v}_{\rm a}^{*} = -\frac{\mu\alpha_{\rm N}\ell_s^2|A_{\ell_s}|^2}{2\rho_{\rm B}\overline{U}}\widehat{\Theta}_{\ell_s}^2 \exp\left[-2(n+\mu n_1)z\right], \quad (5.6.14)$$

$$\overline{w}_{\rm a}^{*} = -\frac{\mu\alpha_{\rm N}\ell_s^2|A_{\ell_s}|^2}{4(n+\mu n_1)\rho_{\rm B} r}\frac{\partial}{\partial r}\left(\frac{r\widehat{\Theta}_{\ell_s}^2}{\overline{U}}\right)\exp\left[-2(n+\mu n_1)z\right], \quad (5.6.15)$$

$$\overline{v}_{\rm a}^{\rm L} = -\frac{\mu\alpha_{\rm N}\left[m^2+(n_H-n)^2\right]|A_{\ell_s}|^2}{2\mathcal{B}\rho_{\rm B}\overline{U}}\widehat{\Theta}_{\ell_s}^2 \exp\left[-2(n+\mu n_1)z\right] + O(\mu^2 a^2, \mathcal{R}\mu a^2, \mu a^3), \quad (5.6.16)$$

$$\overline{w}_{\rm a}^{\rm L} = -\frac{\mu\alpha_{\rm N}\left[m^2+(n_H-n)^2\right]|A_{\ell_s}|^2}{4(n+\mu n_1)\mathcal{B}\rho_{\rm B} r}\frac{\partial}{\partial r}\left(\frac{r\widehat{\Theta}_{\ell_s}^2}{\overline{U}}\right)\exp\left[-2(n+\mu n_1)z\right] + O(\mu^2 a^2, \mathcal{R}\mu a^2, \mu a^3). \quad (5.6.17)$$

The above equations indicate that the TEM and GLM circulations induced by steady, thermally dissipating interfacial Rossby waves have the same vertical and horizontal structures. Both $\overline{v}_{\rm a}^{*}$ and $\overline{v}_{\rm a}^{\rm L}$ are $O(\mu a^2)$ quantities. They are negative because $\overline{U}(r)$ is assumed positive. This means that the mass transport is poleward rather than equatorward.

The difference between $\overline{v}^{\mathrm{L}}_{\mathrm{a}}$ and $\overline{v}^{*}_{\mathrm{a}}$ can be obtained as

$$\overline{v}^{\mathrm{L}}_{\mathrm{a}} - \overline{v}^{*}_{\mathrm{a}} = \frac{\mu\alpha_{\mathrm{N}} n(n_H - n)|A_{\ell_s}|^2}{\mathcal{B}\rho_{\mathrm{B}}\overline{U}(r)} \widehat{\Theta}^{2}_{\ell_s} \exp\left[-2(n+\mu n_1)z\right] + O(\mu^2 a^2, \mathcal{R}\mu a^2, \mu a^3), \quad (5.6.18)$$

which is $O(\mu^2 a^2, \mathcal{R}\mu a^2, \mu a^3)$ for vertically propagating waves ($n = 0,\ n_1 \neq 0$) and $O(\mu a^2)$ for vertically diffracting waves ($n \neq 0,\ n_1 = 0$). The latter has the same order as $\overline{v}^{*}_{\mathrm{a}}$ and $\overline{v}^{\mathrm{L}}_{\mathrm{a}}$ themselves, so that in general it is not negligible. In addition, for vertically diffracting interfacial Rossby waves, Eq.(5.3.59) shows that $0 < n < n_H$. In such circumstances the first term on the right-hand side of Eq.(5.6.18) is positive, implying that the poleward GLM flow is weaker than its TEM counterpart.

Note that $\overline{w}^{*}_{\mathrm{a}}$ and $\overline{w}^{\mathrm{L}}_{\mathrm{a}}$ are of $O(a^2)$ for vertically propagating waves and are $O(\mu a^2)$ for vertically diffracting waves. The difference between them is not negligible, either, when the waves are vertically diffracting.

We have seen that thermally dissipating interfacial Rossby waves induce mean mass transport into rather than out of the polar vortex, giving no support to the 'flowing processor' hypothesis (Tuck *et al.*, 1992, 1993), which requires a significant transport of chemically perturbed air from within the polar vortex to midlatitudes to explain the observed ozone depletion there. In the real atmosphere, the mean inflow has to act against vortex erosion which transfers air out of the polar vortex via Rossby-wave breaking. Thus, the wave-induced mean inflow may not imply real, significant intrusions into the polar vortex.

When gravity-wave drag has some role to play in the winter stratosphere, the mechanical dissipation of Rossby waves should be taken into account. The discussion in Section 5.5 suggests that the effects of mechanical and thermal dissipation on interfacial Rossby waves always try to cancel each other, to some extent. In particular, when mechanical dissipation becomes the dominant factor, positive EP flux divergences $\nabla\cdot\mathbf{S}_{\mathrm{g}}$ can be induced by interfacial Rossby waves. If the contribution of gravity-wave drag to the mean zonal force $\overline{X}_{\mathrm{g}}$ were negligible under such circumstances, an equatorward $\overline{v}^{*}_{\mathrm{a}}$ would be expected from the steady version of Eq.(5.5.2).

In the real stratosphere, however, it would be a formally inconsistent approximation to neglect the gravity-wave drag contribution to the mean zonal force while retaining its effect on Rossby waves. A review of the direct and indirect evidence for the possible importance of orographically induced gravity-wave drag on the mean flow has been given in Palmer *et al.* (1986), and a clear laboratory demonstration has been reported in Delisi and Dunkerton (1989). It is beyond the scope of this study to discuss in detail how to calculate $\overline{X}_g$ using some useful but complicated parameterisation schemes, such as those described in Palmer *et al.* (*op. cit.*) and McFarlane (1987). To conclude our discussion, here we simply mention some relevant evidence identified from model studies. Firstly, results of model comparisons (Palmer *et al.*, 1986; Rosenlof and Holton, 1993) have shown that the lower stratosphere is a region where the effect of orographically induced gravity-wave drag on the mean flow is important. Secondly, Palmer *et al.* (*op. cit.*) further pointed out that including gravity-wave drag in their model resulted, generally speaking, in a large deceleration of westerlies in the lower stratosphere; this is consistent with the observational evidence. A reduction in westerlies in turn tends to induce a poleward meridional circulation. This implies that in Eq.(5.5.2) the effect of the gravity-wave drag on the zonal force $\overline{X}_g$ in the lower stratosphere is opposite to that on the EP flux divergence $\nabla \cdot \mathbf{S}_g$. In addition, as pointed out in Section 5.5, possible positive $\nabla \cdot \mathbf{S}_g$ due to mechanically dissipating interfacial Rossby waves in the midlatitude surf zone are likely to be submerged by the large negative $\nabla \cdot \mathbf{S}_g$ due to Rossby-wave breaking there. Therefore, it is unlikely that a persistent and significant mean outflow can be driven by mechanically dissipating Rossby waves in the real stratosphere.

Several recent studies have addressed aspects of air motion through the stratospheric polar vortex. Results from barotropic models showed that there is little transport across the vortex edge (Bowman, 1993a, Bowman and Chen, 1994). Diagnostic studies of Bowman (1993b), Dahlberg and Bowman (1994), and Chen *et al.* (1994) revealed that there is a transition in the lower stratosphere in late winter around the 400K isentropic surface, above which the polar vortex is nearly completely isolated from midlatitudes and below which more mixing occurs. The mixing in the lower stratosphere is generally believed to be the result of Rossby-wave breaking (Juckes and McIntyre, 1987; Juckes, 1989; Norton, 1994; Waugh *et al.*, 1994; Plumb *et al.*, 1994). On the other hand, trajectory calculations using horizontal winds from the U.K. Meteorological Office data assimilation system and vertical velocities from a radiation scheme showed that the net air motion is poleward and downward throughout the winter (Manney *et al.*, 1994). This is in agreement with our above argument.

5.7 Discussion

The growing interest in the polar vortex requires a closer look at fundamental processes active in the polar region. The polar γ-plane approximation introduced in Section 5.2 is an attempt to simplify the dynamical analysis of the problem. By systematically applying scaling analysis to the model equations for the polar vortex, it is shown that the γ term is the only effect of the spherical shape of the Earth that appears in the equation of quasi-geostrophic potential vorticity in the polar region.

A sharp-edge model was used to analyze the characteristics of Rossby waves associated with the polar vortex. This highly idealized model is deliberately simplified in a way that is guided by intuition about which aspects of the real polar vortex are interesting and need to be understood in a simplified context before there is hope of understanding the same aspects of the real polar vortex. The unusually steep isentropic gradients of PV at the vortex edge is one of the most significant characteristics of the polar vortex observed in the winter lower stratosphere (McIntyre and Palmer, 1983, 1984). High-resolution modelling studies (Juckes and McIntyre, 1987; Juckes, 1989; Salby *et al.*, 1990; Polvani and Plumb, 1992; Yoden and Ishioka, 1993; Waugh *et al.*, 1994; Norton, 1994) have revealed that, through Rossby-wave breaking, the air is usually stripped or eroded from the vortex edge and mixed into the surrounding surf zone. As a result, the isentropic gradients of PV that mark the vortex edge are further steepened. The sharp-edge model studied here is not intended to describe the nonlinear dynamics of the Rossby-wave breaking, vortex erosion, formation of the sharp vortex edge, etc. Instead, it confines attention to the way in which the sharp-edge structure of the polar vortex supports Rossby-wave motions. In the sharp-edge model, the basic PV gradients are idealized as a δ-function centred at the vortex edge, in contrast with the continuously-distributed PV gradients in the Charney–Drazin model. With this idealization, the interfacial Rossby waves can be derived analytically. Unlike the Rossby waves found in the Charney–Drazin model, these interfacial waves are trapped on the vortex edge.

The dynamical differences between the interfacial Rossby waves in the sharp-edge model and the Rossby waves in the Charney–Drazin model can be further highlighted by examining the characteristics of wave-induced EP flux divergences ($\nabla\cdot\mathbf{S}_g$). When steady interfacial Rossby waves in the sharp-edge model are assumed to be dissipated by Rayleigh friction and Newtonian cooling with equal relaxation coefficients, $\nabla\cdot\mathbf{S}_g$ is zero everywhere, in contrast to the negative $\nabla\cdot\mathbf{S}_g$ everywhere in the Charney–Drazin model (Holton, 1983a; Robinson, 1986; Andrews, 1987). When the Newtonian cooling dominates the Rayleigh friction, namely if $\alpha_N > \alpha_R$, $\nabla\cdot\mathbf{S}_g$ associated with steady interfacial Rossby waves in the sharp-edge model is always negative, while the corresponding $\nabla\cdot\mathbf{S}_g$ in the Charney–Drazin model is positive whenever Eq.(4.4.19) is satisfied (see the discussion in §4.4.3). On the other hand, when $\alpha_R > \alpha_N$, $\nabla\cdot\mathbf{S}_g$ is positive in the sharp-edge model and negative in the Charney–Drazin model, as shown by Eqs.(5.5.23) and (4.4.18), respectively.

We have argued that the dissipating interfacial Rossby waves are unlikely to drive a significant mean flow out of the polar vortex. In particular, if the waves are thermally dissipated, the wave-induced GLM (and TEM as well) meridional circulation is poleward in the winter stratosphere.

CHAPTER 6

CONCLUSIONS

Some fundamental problems of wave–mean interaction relevant to the middle atmospheric circulation are studied in this thesis. In the first part (Chapters 2 and 3) we demonstrate how the wave-induced mean motion can be described in terms of the wave-induced PVS transport. In the second part (Chapters 4 and 5), Rossby waves and their mean effects in the middle atmosphere are investigated within the framework of quasi-geostrophic theory.

To demonstrate how, and at what order of accuracy, the wave-induced mean motion can be described in terms of the wave-induced PVS transport, in Chapter 2 we discuss a simple thought experiment in which inertia-gravity waves are dissipated in some layer away from the bottom boundary in a Boussinesq fluid at rest. It is shown that the wave-induced irreversible PVS transport depends crucially on wave dissipation. More precisely, the wave-induced mean PV anomalies are non-zero only within the dissipation layer. When the invertibility principle for the mean PV anomaly field applies from a coarse-grain perspective, the resulting balanced mean motions (namely, the mean motions inverted from the mean PV anomalies) are dissipation-dependent in the sense that they vanish identically in the limit of zero dissipation. Moreover, these dissipation-dependent balanced mean motions are equal to the $O(a^2)$ dissipation-dependent mean motions deduced from the momentum viewpoint. When the invertibility principle applies from a fine-grain perspective, the balanced mean motions include also the $O(a^2)$ mean motions induced by the effect of wave transience. In addition, it is demonstrated that the $O(a^2)$ dissipation-dependent mean motions are cumulatively much larger than the $O(a^2)$ dissipation-independent mean motions as time goes on. Thus, even from a coarse-grain perspective, the PVS description can grasp the essence of the dissipative type of wave-induced mean motions.

It is understood that the wave-induced balanced mean motions are inherently approximate motions. In the simplest case described in Chapter 2, the mean motions inverted from the mean PV anomalies agree only to the leading order with mean motions described in terms of wave-induced momentum transport. Such agreement is unlikely to exist in the higher order mean fields. Nevertheless, model studies (e.g., McIntyre and Norton, 190, 1995) have shown that the approximations involved in the balanced motions are astonishingly good in comparison with what one might guess from the standard approximate inversion theories that restrict attention to some small parameters. In practice, therefore, the concept of balanced mean motion is widely applicable.

The theoretical basis of description of wave-mean interaction in terms of the PVS transport is discussed in Chapter 3. Both integral and differential relationships between the wave-induced PVS transport and momentum transport are derived. It is shown that the wave-induced contribution to the PVS transport is closely related to the rate of dissipation of quasimomentum. This result generalizes Taylor's well-known identity to a stably stratified, rapidly rotating fluid, such as the middle atmosphere. It also strongly supports the idea that the general nature of the dissipative type of wave-induced mean motion can be understood in a succinct and unified way, by viewing all the phenomena in terms of the wave-induced PVS transport.

In Chapter 4, forced Rossby waves are examined in a Charney–Drazin model within the framework of quasi-geostrophic theory. Our attention is focused on the dissipative nature of the Rossby waves and their mean effects. It is shown that dissipative processes in the atmosphere not only act to damp the wave amplitude, but also affect significantly the wave phase structure. For the simplest case in which the basic zonal flow and buoyancy frequency are assumed constant and Rossby waves are dissipated by Rayleigh friction and Newtonian cooling with equal, constant relaxation coefficients, the phase line always tilts westward with increasing altitude. If, on the other hand, the waves are dissipated by Newtonian cooling alone, the phase line tilts eastward when the basic zonal flow exceeds the Rossby critical velocity defined by Eq.(4.3.20). Under such circumstances, the divergence of quasi-geostrophic EP flux, $\nabla\cdot\mathbf{S}_g$, is anomalously-signed (positive) and the TEM meridional velocity is equatorward, while the GLM meridional velocity is poleward. These results, which are found to be qualitatively applicable when the basic zonal flow and buoyancy frequency vary with height, suggest that the existence of anomalously-signed EP flux divergences in the middle atmosphere may be physically possible, and the difference between the TEM and GLM meridional circulation is not always negligible.

In Chapter 5, a sharp-edge model on the polar γ-plane is introduced to study Rossby waves associated with the polar vortex. In this highly idealized model, the gradient of the mean QGPV is described by a δ-function centred at the vortex edge, in contrast to the assumption in the Charney–Drazin model that the gradient of the mean QGPV is continuously distributed. Results show that the vortex edge can support both free travelling and forced Rossby waves that have a horizontal structure decaying exponentially away from the vortex edge. When the polar night jet is strong enough, the free travelling Rossby waves with each zonal wavenumber tend to travel eastward with approximately the same zonal angular phase velocity, resembling many aspects of the behaviour of the 4-day waves observed in the winter stratosphere (Randel and Lait, 1991). Of the waves forced by the topography, only those of planetary scale (usually zonal waves 1, 2 and sometimes 3) can exist under the typical parameter conditions of the winter stratosphere (depending crucially on the strength of the polar-night jet). Again, the sign of the wave-induced quasi-geostrophic EP flux divergence depends on the relative strengths of the mechanical and the thermal dissipation (in a certain well-defined sense of the dissipation strength). When the waves are assumed to be dissipated by Rayleigh friction and Newtonian cooling with equal relaxation coefficients, i.e., $\alpha_R = \alpha_N$ where α_R and α_N are respectively the Rayleigh friction coefficient and the Newtonian cooling coefficient, $\nabla \cdot \mathbf{S}_g$ becomes zero everywhere. If $\alpha_R > \alpha_N$, on the other hand, $\nabla \cdot \mathbf{S}_g$ is positive, and *vice versa.* Note that these results are in sharp contrast to those obtained from the Charney–Drazin model. The wave-induced meridional circulation and its relevance to current questions about the effectiveness of the Rossby-wave restoring mechanism in inhibiting chemical transport across the edge of the polar vortex are also discussed. Our result gives no support to the 'flowing processor' hypothesis, which requires a significant transport of chemically perturbed air from within the polar vortex to mid-latitudes to explain the observed ozone depletion.

References

Abbe, C., 1893: *The Mechanics of the Earth's Atmosphere. A Collection of Translations.* Second collection in Smithson. Misc. Collect. No.34. Smithsonian Institution, Washington, D. C. 324pp.

Anderson, J. G. W. H. Brune and M. H. Proffitt, 1989: Ozone destruction by chlorine radicals within the Antarctic vortex: the spatial and temporal evolution of ClO-O_3 anticorrection based on in situ ER-2 data. *J. Geophys. Res.,* **94**, 11465–11479.

Andrews, D. G., 1980: On the mean motion induced by transient inertio-gravity waves. *Pure and Appl. Geophys.,* **118**, 177–188.

Andrews, D. G., 1987: On the interpretation of the Eliassen-Palm flux divergence. *Q. J. R. Meteorol. Sci.,* **113**, 323–338.

Andrews, D. G., J. R. Holton and C. B. Leovy, 1987: *Middle Atmosphere Dynamics.* Academic Press, 489pp.

Andrews, D. G. and M. E. McIntyre, 1976a: Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. *J. Atmos. Sci.,* **33**, 2031–2048.

Andrews, D. G. and M. E. McIntyre, 1976b: Planetary waves in horizontal and vertical shear: asymptotic theory for equatorial waves in weak shear. *J. Atmos. Sci.,* **33**, 2049–2053.

Andrews, D. G. and M. E. McIntyre, 1978a: Generalized Eliassen-Palm and Charney-Drazin theorems for waves on axismmetric mean flows in compressible atmospheres. *J. Atmos. Sci.*, 35, 175–185.

Andrews, D. G. and M. E. McIntyre, 1978b: An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.,* **89**, 609–646.

Andrews, D. G. and M. E. McIntyre, 1978c: On wave-action and its relatives. *J. Fluid Mech.,* **89**, 647–664.

Balsley, B. B. and R. Garello, 1985: The kinetic energy density in the troposphere, stratosphere and mesosphere: A preliminary study using the Poker Flat MST radar in Alaska. *Radio Sci.* **20**, 1355–1361.

Batchelor, G. K., 1953: The condition for dynamical similarity of motions of a frictionless perfect-gas atmosphere. *Quart. J. Roy. Meteor. Soc.,* **79**, 224–235.

Batchelor, G. K., 1967: *An Introduction to Fluid Dynamics.* Cambridge University Press, 615pp.

Béland, M., 1976: Numerical study of the nonlinear Rossby wave critical level development in a barotropic zonal flow. *J. Atmos. Sci.,* **33**, 2066–2078.

Béland, M., 1978: The evolution of a nonlinear Rossby wave critical level: effects of viscosity. *J. Atmos. Sci.,* **35**, 1802–1815.

Benney, D. J. and R. F. Bergeron, 1969: A new class of nonlinear waves in parallel flows. *Stud. Appl. Math.,* **48**, 181–204.

Bolin, B., 1950: On the influence of the Earth's orography on the general character of the westerlies. *Tellus*, **2**, 184–195.

Booker, J. R. and F. P. Bretherton, 1967: The critical layer for internal gravity waves in a shear flow. *J. Fluid Mech.,* **27**, 513–539.

Boussinesq, J., 1877: Essai sur la theorie des eaux courantes. *Mem. pres. par div. savants a l'Academie Sci.*, Paris, **23**, 1–680.

Boussinesq, J., 1903: *Theéorie analytique de la chaleur*, vol.2. Gauthier-Villars., Paris, 172pp.

Bowman, K. P., 1993a: Barotropic simulation of large-scale mixing in the Antarctic polar vortex. *J. Atmos. Sci.,* **50**, 2901–2914.

Bowman, K. P., 1993b: Large-scale isentropic mixing properties of the Antarctic polar vortex from analyzed winds. *J. Geophys. Res.*, **98**, 23013–23027.

Bowman, K. P. and P. Chen, 1994: Mixing by barotropic instability in a nonlinear model. *J. Atmos. Sci.,* in press.

Boville, B. A., 1987: The validity of the geostrophic approximation in the winter stratosphere and troposphere. *J. Atmos. Sci.,* **44**, 443–457.

Boville, B. A. and W. J. Randel, 1991: Equatorial waves in a stratospheric GCM: Effects of vertical resolution. *J. Atmos. Sci.,* **48**, 785–801.

Boyd, J., 1976: The noninteraction of waves with the zonally-averaged flow on a spherical earth and the interrelationships of eddy fluxes of energy, heat and momentum. *J. Atmos. Sci.,* **33**, 2285–2291.

Bretherton, F. P., 1966a: Critical layer instability in baroclinic flows. *Q. J. R. Met. Soc.,* **92**, 325–334.

Bretherton, F. P., 1966b: Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. *Q. J. Roy. Meteorol. Soc.,* **92**, 335–345.

Bretherton, F. P., 1969: On the mean motion induced by internal gravity waves. *J. Fluid Mech.,* **36**, 785–803.

Bretherton, F. P., 1971: The general linearised theory of wave propagation. *Lectures in Applied Mathematics,* **13**, 61–102. Amer. Maths. Soc.

Brillouin, L., 1925: On radiation stresses. *Annales de Physique,* **4**, 528–586. (in French).

Bronshtein, I. N. and K. A. Semendyayev, 1979: *Handbook of Mathematics* (English translation edited by K. A. Hirsch.) Van Nostrand Reinhold Company Inc., New York, 973pp.

Brown, R. A., 1991: *Fluid Mechanics of the Atmosphere.* Academic Press, 489pp.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *J. Meteorol.,* **4**, 135–162.

Charney, J. G., 1948: On the scale of atmospheric motions. *Geofysiske Publ.*, **17** (2), 3–17.

Charney, J. G. and P. G. Drazin, 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. *J. Geophys. Res.*, **66**, 83–109.

Charney, J. G. and A. Eliassen, 1949: A numerical method for predicting the perturbations of the middle latitude westerlies. *Tellus*, **1**, 38-54.

Charney, J. G. and M. E. Stern, 1962: On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19**, 159–172.

Chen, P., 1994: The permeability of the Antarctic vortex edge. *J. Geophys. Res.*, **99**, 20563–20571.

Chen, P., J. R. Holton, A. O'Neill, and R. Swinbank, 1994: Quasi-horizontal transport and mixing in the Antarctic stratosphere. *J. Geophys. Res.*, **99**, 16851–16866.

Chen, P. and W. A. Robinson, 1992: Propagation of planetary waves between the troposphere and stratosphere. *J. Atmos. Sci.*, **49**, 2533–2545.

Dahlberg, S. P. and K. P. Bowman, 1994: Climatology of large-scale isentropic mixing in the Arctic winter stratosphere from analyzed winds. *J. Geophys. Res.*, accepted.

Danielsen, E. F. and R. S. Hipskind, 1980: Stratospheric-tropospheric exchange at polar latitudes in summer. *J. Geophys. Res.*, **85**, 393–400.

Danielsen, E. F., R. S. Hipskind, S. E. Gaines, G. W. Sachse, G. L. Gregory, and G. F. Hill, 1987: Three-dimensional analysis of potential vorticity associated with tropopause folds and observed variations of ozone and carbon monoxide. *J. Geophys. Res.*, **92**, 2103–2111.

Davis, R. E., 1969: On the high Reynolds number flow over a wavy boundary. *J. Fluid Mech.*, **36**, 337–346.

Delisi, D. P. and T. J. Dunkerton, 1989: Laboratory observations of gravity wave critical-layer flows. *Pure Appl. Geophys.*, **130**, 445–461.

Dickinson, R. E., 1968: Planetary Rossby waves propagating vertically through weak westerly wind waveguides. *J. Atmos. Sci.*, **25**, 984–1002.

Dickinson, R. E., 1969a: Theory of planetary wave-zonal flow interaction. *J. Atmos. Sci.,* **26**, 73–81.

Dickinson, R. E., 1969b: Vertical propagation of planetary Rossby waves through an atmosphere with Newtonian cooling. *J. Geophys. Res.,* **74**, 929–938.

Dickinson, R. E., 1970: Development of a Rossby wave critical level. *J. Atmos. Sci.,* 27, 627–633.

Dickinson, R. E., 1978: Rossby waves—long-period oscillations of oceans and atmospheres *Ann. Rev. Fluid Mech.*, **10**, 159–195.

Dunkerton, T. J., 1978: On the mean meridional mass motions of the stratosphere and mesosphere. *J. Atmos. Sci.,* **35**, 2325–2333.

Dunkerton, T. J., 1985: A two-dimensional model of the quasi-biennial oscillation. *J. Atmos. Sci.,* **42**, 1151–1160.

Dunkerton, T. J., 1987: Resonant excitation of hemispheric barotropic instability in the winter mesosphere. *J. Atmos. Sci.,* **44**, 2237–2251.

Dunkerton, T. J., C.-P. F. Hsu and M. E. McIntyre, 1981: Some Eulerian and Lagrangian diagnostics for a model stratospheric warming. *J. Atmos. Sci.,* **38**, 2819–2843.

Dritschel, D. G. and R. Saravanan, 1994: Three-dimensional quasi-geostrophic contour dynamics, with an application to stratospheric vortex dynamics. *Q. J. R. Met. Soc.,* **120**, 1267–1298.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Elbern, H. and P. Spheth, 1993: Energy of Rossby waves as a part of global atmospheric oscillations. *Tellus*, **45**, 168–192.

Eliassen, A. and E. Palm, 1961: On the transfer of energy in stationary mountain waves.

Geofys Publ., **22** (3), 1–23.

Edmon, H. J., Jr., B. J. Hoskins and M. E. McIntyre, 1980: Eliassen-Palm cross sections for the troposphere. *J. Atmos. Sci.,* **37**, 2600–2616.

Ertel, H., 1942: Ein neuer hydrodynamischer Wirbelsatz. *Meteorol. Z.* **59**, 271–281. [English translation in W. Schröder: *Geophysical Hydrodynamics and Ertel's Potential Vorticity* (Selected Papers of Hans Ertel), Interdivisional Commission of History, No. 12, International Association of Geomagnetism and Aeronomy, Hechelstrasse 8, D-2820 Bremen-Rönnebeck, Germany, 33-40, 1991.]

Farman, J. C., R. J. Murgatroyd, A. M. Silnickas, and B. A. Thrush, 1985: Ozone photochemistry in the Antarctic stratosphere in summer. *Q. J. Roy. Meteorol. Soc.,* **111**, 1013–1025.

Fels, S. B., 1985: Radiative-dynamical interactions in the middle atmosphere. In *Issues in Atmospheric and Oceanic modelling* (ed. S. Manabe), *Advances in Geophysics* **28A**, Academic Press, 277–300.

Fjørtoft, R., 1950: Application of integral theorems in deriving criteria of stability for laminar flows and for the baroclinic circular vortex. *Geofys. Publ.,* **17**(6), 1–52.

Fleming, E. L., S. Chandra, J. J. Barnett, and M. Corey, 1990: Zonal mean temperature, pressure, zonal wind and geopotential height as functions of latitude. *Adv. Space Res.*, **10**, No.12, 11–59.

Foote, J. R. and C. C. Lin, 1950: Some recent investigations in the theory of hydrodynamic stability. *Quart. Appl. Math.*, **8**, 265–280.

Ford, R. and M. E. McIntyre, 1995: Balance and the slow quasi-manifold: some explicit results. Manuscript in preparation.

Fritts, D. C., 1984: Gravity wave saturation in the middle atmosphere: a review of theory and observations. *Rev. Geophys. Space Phys.*, **22**, 275–308.

Fritts, D. C., 1989: A review of gravity wave saturation processes, effects, and variability in the middle atmosphere. *Pure Appl. Geophys.,* **130**, 343–371.

Gao, S.-T., S.-Y. Tao and Y.-H. Ding, 1990: The generalized E-P flux of wave–mean flow interactions. *Science in China,* **33B**, 704–715.

Geisler, J. E. and R. E. Dickinson, 1974: Numerical study of an interacting Rossby wave and barotropic zonal flow near a critical level. *J. Atmos. Sci.,* **31**, 946–955.

Geller, M. A., M.-F. Wu and M. E. Gelman, 1983: Troposphere-stratosphere (surface-55 km) monthly winter general circulation statistics for the Northern Hemisphere—four year averages. *J. Atmos. Sci.,* **40**, 1334–1352.

Geller, M. A., M.-F. Wu and M. E. Gelman, 1984: Troposphere-stratosphere (surface-55 km) monthly winter general circulation statistics for the Northern Hemisphere—interannual variations. *J. Atmos. Sci.,* **41**, 1726–1744.

Gill, A. E., 1982: *Atmosphere-Ocean Dynamics.* Academic Press, 662pp.

Gordon, J. P., 1973: Radiation forces and momenta in dielectric media. *Phys. Rev.,* **A8**, 14–21.

Green, J. S. A., 1970: Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Q. J. Roy. Meteorol. Soc.,* **96**, 157–185.

Hamilton, K., 1982: Some features of the climatology of the northern hemisphere stratosphere revealed by NMC upper atmosphere analyses. *J. Atmos. Sci.,* **39**, 2737–2749.

Hartley, W. N., 1881: On the absorption of solar rays by atmospheric ozone. *J. Chem. Soc.,* **39**, 111–128.

Hartmann, D. L., 1985: Some aspects of stratospheric dynamics. *Adv. Geophys.,* **28A**, 219–247.

Hartmann, D. L., C. R. Mechaso and K. Yamazaki, 1984: Observations of wave-mean flow interactions in the Southern Hemisphere. *J. Atmos. Sci.,* **41**, 351–362.

Hauchecorne, A., M. L. Chanin, and R. Wilson, 1987: Mesospheric temperature inversion and gravity wave breaking. *Geophys. Res. Lett.,* **14**, 933–936.

Haynes, P. H., 1985: Nonlinear instability of a Rossby-wave critical layer. *J. Fluid Mech.,* **161**, 493–511.

Haynes, P. H., 1988: Forced, dissipative generalizations of finite-amplitude wave-activity conservation relations for zonal and non-zonal basic flows. *J. Atmos. Sci.,* **45**, 2352–2362.

Haynes, P. H., 1989: The effect of barotropic instability on the nonlinear evolution of a Rossby wave critical layer. *J. Fluid Mech.,* **207**, 231–266.

Haynes, P. H., 1990: High-resolution three-dimensional modelling of stratospheric flows: quasi-two-dimensional turbulence dominated by a single vortex. In *Topological Fluid Mechanics* (edited by H. K. Moffatt), 345–354, Cambridge University Press.

Haynes, P. H., C. J. Marks, M. E. McIntyre, T. G. Shepherd and K. P. Shine, 1991: On the "downward control" of extratropical diabatic circulations by eddy-induced mean zonal forces. *J. Atmos. Sci.,* **48**, 651–678.

Haynes, P. H. and M. E. McIntyre, 1987: On the evolution of vorticity and potential vorticity in the presence of diabatic heating and frictional or other forces. *J. Atmos. Sci.,* **44**, 828–841.

Haynes, P. H. and M. E. McIntyre, 1990: On the conservation and impermeability theorems for potential vorticity. *J. Atmos. Sci.,* **47**, 2021–2031.

Hazel, P., 1967: The effect of viscosity and heat conduction on internal gravity waves at a critical level. *J. Fluid Mech.,* **30**, 775–783.

Hearth, D. F., E. Hilsenrath, A. J. Krueger, W. Nordberg, C. Prabhakara, and J. S. Theon, 1974: The global structure of the stratosphere and mesosphere with sounding rockets and with remote sensing techniques from satellites. *Structure and Dynamics of the Upper Atmosphere* (edited by P. Verniani), 131–198, Elsevier.

Helmholtz, H. von, 1889: Über atmosphaerische Bewegungen. II. Zur Theorie von Wind und Wellen. *Sitzungsberichte Akad. Wissenschaften Berlin*, 761–780. [English transl. in Abbe (1893, pp.78–93).]

Hines, C. O., 1972: Gravity waves in the atmosphere. *Nature,* **239**, 73–78.

Hirota, I. and T. Niki, 1985: A statistical study of inertia-gravity waves in the middle atmosphere. *J. Meteor. Soc. Japan,* **63**, 1055–1066.

Hirota, I. and T. Niki, 1986: Inertia-gravity waves in the troposphere and stratosphere observed by the MU radar. *J. Meteor. Soc. Japan,* **64**, 995–999.

Holton, J. R., 1975: *The dynamic meteorology of the stratosphere and mesosphere.* American Meteorological Society, 218pp.

Holton, J. R., 1982: The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. *J. Atmos. Sci.,* **39**, 791–799.

Holton, J. R., 1983a: The stratosphere and its links to the troposphere. In *Large-Scale Dynamical Processes in the Atmosphere* (edited by B. J. Hoskins and R. P. Pearce), 277-303, Academic Press, London.

Holton, J. R., 1983b: The influence of gravity wave breaking on the general circulation of the middle atmosphere. *J. Atmos. Sci.,* **40**, 2497–2507.

Holton, J. R., 1992: *An Introduction to Dynamic Meteorology.* Academic Press, 511pp.

Holton, J. R. and R. S. Lindzen, 1972: An updated theory of the quasi-biennial cycle of the tropical stratosphere. *J. Atmos. Sci.,* **29**, 1076–1080.

Holton, J. R. and M. Wehrbein, 1980: The role of forced planetary waves in the annual cycle of the zonal mean circulation of the middle atmosphere. *J. Atmos. Sci.,* **37**, 1968–1983.

Hoskins, B. J., 1991: Towards a PV-θ view of the general circulation. *Tellus,* **43**, 27–35.

Hoskins, B. J., M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential-vorticity maps. *Q. J. Roy. Meteorol. Soc.,* **111**, 877–946. Also

111, 402–404.

Houghton, J. T., 1978: The stratosphere and the mesosphere. *Q. J. Roy. Meteorol. Soc.,* **104**, 1–29.

Iwasaki, T., 1989: A diagnostic formulation for wave-mean flow interactions and Lagrangian-mean circulation with a hybrid vertical coordinate of pressure and isentropes. *J. Meteorol. Soc. Japan,* **67**, 293–312.

Juckes, M. N., 1989: A shallow water model of the winter stratosphere. *J. Atmos. Sci.,* **46**, 2934–2955.

Juckes, M. N. and M. E. McIntyre, 1987: A high resolution, one-layer model of breaking planetary waves in the stratosphere. *Nature,* **328**, 590–596.

Kelvin, Lord, 1880: Vibrations of a columnar vortex. *Phil. Mag.,* **10**, 155-168. Also *Mathematical and Physical papers* (1910), vol.IV. 152–165, Cambridge University Press.

Keyser, D., Rotunno, R., 1990: On the formation of potential-vorticity anomalies in upper-level jet-front systems. *Mon. Wea. Rev.,* **118**, 1914–1921.

Killworth, P. D. and M. E. McIntyre, 1985: Do Rossby-wave critical layers absorb, reflect or over-reflect?. *J. Fluid Mech.,* **161**, 449–492.

Kleinschmidt, E., 1950a: Über Aufbau und Entstehung von Zyklonen (1 Teil). *Met. Runds.,* **3**, 1–6.

Kleinschmidt, E., 1950b: Über Aufbau und Entstehung von Zyklonen (2 Teil). *Met. Runds.,* **3**, 54–61.

Kleinschmidt, E., 1951: Über Aufbau und Entstehung von Zyklonen (3 Teil). *Met. Runds.,* **4**, 89–96.

Ko, M. K. W., K. K. Tung, D. K. Weisenstein, and N. D. Sze, 1985: A zonal-mean model of stratospheric tracer transport in isentropic coordinates: Numerical simulations for nitrous oxide and nitric acid. *J. Geophys. Res.,* **90**, 2313–2329.

Kuo, H. L., 1949: Dynamic instability of two-dimensional nondivergent flow in a barotropic atmosphere. *J. Meteor.*, **6**, 105–122.

Kuo, H. L., 1951: Dynamic aspects of the general circulation and the stability of zonal flow. *Tellus*, **3**, 268–284.

Kuo, H. L., 1953: The stability properties and structure of disturbances in a baroclinic atmosphere. *J. Meteor.*, **10**, 235–243.

Kurganskiy, M. V. and M. S. Tatarskaya, 1987: The potential vorticity concept in meteorology: a review. *Izv. Atmos. Oc. Phys.* (English translation), **23**, 587–606.

Lait, L. R. and J. L. Stanford, 1988: Fast, long-lived features in the polar stratosphere. *J. Atmos. Sci.*, **45**, 3800–3809.

LeBlond, P. H., 1964: Planetary waves in a symmetrical polar basin. *Tellus*, **16**, 503–512.

Lighthill, M. J., 1978: *Waves in Fluids.* Cambridge University Press, 504pp.

Lin, B.-D., 1982: The behavior of winter stationary planetary waves forced by topography and diabatic heating. *J. Atmos. Sci.*, **39**, 1206–1226.

Lin, C. C., 1955: *The Theory of Hydrodynamic Stability.* Cambridge University Press, 151pp.

Lindzen, R. S., 1981: Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.*, **86**, 9707–9714.

Lindzen, R. S. and J. R. Holton, 1968: A theory of the quasi-biennial oscillation. *J. Atmos. Sci.*, **25**, 1095–1107.

Lindzen, R. S. and C. Y. Tsay, 1975: Wave structure of the tropical stratosphere over the Marshall Islands area during 1 April – 1 July 1958. *J. Atmos. Sci.*, **32**, 2008–2021.

Lyjak, L. V. and A. K. Smith, 1987: Lagrangian mean circulations in the stratosphere. *J. Atmos. Sci.*, **44**, 2252–2266.

Mahlman, J. D. and L. J. Umscheid, 1984: Dynamics of the middle atmosphere: successes and problems of the GFDL 'SKYHI' general circulation model. In *Dynamics of the Middle Atmosphere* (edited by J. Holton and T. Matsuno), 501–525, Tokyo, Terrapub, Dordrecht, Reidel.

Mahlman, J. D. and L. J. Umscheid, 1987: Comprehensive modeling of the middle atmosphere: the influence of horizontal resolution. In *Transport Processes in the Middle Atmosphere* (edited by G. Visconti and R. R. Garcia), 251–266, Reidel.

Manney, G. L., R. W. Zurek, A. O'Neill, and R. Swinbank, 1994: On the motion of air through the stratospheric polar vortex. *J. Atmos. Sci.*, **51**, 2973–2994.

Marks, C. J., 1989: Some features of the climatology of the middle atmosphere revealed by Nimbus 5 and 6. *J. Atmos. Sci.*, **46**, 2485–2508.

Matsuno, T., 1970: Vertical propagation of stationary planetary waves in the winter Northern Hemisphere. *J. Atmos. Sci.*, **27**, 871–883.

Matsuno, T., 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.,* **28**, 1479–1494.

Matsuno, T. and K. Nakamura, 1979: The Eulerian- and Lagrangian-mean meridional circulations in the stratosphere at the time of a sudden warming. *J. Atmos. Sci.,* **36**, 640–654.

McFarlane, N. A., 1987: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.,* **44**, 1775–1800.

McIntyre, M. E., 1973: Mean motions and impulse of a guided internal gravity wave packet. *J. Fluid Mech.,* **60**, 801–811.

McIntyre, M. E., 1980a: An introduction to the generalized Lagrangian-mean description of wave, mean-flow interaction. *Pure Appl. Geophys.,* **118**, 152–176.

McIntyre, M. E., 1980b: Towards a Lagrangian-mean description of stratospheric circulations and chemical transports. *Phil. Trans. Roy. Soc. Lond.,* **A296**, 129–148.

McIntyre, M. E., 1981: On the "wave momentum" myth. *J. Fluid Mech.,* **106**, 331–347.

McIntyre, M. E., 1982: How well do we understand the dynamics of stratospheric warmings? *J. Meteorol. Soc. Japan,* **60**, 37–65.

McIntyre, M. E., 1987: Dynamics and tracer transport in the middle atmosphere: an overview of some recent developments. In *Transport Processes in the Middle Atmosphere, Proc. NATO Workshop, November 1986, Erice, Sicily* (edited by G. Visconti and R. R. Garcia), 267–296, Reidel.

McIntyre, M. E., 1989: On the Antarctic ozone hole. *J. Atm. Terr. Phys.,* **51**, 29–43.

McIntyre, M. E., 1992: Atmospheric dynamics: some fundamentals, with observational implications. In *Proc. Internat. School Phys. "Enrico Fermi", CXV Course* (edited by J. C. Gille and G. Visconti), 313–386, Amsterdam, Oxford, New York, Toronto, North-Holland.

McIntyre, M. E., 1993: On the role of wave propagation and wave breaking in atmosphere–ocean dynamics (Sectional Lecture). In *Proc. XVIII Int. Congr. Theor. Appl. Mech.,* Haifa (edited by S. R. Bodner, J. Singer, A. Solan, and Z. Hashin), 281–304, Elsevier.

McIntyre, M. E. and W. A. Norton, 1990: Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. *J. Fluid Mech.*, **212**, 403–435.

McIntyre, M. E. and W. A. Norton, 1995: Potential-vorticity inversion on a hemisphere. *J. Atmos. Sci.,* (to appear).

McIntyre, M. E. and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. *Nature,* **305**, 593–600.

McIntyre, M. E. and T. N. Palmer, 1984: The "surf zone" in the stratosphere. J. Atm. Terr. Phys., 46, 825–849.

McIntyre, M. E. and T. N. Palmer, 1985: A note on the general concept of wave breaking for Rossby and gravity waves. *Pure Appl. Geophys.,* **123**, 964–975.

McIntyre, M. E. and J. A. Pyle, 1993: Model studies of dynamics, chemistry and transport in the Antarctic and Arctic stratospheres. In *University Research in Antarctica, 1989–92* (edited by R. B. Heywood), 17–34, British Antarctic Survey, Cambridge.

McIntyre, M. E. and T. G. Shepherd, 1987: An exact local conservation theorem for finite-amplitude disturbances to nonparallel shear flows, with remarks on Hamiltonian structure and on Arnol'd's stability theorems. *J. Fluid Mech.,* **181**, 527–565.

Mechoso, C. R., D. L. Hartmann, and J. D. Farrara, 1985: Climatology and interannual variability of wave, mean-flow interaction in the Southern Hemisphere. *J. Atmos. Sci.,* **42**, 2189–2206.

Moffatt, H. K., 1978: *Magnetic Field Generation in Electrically Conducting Fluids.* Cambridge University Press, Cambridge, 343pp.

Nakamura, T., T. Tsuda, M. Yamamoto, S. Fukao, and S. Kato, 1993: Characteristics of gravity waves in the mesosphere observed with the middle and upper atmosphere radar. *J. Geophys. Res.,* **98**, 8911–9823.

Newman, P. A., 1986: The final warming and polar vortex disappearance during the Southern Hemisphere spring. *Geophys. Res. Lett.,* **13**, 1228–1231.

Nof, D., 1990: Modons and monopoles on a γ-plane. *Geophys. Astrophys. Fluid Dyn.,* **52**, 71–87.

Norton, W. A., 1988: *Balance and potential vorticity inversion in atmospheric dynamics.* PhD Thesis, University of Cambridge. 167pp

Norton, W. A., 1994: Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following coordinate system and a technique for advecting material contours. *J. Atmos. Sci.,* **51**, 654–673.

Oberbeck, A., 1879: Ueber die Wärmleitung der Flüssigkeiten bei Berücksichtigung der Strömungen infolge von Temperaturdifferenzen. *Ann. Phys. Chem.*, **7**, 271-292.

Palmer, T. N., G. J. Shutts, and R. Swinbank, 1986: Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Q. J. Roy. Meteorol. Soc.,* **112**, 1001–1039.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics.* Springer-Verlag, 710pp.

Phillips, N. A., 1963: Geostrophic motion. *Revs. Geophys. Space Phys.,* **1**, 123–176.

Pierce, B. R. and T. D. A. Fairlie, 1993: Chaotic advection in the stratosphere: implications for the dispersal of chemically perturbed air from the polar vortex. *J. Geophys. Res.,* **98**, 18589–18595.

Plumb, R. A. and A. D. McEwan, 1978: On the meridional structure of long-lived tropospheric constituents. *J. Geophys. Res.,* **93**, 15897–15913.

Plumb, R. A., D. W. Waugh, R. J. Atkinson, P. A. Newman, L. R. Lait, M. R. Schoeberl, E. V. Browell, A. J. Simmons, and M. Loewenstein, 1994: Intrusions into the lower stratospheric Arctic vortex during the winter of 1991–1992. *J. Geophys. Res.,* **99**, 1089–1105.

Polvani, L. M. and R. A. Plumb, 1992: Rossby wave breaking, microbreaking, filamentation and secondary vortex formation: the dynamics of a perturbed vortex. *J. Atmos. Sci.,* **49**, 462–476.

Prata, A. J., 1984: The 4-day wave. *J. Atmos. Sci.,* **41**, 150–155.

Proffitt, M. H., K. K. Kelly, J. A. Powell, M. Loewenstein, J. R. Podolske, S. E. Strahan and K. R. Chan, 1989: Evidence for diabatic cooling and poleward transport within and around the 1987 Antarctic ozone hole. *J. Geophys. Res.,* **94**, 16797–16813.

Proffitt, M. H., J. J. Margitan, K. K. Kelly, M. Loewenstein, J. R. Podolske and K. R. Chan, 1990: Ozone loss in the Arctic polar vortex inferred from high-altitude aircraft measurements. *Nature,* **347**, 31–36.

Prudnikov, A. P., Yu. A. Brychkov, and O. I. Marichev, 1986: *Integrals and Series. Volume 2: Special Functions.* Translated from the Russian by N. M. Queen. Gordon and Breach Science Publishers, 750pp.

Pryce, J. D., 1993: *Numerical Solution of Sturm-Liouville Problems.* Clarendon Press, Oxford, 322pp.

Randel, W. J., 1987: The evaluation of winds from geopotential height data in the stratosphere. *J. Atmos. Sci.,* **44**, 3097–3120.

Randel, W. J. and L. R. Lait, 1991: Dynamics of the 4-day wave in the Southern Hemisphere polar stratosphere. *J. Atmos. Sci.,* **48**, 2496–2508.

Rayleigh, Lord, 1880: On the stability, or instability, of certain fluid motions. *Proc. Lond. Math. Soc.,* **11**, 57–70.

Rayleigh, Lord, 1896: *The theory of sound,* vol. 2. Dover (reprinted 1945), 504pp. (see §352.)

Reed, R. J., W. J. Campbell, L. A. Rasmusson, and D. C. Rogers, 1961: Evidence of the downward-propagating annual wind reversal in the equatorial stratosphere. *J. Geophys. Res.,* **66**, 813–818.

Robinson, W. A., 1986: The application of the quasi-geostrophic Eliassen-Palm flux to the analysis of stratospheric data. *J. Atmos. Sci.*, **43**, 1017–1023.

Rood, R. B. and M. R. Schoeberl, 1983: A mechanistic model of Eulerian, Lagrangian mean, and Lagrangian ozone transport by steady planetary waves. *J. Geophys. Res.,* **88**, 5208–5218.

Rosenlof, K. H. and J. R. Holton, 1993: Estimates of the stratospheric residual circulation using the downward control principle. *J. Geophys. Res.,* **98**, 10465–10479.

Rossby, C.-G., 1936: Dynamics of steady ocean currents in the light of experimental fluid mechanics. In *Papers in Physical Oceanography and Meteorology,* Massachusetts Institute of Technology and Woods Hole Oceanic Institution. **5** (1), 1–43.

Rossby, C.-G., Collaborators, 1939: Relation between variations in the intensity of the zonal circulation of the semi-permanent centers of action. *J. Mar. Res.*, **2**, 38–55.

Rossby, C.-G., 1940: Planetary flow patterns in the atmosphere. *Q. J. Roy. Meteorol. Soc.,* **66**, 68–87.

Rossby, C.-G., 1945: On the propagation of frequencies and energy in certain types of oceanic and atmospheric waves. *J. Meteor.*, **2**, 187-204.

Rossby, C-G., 1949: On a mechanism for the release of potential energy in the atmosphere. *J. Meteor.,* **6**, 163–180.

Rowland, F. S., 1991: Stratospheric ozone depletion. *Ann. Rev. Phys. Chem.,* **42**, 731–768.

Salby, M. L., D. O'Sullivan, R. R. Garcia, and P. Callaghan, 1990: Air motions accompanying the development of a planetary wave critical layer. *J. Atmos. Sci.,* **47**, 1179–1204.

Sawyer, J. R., 1961: Quasi-periodic wind variations with height in the lower stratosphere. *Quart. J. Roy. Meteor. Soc.,* **87**, 24–33.

Schoeberl, M. R. and M. A. Geller, 1977: A calculation of the structure of stationary planetary waves in winter. *J. Atmos. Sci.,* **34**, 1235–1255.

Schoeberl, M. R. and D. L. Hartmann, 1991: The dynamics of the stratospheric polar vortex and its relation to springtime ozone depletions. *Science*, **251**, 46–52.

Scinocca, J. S. and T. G. Shepherd, 1992: Nonlinear wave-activity conservation laws and Hamiltonian structure for the two-dimensional anelastic equations. *J. Atmos. Sci.,* **49**, 3–25.

Shapiro, M. A., 1976: The role of turbulent heat flux in the generation of potential vorticity in the vicinity of upper-level jet stream systems. *Mon. Wea. Rev.,* **104**, 892–906.

Shapiro, M. A., 1978: Further evidence of the mesoscale and turbulent structure of upper level jet stream-frontal zone systems. *Mon. Wea. Rev.,* **106**, 1100–1111.

Simmons, A. J., 1974: Planetary-scale disturbances in the polar winter stratosphere.

Quart. J. Roy. Meteor. Soc., **100**, 76–108.

Smagorinsky, J., 1953: The dynamical influence of largescale heat sources and sinks on the quasi-stationary mean motions of the atmosphere. *Q. J. Roy. Meteorol. Soc.*, **79**, 342–366.

Spiegel, E. A. and Veronis, G., 1960: On the Boussinesq approximation for a compressible fluid. *Astrophys. J.*, **131**, 442–447.

Solomon, S., J. T. Kiehl, R. R. Garcia, and W. Grose, 1986: Tracer transport by the diabatic circulation deduced from satellite observations. *J. Atmos. Sci.*, **43**, 1603–1617.

Stewartson, K., 1978: The evolution of the critical layer of a Rossby wave. *Geophys. Astrophys. Fluid Dyn.*, **9**, 185–200.

Stokes, G. G., 1847: On the theory of oscillatory waves. *Trans. Camb. Phil. Soc.*, **8**, 441–473.

Stolarski, R. S., P. Bloomfield, R. D. McPeters and J. R. Herman, 1991: Total ozone trends deduced from Nimbus 7 TOMS data. *Geophys. Res. Lett.*, **18**, 1015–1019.

Stolarski, R. S., R. Bojkov, L. Bishop, C. Zerefos, J. Staehlin and J. M. Zawodny, 1992: Measured trends in stratospheric ozone. *Science*, **256**, 342–349.

Takahashi, M. and M. Uryu, 1981: The Eulerian- and Lagrangian-mean flows induced by stationary, dissipating planetary waves. *J. Meteor. Soc. Japan*, **59**, 510–534.

Takahashi, M. and B. A. Boville, 1992: A three-dimensional simulation of the equatorial quasi-biennial oscillation. *J. Atmos. Sci.*, **49**, 1020–1035.

Taylor, G. I., 1915: Eddy motion in the atmosphere. *Phil. Trans. R. Soc. Lond. A* **215**, 1–23.

Thorpe, A. J., 1985: Diagnosis of balanced vortex structure using potential vorticity. *J. Atmos. Sci.*, **42**, 397–406.

Tao, X., 1991: Atmospheric responses to lower tropospheric forcing in a multi-layer isentropic model. PhD Thesis, Princeton University. 267pp.

Truesdell, C., 1951: Proof that Ertel's vorticity theorem holds in average for any medium suffering no tangential acceleration on the boundary. *Geofisica pura e applicata* (*Pure Appl. Geophys.*), **19**, 167–169.

Tuck, A. F., 1989: Synoptic and chemical evolution of the Antarctic vortex in late winter and early spring. *J. Geophys. Res.,* **94**, 11687–11737.

Tuck, A. F., T. Davies, S. J. Hovde, M. Noguer-Alba, D. W. Fahey, S. R. Kawa, K. K. Kelly, D. M. Murphy, M. R. Proffitt, J. J. Margitan, M. Loewenstein, J. R. Podolske, S. E. Strahan and K. R. Chan, 1992: Polar stratospheric cloud processed air and potential vorticity in the Northern Hemisphere lower stratosphere at mid-latitudes during winter. *J. Geophys. Res.,* **97**, 7883–7904.

Tuck, A. F., J. M. Russell III, and J. E. Harries, 1993: Stratospheric dryness: Antiphased desiccation over Micronesia and Antarctica. *Geophys. Res. Lett.*, **20**, 1227–1230.

Uryu, M., 1980: Acceleration of mean zonal flow by planetary waves. *Pure Appl. Geophys.,* **118**, 661–693.

Ushimaru, S. and H. Tanaka, 1990: Characteristics of internal inertial gravity waves and inertial waves in the lower stratosphere observed by the MU radar. *J. Meteor. Soc. Japan,* **67**, 325–334.

van Loon, H., R. L. Jenne, and K. Labitzke, 1973: Zonal harmonic standing waves. *J. Geophys. Res.,* **78**, 4463-4471.

Venne, D. E. and J. L. Stanford, 1979: Observations of a four-day temperature wave in the polar-winter stratosphere. *J. Atmos. Sci.,* **36**, 2016–2019.

Venne, D. E. and J. L. Stanford, 1982: An observational study of high-latitude stratospheric planetary waves in winter. *J. Atmos. Sci.,* **39**, 1026–1034.

Verkley, W. T. M., 1990: On the beta plane approximation. *J. Atmos. Sci.,* **47**, 2453–2460.

Veryard, R. G. and R. A. Ebdon, 1961: Fluctuations in tropical stratospheric winds.

Meteor. Mag., **90**, 125–143

Warn, T. and H. Warn, 1976: On the development of a Rossby wave critical level. *J. Atmos. Sci.,* **33**, 2021–2024.

Warn, T. and H. Warn, 1978: The evolution of a nonlinear critical level. *Stud. in Appl. Math.,* **59**, 37–71.

Watson, G. N., 1944: *A Treatise on the Theory of Bessel Functions,* (3rd ed.). Cambridge University Press, 804pp.

Waugh, D. W., R. A. Plumb, R. J. Atkinson, M. R. Schoeberl, L. R. Lait, P. A. Newman, M. Loewenstein, D. W. Toohey, L. M. Avallone, C. R. Webster, and R. D. May, 1994: Transport of material out of the stratospheric Arctic vortex by Rossby wave breaking. *J. Geophys. Res.,* **99**, 1071–1088.

Wayne, R. P., 1991: *Chemistry of Atmospheres.* Clarendon Press, 447pp.

Wilson, R., M. L. Chanin, and A. Hauchecorne, 1991: Gravity waves in the middle atmosphere observed by Rayleigh lidar, 1. case studies. *J. Geophys. Res.,* **96**, 5153–5167.

WMO, 1985: Atmospheric ozone 1985: Assessment of our understanding of the processes controlling its present distribution and change. *World Meteorological Organization Global Ozone Research and Monitoring Project Report No. 16.* Geneva, World Meteorological Organisation, 1095+86pp.

Yang, H., 1991: *Wave packets and their bifurcations in geophysical fluid dynamics* Springer-Verlag, 247pp.

Yoden, S. and K. Ishioka, 1993: A numerical experiment on the breakdown of a polar vortex due to forced Rossby waves. *J. Meteor. Soc. Japan,* **71**, 59–72.

Zeytounian, R., 1990: *Asymptotic Modeling of Atmospheric Flows.* Translated from French by Lesly Bry. Springer-Verlag, 396pp.