Influences of Inertia-Gravity Waves on the Permeability of the Antarctic Polar Vortex Edge

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de Rector Magnificus, prof.dr. R.A. van Santen, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op woensdag 8 januari 2003 om 16.00 uur

door

Michel Christiaan Öllers

deboren te Schaesberg
Dit proefschrift is goedgekeurd door de promotoren:

prof.dr. H. Kelder
en
prof.dr. F.W. Sluijter

Copromotor:
dr.ir. L.P.J. Kamp

Druk: Universiteitsdrukkerij Technische Universiteit Eindhoven

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Öllers, Michel
Influences of inertia-gravity waves on the permeability of the Antarctic polar vortex edge,-
Michel Öllers,-
Technische Universiteit Eindhoven, 2003, -Proefschrift-
ISBN 90-386-1565-5
NUR 912
Trefwoorden: atmosferische dynamica / wervelstructuren / stratosfeer / golven / Antarctica
Subject headings: atmospheric dynamics / vortex / stratosphere / waves / Antarctica
Contents

1 Introduction 1

1.1 The structure of the middle atmosphere 1

1.1.1 Basic vertical structure of the atmosphere 1

1.1.2 Thermal and dynamical structure of the middle atmosphere 3

1.1.3 The Brewer-Dobson circulation 3

1.2 The ozone hole 5

1.2.1 Ozone production and destruction 6

1.2.2 The discovery of the ozone hole 7

1.2.3 Chemical background of the ozone hole 7

1.2.4 Dynamical background of the ozone hole 9

1.3 Transport and mixing in the Antarctic polar vortex: methods and results 14

1.3.1 Large-scale transport and mixing processes 14

1.3.2 Small-scale transport and mixing processes 17

1.4 Some basic features of inertia-gravity waves 18

1.4.1 Pure gravity waves 18

1.4.2 Pure inertial waves 20

1.4.3 Inertia-gravity waves 21

1.5 Central issues of this thesis 22

2 A study of the leakage of the Antarctic polar vortex in late austral winter and spring using isentropic and 3-D trajectories 25

2.1 Introduction 26

2.2 Trajectory model 28

2.3 Experimental setup 28

2.3.1 Initial position of air parcels inside/outside the vortex 29

2.3.2 Exchange diagnostics 29

2.3.3 Isentropic mode 30

2.3.4 3-D mode 31

2.4 Isentropic trajectory results 32
### A study of inertia-gravity wave induced exchange across the Antarctic polar vortex edge in late austral winter using 3-D trajectories

3.1 Introduction ............................. 48  
3.2 Trajectory model and data ...................... 49  
3.3 Imposed inertia-gravity wave field ...................... 50  
3.4 Experimental design and exchange diagnostics ........... 52  
3.5 Trajectory results ........................... 54  
3.5.1 Mass exchange rates .............................. 54  
3.5.2 Features of air mass distributions in the presence of IGWs 58  
3.5.3 Effect of IGWs on diabatic descent ............ 62  
3.6 Conclusions ............................ 64

### Simple analytical models of propagating inertia-gravity waves in barotropic shear layers

4.1 Inertia-gravity waves incident upon a piecewise continuous wind profile 71  
4.2 Inertia-gravity waves incident upon a hyperbolic tangent wind profile 76  
4.2.1 Reduction to the hypergeometric equation ............ 77  
4.2.2 Reflection and transmission coefficients ............ 78  
4.3 Inertia-gravity waves incident upon a hyperbolic cosine wind profile 80

### Propagation properties of inertia-gravity waves through a barotropic shear layer and application to the Antarctic polar vortex

5.1 Introduction ................................ 96  
5.2 Model formulation .......................... 98  
5.2.1 Basic model equations ....................... 98  
5.2.2 Conservation of momentum fluxes ............... 99  
5.3 Numerical results ........................... 101  
5.3.1 Numerical solution method ................. 102  
5.3.2 Choice of parameters ................. 102  
5.4 Results ................................. 103  
5.4.1 Inertially stable background flow ............... 103
## Contents

5.4.2 Inertially unstable background flow ........................................ 108
5.5 Conclusions ............................................................................ 112
5.6 Appendix: Jump in momentum flux across the $\Omega$-critical level ... 113

6 Summary, conclusions, and outlook ........................................... 115
6.1 Summary and conclusions ....................................................... 115
6.2 Outlook ................................................................................ 119

Bibliography .............................................................................. 123

Samenvatting ............................................................................. 132

Dankwoord ............................................................................... 137

Curriculum Vitae ...................................................................... 139
Chapter 1

Introduction

The work that is described in this thesis focuses on the role of inertia-gravity waves in the permeability of the Antarctic polar vortex. In this chapter, a short overview is presented of some of the main properties of the atmosphere. Then, a historical overview of the discovery of the ozone hole is given, together with the chemical and dynamical background of the ozone hole. Thereafter, a summary of the numerical techniques used to investigate the permeability of the Antarctic polar vortex is given, as well as some major results that have been achieved over the years. Also, some basic features of inertia-gravity waves are given, and this chapter is concluded by a description of the major research goals.

1.1 The structure of the middle atmosphere

1.1.1 Basic vertical structure of the atmosphere

The atmosphere can be divided into several layers (Figure 1.1). The lowermost layer from the surface to approximately 11 km is the troposphere. The troposphere is the part of the atmosphere we live in and where practically all the familiar weather phenomena occur. The troposphere contains nearly all water that is present in the atmosphere and is further characterized by a decrease in temperature with height at a rate of about 6-10 K km$^{-1}$. The troposphere is bound by the tropopause, which is defined as the lowest level at which the temperature lapse rate decreases to 2 K km$^{-1}$ or less, provided that the average lapse rate between this level and the upper levels within 2 km does not exceed 2 K km$^{-1}$ (definition according to the World Meteorological Organization, see also Ambaum (1997)). The layer above the tropopause between approximately 11 and 50 km is referred to as the stratosphere. The stratosphere contains the bulk of ozone molecules. These absorb solar ultraviolet radiation
(wavelength between approximately 280-400 nm). The associated heating results in an increase of temperature with height. The stratosphere is very dry and is dominated by radiative processes. The stratopause forms the transition to the mesosphere in which temperature again decreases with height. Its top is at about 85 km and is called the mesopause. The stratosphere and the mesosphere together are commonly referred to as the middle atmosphere. The thermopause marks the boundary between the mesosphere and the thermosphere. In the thermosphere temperature again increases with altitude. Ionisation of molecules by extreme ultraviolet radiation (wavelengths shorter than approximately 200 nm) means that electromagnetic effects become important in the dynamics. The atmosphere above the mesopause is referred to as the upper atmosphere.

Figure 1.1: Vertical mean temperature distribution in the lowest 100 km of the atmosphere. Based on data from Fleming et al. (1990). After Andrews (2000).
1.1.2 Thermal and dynamical structure of the middle atmosphere

In Figure 1.2 the zonal mean temperature distribution in September is shown. In the troposphere temperature decreases with altitude and latitude. A temperature minimum is present just above the equatorial tropopause. In the stratosphere the temperature increases with altitude, and above approximately 30 mb, decreases from the summer pole to the winter pole. Extremely low temperatures are found near the pole in the winter hemisphere, caused by the lack of heating by sunlight during the polar night. The difference in solar insulation causes a temperature difference between equator and pole, that in combination with the rotation of the earth, drives air motions that are mainly zonal. The zonal mean wind distributions are shown in Figure 1.3. Stable features of the upper troposphere are the subtropical westerly jetstreams. The decrease in temperature with latitude in the stratosphere, in combination with the thermal wind relationship, leads to the existence of a strong polar night jet in the winter stratosphere.

1.1.3 The Brewer-Dobson circulation

In the previous section it was shown that the circulation in the stratosphere is mainly zonal. The presence of planetary waves disturbs this circulation pattern. Wave dissipation causes transport of mass in the poleward direction in the stratosphere. Rising air in the tropics brings chemical species from the troposphere into the lower stratosphere. These are subsequently transported poleward, followed by, on average, a slow subsidence at mid and high latitudes (Figure 1.4). This circulation pattern was first described by Brewer (1949) and Dobson (1956) and is referred to as the Brewer-Dobson circulation. Ozone has its most important sources in the tropical stratosphere. The poleward transport of ozone by the Brewer-Dobson circulation explains why so much ozone is found at middle and high latitudes. Even though most stratospheric ozone is produced in the tropics, due to the lower tropopause and hence larger thickness of the stratosphere at high latitudes (the lower tropopause), the ozone column maximizes there (Figure 1.5). However, in winter and spring above the polar stratosphere of the Southern Hemisphere, the presence of the ozone hole violates the picture above. The dramatic destruction of ozone within the ozone hole leads to low ozone values there. At high latitudes ozone has a very long lifetime, because the destruction of ozone depends on the presence of ultraviolet radiation which is virtually absent in winter at high latitudes (polar night) (e.g. Brasseur et al., 1999).
Figure 1.2: Vertical distribution of the observed monthly and zonally averaged temperature (K) for July on both hemispheres. After Fleming et al. (1990).

Figure 1.3: Zonally averaged geostrophic winds (ms⁻¹) for July based on the monthly and zonally averaged temperatures in Figure 1.2.
Figure 1.4: Schematic representation of the Brewer-Dobson circulation. Chemical species such as ozone are transported from their source region through the equatorial tropopause by an upward flow. At middle latitudes the transport is characterized by a stratospheric poleward flow. At high latitudes the flow is downward and the ozone column is large. These large ozone columns over the poles are due to the lower tropopause and hence relatively larger thickness of the polar stratosphere compared to the tropical stratosphere.

1.2 The ozone hole

The ozone layer protects the biosphere from potentially damaging ultraviolet radiation. Destruction of stratospheric ozone, thins the ozone layer, which can lead to enhanced ultraviolet radiation reaching the earth surface. It is well known now that ultraviolet radiation can have severe health effects on the skin, eyes and immune system of humans and animals. It also affects agricultural crops and the flora in general. Although the ozone concentration in the atmosphere varies on a seasonal time scale and with latitude, spaceborne and other measurements have reported downward trends in stratospheric ozone that are largest in winter and early spring and most pronounced in the polar regions (e.g. Randel and Wu, 1995). The ‘ozone hole’ that reappears every year in early spring over the Antarctic continent is the most striking example of this. Man-made chemical compounds containing chlorine, fluorine and carbon (chlorofluorocarbons or CFCs) are responsible for the depletion of ozone in
Figure 1.5: Column amount of total ozone (Dobson units) as a function of latitude and season obtained from TOMS observations averaged for the period 1979-1986. 1 Dobson unit is defined as the height of the ozone column, in hundredths of a millimetre, if all ozone molecules in this column were brought to a pressure of 1 atm. and a temperature of 273 K. After Brasseur et al. (1999).

1.2.1 Ozone production and destruction

In the stratosphere ozone is generated when molecular oxygen ($O_2$) photodissociates due to ultraviolet radiation from the sun. The oxygen atoms (O) recombine with molecular oxygen ($O_2$) to form ozone ($O_3$). Ozone itself is photodissociated to one oxygen atom (O) and molecular oxygen ($O_2$) (Chapman, 1930). This creates a balance between production and destruction. The bulk of ozone is found in the stratosphere between 15 and 30 km. The ozone production/destruction schemes of Chapman lead to an overestimation of the total ozone column. Obviously other chemical processes are active that can destroy ozone. Today it is known that several catalysts speed up the process of ozone destruction. Catalysts promote chemical
reaction chains without being ‘consumed’ themselves. A typical catalytic reaction chain for ozone destruction is:

\[
X + O_3 \rightarrow XO + O_2
\]

\[
XO + O \rightarrow X + O_2
\]

Net: \(O_3 + O \rightarrow 2O_2\)

Various catalytic cycles that destroy ozone have been identified in the stratosphere (e.g. Brasseur et al. (1999)). For instance chlorine, bromine, nitrogen oxides and hydroxyl.

1.2.2 The discovery of the ozone hole

The first observational evidence that indicated the presence of an Antarctic ozone hole was presented by Chubachi (1984). Ground based observations from Syowa station in Antarctica from February 1982 through January 1983 showed low ozone values in the months of September and October. However, Chubachi did not point out that the measured ozone values for September and October were anomalously low compared to previous years. It was Farman et al. (1985) who first identified a remarkable downward trend in ozone above Antarctica based on a long record of ozone measurements at Halley Bay (Figure 1.6). They discovered a dramatic 30\% decrease in the total ozone amount in October for the period 1980-1984 compared to the period 1957-1973. The ground based measurements of Farman et al. (1985) were soon after confirmed by the satellite based measurements of Stolarski et al. (1986). Besides the dramatic decline in total ozone since 1980, they also showed that the low ozone values occurred over much of the area above the Antarctic continent thereby creating an Antarctic ozone hole (Figure 1.7).

1.2.3 Chemical background of the ozone hole

Several explanations have been put forward to explain the spectacular decrease in total ozone in austral spring. Farman et al. (1985) argued that the cold temperatures in the polar stratosphere together with an increase in active chlorine radicals released from man-made CFCs were the cause. Later other dynamical and chemical explanations were suggested of which most proved to be incorrect (see e.g. Solomon (1999) for an overview). Solomon et al. (1986) were the first to reveal some of the essential clues that could explain the severe ozone loss above Antarctica. Their theory involves heterogeneous chemical reactions that occur in so-called polar stratospheric clouds (PSCs) (McCormick et al., 1982). Although PSCs even today still have not revealed all of their mysteries, it is widely accepted that PSCs form due to the heterogeneous
nucleation of nitric acid particles and water vapor on pre-existing aerosols (e.g. Ham- mill and Toon (1991), Peter (1997) and Tolbert and Toon (2001)). During the austral winter months, temperatures within the ozone hole can drop to values as low as 185 K (-88° C) which is cold enough for PSCs to form. The lack of sunlight prevents ozone destruction. However, during spring, when part of the polar vortex becomes sunlit, PSCs provide a reactive surface on which ozone can be destroyed catalytically. Active chlorine radicals play a major role in the catalytic destruction of ozone on PSCs. In fact, Molina and Rowland (1974) already recognized the potential danger of chlorine compounds released from CFCs in the destruction of ozone. After all, the theory of Farman and his co-workers was close to the essence but they supposed the wrong chemical mechanism. Chlorine radicals are not only produced due to the slow process of photochemical dissociation of CFCs, PSCs themselves are a source of chlorine radicals. This can be illustrated by looking at one of the most important heterogeneous reactions in this context (Solomon et al., 1986) i.e.

$$\text{ClONO}_2 + \text{HCl} \rightarrow \text{HNO}_3 + \text{Cl}_2$$  \hspace{1cm} (1.2)
in which the long-lived chlorine compounds chlorine nitrate and hydrogen chloride (reservoir species) are converted into more reactive nitric acid and molecular chlorine. Molecular chlorine is quickly photolyzed in austral spring into active chlorine radicals which catalytically destroy ozone according to (1.1). Moreover, subsidence of PSC particles effectively removes nitrate and thereby the nitrogen oxides needed to recreate the reservoir compound ClONO₂ (denitrification).

It was in fact the complexity of the physical and chemical processes needed to explain the dramatic decline in ozone within the ozone hole that surprised many researchers involved in early ozone hole research. Before the discovery of the ozone hole, it was generally thought that only homogeneous gas-phase chemistry could destroy ozone.

1.2.4 Dynamical background of the ozone hole

Dynamical mechanisms play an important role in understanding the confinement and strong cooling of the air masses within the ozone hole. At the onset of the polar night in late austral autumn and early winter, the temperature difference between the
pole and the equator increases. The strong temperature gradient, due to geostrophic equilibrium causes the development of a strong westerly jet in the stratosphere with wind velocities up to 80 ms$^{-1}$. This jet is commonly referred to as the polar night jet (Figure 1.8). The area enclosed by the polar night jet is generally referred to as the Antarctic polar vortex.

The Antarctic polar vortex is often analysed in terms of the conserved quantity potential vorticity (PV). The vortex edge is characterised by a circumpolar zone with large gradients in PV (Figure 1.9). PV can be defined in terms of potential temperature, which in its turn is defined as (e.g. Holton, 1992):

$$\theta = T \left( \frac{p}{p_0} \right)^{\kappa},$$

where $T$ is the temperature, $p$ is the pressure, $p_0$ is a reference pressure (usually 1000 hPa) and $\kappa$ is the ratio of specific heat at constant volume to that at constant pressure and is taken to be equal to $5/7$. In the stratosphere $\theta$ is a monotonic function with height, and can therefore be used as a vertical coordinate. PV is then defined as

$$PV = -\frac{\zeta_\theta + f}{g} \frac{\partial p}{\partial \theta}.$$  

$\zeta_\theta + f$ is known as the absolute vorticity consisting of the relative vorticity $\zeta$ and the earth rotation $f$. The relative vorticity $\zeta$ is evaluated on a surface of constant potential temperature (isentropic surface). It equals the curl of the wind field $(\nabla \times \vec{v}) \cdot \vec{k}$, where $\vec{v} = u\hat{i} + v\hat{j} + w\hat{k}$ defines the wind field with $\hat{i}$, $\hat{j}$ and $\hat{k}$ being the unit vectors in eastward, northward and upward direction, respectively. $g$ is the gravitational acceleration. PV is conserved under inviscid and adiabatic conditions (e.g. Haynes and McIntyre, 1987). This means that the absolute vorticity $\zeta_\theta + f$ can only change in proportion to the thickness $-(1/g)\partial p/\partial \theta$ between two isentropic surfaces (Salby, 1996). In the stratosphere PV can be considered to be conserved on isentropic surfaces over a period of a couple of days to more than a week. Therefore isentropic surfaces act as material surfaces (e.g. Haynes and McIntyre, 1987, 1990) on which the motion of air parcels can be followed in time. PV can be regarded as an active tracer since it depends intimately on the flow field itself. This is in contrast to passive tracers, which are merely advected by the background flow field. Diabatic heating or cooling processes can induce vertical air motions and violate the conservation principle of PV. The large gradient in PV that marks the edge of the polar vortex acts as a very effective barrier to cross-edge transport of trace constituents such as ozone from the polar vortex core to midlatitudes and vice versa (McIntyre, 1989). The effective isolation of the air inside the vortex and the lack of sunlight during the polar night leads to strong radiative cooling. The diabatic cooling rate is height dependent with largest rates in the upper vortex and decreasing downwards. Descent rates
The ozone hole

Figure 1.8: Zonal wind vector field (m/s) on the 450 K isentropic level for October 1, 1998 based on analysis data from the European Centre for Medium-range Weather Forecasts (ECMWF). The polar night jet is approximately the high wind speed region between 50° S and 60° S.

...depend on the season with generally larger rates in early winter than in spring (e.g. Manney et al. (1994); Öllers et al. (2002b)). Furthermore, studies by e.g. Manney et al. (1994); Wauben et al. (1997a); Kawamoto and Shiotani (2000); Rosenfield and Schoeberl (2002) and Öllers et al. (2002b) found that diabatic descent rates also show a clear year-to-year variability. The development of the polar vortex edge starts first at high altitudes in June and July and shifts downward as time progresses.

Planetary waves play an important role in the formation and break-down of the vortex edge. Upward propagating planetary waves from the troposphere increase in amplitude due to the decreasing density of the air. Eventually they can become saturated and break and disturb the polar vortex. On the other hand, planetary waves break when their phase speed equals the speed in the vortex edge. At these so-called critical levels energy exchange between the wave and the background flow can take place, and the wave may eventually break. As the wind velocities in the vortex edge further increase during winter, the wave phase velocities are no longer able to match the velocities in the vortex edge. The zone of wave breaking then shifts equatorward...
Figure 1.9: Potential vorticity (0.1 PVU) on the 450 K isentropic level for October 1, 1998 based on ECMWF analysis data. The polar vortex edge is characterized by the zone of large gradients in PV.

(Bowman, 1993a, 1996). In late winter and early spring the polar vortex edge is then characterized by a strong gradient in PV surrounded by a mixing zone (‘surf zone’) of breaking planetary waves at midlatitudes (McIntyre and Palmer, 1983; Juckes and McIntyre, 1987).

By late winter and early spring temperatures inside the polar vortex can drop as low as 180 K (-93°C), cold enough for the formation of PSCs. In early austral spring (September-October) the sun starts to warm the polar atmosphere which initiates the rapid destruction of ozone on PSCs. Now, record low ozone values can be observed within the ozone hole (Figure 1.10). By mid to late spring (November-December), due to the solar heating, the equator to pole temperature gradient starts to decrease, accompanied by a weakening of the PV gradient at the edge of the polar vortex. The lower wind velocities in the polar night jet make the polar vortex more susceptible to disturbances caused by (breaking) planetary waves. Eventually, the polar vortex
The ozone hole

Figure 1.10: Image of assimilated total ozone based on GOME measurements for October 1, 1998. The ozone hole corresponds to the black area over the Antarctic continent. Note that the shape of the ozone hole corresponds well with the shape and location of the polar vortex edge in Figure 1.9.

breaks down around December, accompanied by the shedding of large fragments of chemically perturbed air from the inner vortex into midlatitudes. The equatorward drift of these fragments causes a ‘dilution-effect’ that can manifest itself in low ozone values at midlatitudes (Atkinson et al., 1989).

In the Arctic stratosphere there also exists a polar vortex in winter and spring. However, this polar vortex is not as pronounced as its Antarctic counterpart. In the Arctic the presence of large-scale topographic features such as mountain ranges and larger land-sea temperatures contrasts, induces planetary waves that more frequently disturb the polar vortex (e.g. Waugh et al., 1994a; Plumb et al., 1994; Polvani and Saravanan, 2000). The greater planetary wave activity in the midlatitude Northern Hemisphere induces a stronger Brewer-Dobson circulation and thereby a larger transport of ozone from the tropics poleward and downward to the Arctic regions (section 1.1.3). Also, the disturbance by these planetary waves enhances the amount of mixing of air between midlatitudes and the Arctic vortex. As a consequence, temperatures
inside the Arctic vortex are not as low as in the Antarctic. The higher temperatures prevent the large-scale formation of PSCs, which are responsible for the dramatic destruction of ozone during the Antarctic spring.

1.3 Transport and mixing in the Antarctic polar vortex: methods and results

1.3.1 Large-scale transport and mixing processes

The dynamics of the Antarctic polar vortex has received much attention since the discovery of the ozone hole in the mid 1980s. Especially the degree of permeability of the Antarctic polar vortex edge has aroused a lot of interest. A greater permeability of the vortex edge implies the possibility of dilution of ozone depleted air from the ozone hole with air from midlatitudes. In the last two decades, a column ozone decline was observed at midlatitudes in the Northern Hemisphere. Ozone decline in the lower stratosphere is influenced by local ozone loss that is enhanced by volcanic aerosol particles and transport from other regions. It is now believed that halogens like chlorine are the primary cause of midlatitude ozone depletion (WMO-UNEP, scientific assessment of ozone depletion 1998/2002; e.g. Randel and Wu (1995)). Dilution of ozone-poor air at midlatitudes occurs every year in spring during the break-up of the vortex, but it may also occur at earlier stages caused by either large-scale processes (planetary waves) or small-scale processes (inertia-gravity waves). Over the years two rival views have developed (Randel, 1993) that center around the question whether air within the polar vortex is effectively isolated horizontally by an impermeable vortex edge and vertically by slow vertical motions (‘containment vessel’) (Hartmann et al., 1989a; Schoeberl et al., 1992) or whether there is rapid flow-through with air leaking at the bottom of the polar vortex (Proffitt et al., 1989; Tuck, 1989). In order to find a way out of this controversy, a large number of studies have been performed. Also, measurement campaigns like Strateole will provide important information about the dynamical and chemical processes that play a role in the ozone hole. In the near future, a series of long duration stratospheric balloons will be launched from McMurdo station, to study the dynamics of the polar vortex at the end of the Southern Hemisphere winter. Below, a summary of the major results that have been obtained over the years using different numerical techniques will be given.

There are a number of techniques that are commonly used in the study of transport and mixing across PV barriers. One such barrier is the Antarctic polar vortex edge, but also the tropopause is an example of this (e.g. Meloen, 2002). These techniques use wind and temperature fields from weather forecast or climate models as
Transport and mixing in the Antarctic polar vortex: methods and results

input. In weather forecast models meteorological observations are routinely assimilated so that they give a realistic description of the state of the atmosphere.

One such technique is to use wind and temperature fields to follow the trajectories of air parcels in space and time with the help of trajectory models. In trajectory models one computes the displacements of air parcels. Since the gridpoints and the availability in time of the wind and temperature data do generally not coincide with the location and time of the air parcels, interpolation in space and time is needed. Studies by e.g. Bowman (1993a); Dalhberg and Bowman (1994); Bowman and Chen (1994) have used such air parcel trajectories on isentropic surfaces. They have shown that the polar vortex edge acts as a nearly impermeable barrier to cross-edge exchange above approximately the 425 K isentropic surface. Other studies have used 3D-trajectories that also take into account (vertical) cross-isentropic motions (Manney et al., 1994; Trouniday et al., 1995; Paparella et al., 1997). Cross-isentropic flow is directly proportional to the amount of diabatic heating or cooling. The way in which these effects are accounted for in trajectory models differs. Manney et al. (1994) and Trouniday et al. (1995) use a radiation model to compute heating or cooling rates, whereas Paparella et al. (1997) use the vertical wind component from the ECMWF model. Overall, they come to the same conclusion concerning the amount of quasi-horizontal transport across the vortex edge. Manney et al. (1994) found that vertical air motions within the polar vortex are of great importance in the contribution to the leakage of air through the bottom of the polar vortex. Large vertical descent rates may eventually bring air parcels into the lower polar vortex below 425 K. Below these levels the vortex edge is weak and air then mixes more easily between polar and midlatitudes.

Contour advection with surgery (CAS) (Norton, 1994; Waugh and Plumb, 1994; Baker and Cunnold, 2001) is another technique that is commonly used in studies that are engaged in large-scale quasi-horizontal transport and mixing in the atmosphere. Basically, CAS involves the following of the evolution of material contours. These material contours consist of a finite number of parcels, but as the contours stretch or shrink due to the varying wind field, the number of parcels that make up the contour vary in order to represent the contour in the best possible way. CAS is in fact a sophisticated trajectory calculation. Since PV is materially conserved in the stratosphere over a period of weeks, PV contours act as material contours and their evolution, especially in the vortex edge region, can be followed. CAS is also a useful tool to identify barriers to transport. For example, the vortex edge region is a region where the material contours show little stretching and thus little mixing and exchange of air parcels to midlatitudes occurs (e.g. Pierce and Fairlie, 1993; Chen et al., 1994; Chen, 1994; Mariotti et al., 1997, 2000). In highly dynamical environments such as the Antarctic polar vortex it regularly occurs that small-scale features or thin fil-
Figure 1.11: An example of a reconstructed ozone map using CAS for August 10, 1994, 2300 UTC at 470 K. Clearly visible are the filamentary structures wrapped around the polar vortex. The data used are a combination of ECMWF analyses and ozone fields obtained from the Microwave Limb Sounder on board of the Upper Atmosphere Research Satellite. Adapted from Mariotti et al. (2000). Courtesy of Vincent Daniel.

In such cases the number of parcels that is needed to represent a contour grows excessively and ‘surgery’ is then needed to remove certain parts of the contour. This procedure keeps the calculation within practically performable ranges. The CAS technique has proven to be very useful in representing filamentary structures that are peeled off the main polar vortex (Figure 1.11). Waugh et al. (1994a) have shown for the Arctic vortex that certain filaments predicted by CAS calculations correspond well with the occurrence of similar filaments observed from aircraft data.

In contrast to the Lagrangian approach above, the wind and temperature fields can also serve as input in Eulerian tracer or chemical transport models (CTMs) (e.g. Heimann, 1995). CTMs solve the continuity equation for an arbitrary number of atmospheric tracers on a global Eulerian grid. Tracer transport is accomplished by calculating the air mass fluxes between the three-dimensional gridboxes. An example of a study who used a CTM is Wauben et al. (1997a). They calculated the three-dimensional evolution of passive tracers released inside the polar vortex for late austral to early spring 1990-1993. Based on their computed exchange rates they
conclude that the polar vortex edge is fairly isolated. A drawback of their study is the use of coarse resolution data which provide an additional numerical source of leakage. Lee et al. (2001) studied mixing processes and ozone loss in the Antarctic polar vortex by following the evolution of an artificial tracer advected by high-resolution wind fields. They identified two distinct regions of strong mixing in the vortex core and at midlatitudes. A third isolated region of weak mixing coincident with the vortex edge existed in between.

Typical drawbacks of the methods (e.g. Stohl et al., 2001) described above are mostly related to the limited horizontal and vertical resolution of the available wind and temperature data and the frequency for which these data are available. For example the accuracy of air parcel trajectories is influenced by errors in the numerical integration scheme as well by errors in the wind data. A typical limitation of the CAS method is that it provides a qualitative picture of the evolution of material contours. Recently, some limitations on the use of the CAS method to investigate vortex dynamics have been put forward (Baker and Cunnold, 2001).

### 1.3.2 Small-scale transport and mixing processes

In the previous section some aspects of large-scale transport and mixing found in various numerical models were discussed. Filamentation of the outer vortex edge is caused by breaking planetary waves. These filaments can be torn off the main vortex and mixed at midlatitudes. Erosion of the outer vortex edge steepens the gradient in PV, making the vortex edge even more impermeable to large-scale cross-edge exchange. Not only large-scale processes can induce cross-edge transport; there is some evidence that smaller scale transport processes related to (inertia-)gravity waves also can induce transport of minor trace constituents across the vortex edge. For example, observations of vertically laminated layers of ozone have been made during several Arctic measurement campaigns (e.g. Reid and Vaughan, 1991; Reid et al., 1993). From these studies it has become clear that there is a close relationship between filaments and laminae. In fact they are two sides of the same coin. The vertical shear of the background wind field creates shallow vertical layers of tracers (laminae) whose horizontal manifestations are filaments. The filaments in these studies are mostly caused by differential quasi-horizontal advection due to planetary waves. Laminae are local features since they are usually observed in vertical ozone profiles obtained from ozonosonde data, but they can be part of a horizontally much larger feature, namely a filament. Danielsen et al. (1991), however, showed that low-frequency inertia-gravity waves can also produce horizontally filamented structures near jetstreams. Others (Teitelbaum et al., 1994; Reid et al., 1994; Teitelbaum et al., 1996) found signatures of (inertia-)gravity wave induced laminae in ozone profiles above the Arctic regions. Only very few studies have reported on observations of
Introduction

18 ozone laminae in the neighborhood of the Antarctic polar vortex (Deshler et al., 1990). Recently Pfenninger et al. (1999); Guest et al. (2000) and Yoshiki and Sato (2000) studied (inertia-)gravity wave characteristics and conclude that tropospheric mid-latitude jetstreams and the polar night jet are sources of these waves. The study by Pierce et al. (1994) is probably the only one that addressed in detail the effects of an idealized inertia-gravity wave field upon the mixing and mass exchange across the Antarctic polar vortex edge. Pierce et al. (1994) found that such an idealized wave field can induce irreversible mixing in the vicinity of the polar vortex edge. Quantification of mass exchange across the Antarctic polar vortex edge due to inertia-gravity waves will be discussed later in this thesis.

1.4 Some basic features of inertia-gravity waves

This section will give a short overview of some important features of gravity, inertial and inertia-gravity waves. For an atmosphere at rest the condition of hydrostatic equilibrium $\frac{dp}{dz} = -\rho g$ is satisfied. Hydrostatic equilibrium leads to a mean stratification in which density decreases with height and consequently lighter air is on top of heavier air, creating a statically stable atmosphere. In an incompressible atmosphere, this is always the case. However, in the real (compressible) atmosphere convective processes and orographic influences in the troposphere and radiative processes in the stratosphere can bring the atmosphere out of this equilibrium and induce vertical motions. If an air column is brought out of equilibrium, a restoring mechanism will act such as to bring the column back in equilibrium. Such a restoring mechanism is always associated with waves. We will consider here the restoring forces that act on a single air parcel when brought out of its equilibrium position. More detailed treatments of the properties of gravity, inertial and inertia-gravity waves can be found in standard textbooks (e.g. Andrews et al., 1987; Holton, 1992).

1.4.1 Pure gravity waves

Consider an air parcel of density $\rho$ and pressure $p$ at some level $z$ (Figure 1.12). The environment of the air parcel has density $\rho_0$ and pressure $p_0$ and is in hydrostatic equilibrium $\frac{dp_0}{dz} = -\rho_0 g$. Suppose now that the air parcel is vertically displaced from its initial level $z$ to level $z + \delta z$. Further, suppose that the air parcel is thermally insulated from its environment. The equation of motion for the air parcel of unit volume can be written as

$$\rho \frac{d^2 \delta z}{dt^2} = \frac{\partial p}{\partial z} - \rho g. \quad (1.5)$$
Some basic features of inertia-gravity waves

\[
\begin{array}{c}
z + \delta z \\
\theta_0 + \left(\frac{d\theta_0}{dz}\right)
\end{array}
\]

\[
\begin{array}{c}
\delta z \\
p \\
p
\end{array}
\]

\[
\begin{array}{c}
-\rho \frac{N^2 \delta z}{dz}
\end{array}
\]

\[
\begin{array}{c}
z \\
\theta_0
\end{array}
\]

**Figure 1.12:** Schematic representation of an air parcel that is displaced over a distance \(\delta z\). The displacement from the parcel’s equilibrium position induces a restoring force \(-\rho N^2 \delta z\) per unit volume. The parcel’s density and pressure are denoted by \(\rho\) and \(p\), respectively, whereas the density and pressure of the environment are denoted by \(\rho_0\) and \(p_0\).

We assume that the air parcel does not influence its environment. Through compression or expansion the air parcel pressure automatically adjust to the pressure of the environment i.e. \(p = p_0\). Using the condition of hydrostatic equilibrium of the environment, equation (1.5) can be written as

\[
\frac{d^2 \delta z}{dt^2} = \left(\frac{\rho_0 - \rho}{\rho}\right) g. \tag{1.6}
\]

Now, expression (1.3) for the potential temperature and the ideal gas law can be used to express density in (1.6) in terms of potential temperature i.e.

\[
\frac{d^2 \delta z}{dt^2} = \frac{\delta \theta}{\theta_0} g, \tag{1.7}
\]

where \(\delta \theta\) represents the deviation of the potential temperature of the air parcel from its environmental value \(\theta_0\) at the initial level \(z\). The environmental potential temperature \(\theta_0(z + \delta z)\) can be written as

\[
\theta_0(z + \delta z) \approx \theta_0(z) + \frac{d\theta_0}{dz} \delta z. \tag{1.8}
\]

Under adiabatic conditions the potential temperature of the air parcel is conserved such that

\[
\delta \theta = \theta_0(z) - \theta_0(z + \delta z) = -\frac{d\theta_0}{dz} \delta z. \tag{1.9}
\]
Substitution of (1.9) in (1.7) yields

\[
\frac{d^2 \delta z}{dt^2} = -\omega_B^2 \delta z, \tag{1.10}
\]

where

\[
\omega_B^2 \equiv N^2 = \frac{g d\theta_0}{\theta_0 dz}, \tag{1.11}
\]

is the Brunt-Väisälä frequency. In a stably stratified atmosphere \(d\theta_0/dz > 0\), in an unstably stratified atmosphere \(d\theta_0/dz < 0\), and if \(d\theta_0/dz = 0\) the atmosphere is said to be neutrally stable. The Brunt-Väisälä frequency is thus a measure of the static stability of the atmosphere. An air parcel that is brought out of static equilibrium will experience a restoring force \(\rho N^2 \delta z\) per unit volume. This force tends to bring the parcel back to its equilibrium position. The general solution of equation (1.10) \(\delta z = \exp(\pm iNt)\), describes a harmonic oscillation at the Brunt-Väisälä frequency \(N\). The air parcel moves in a straight line. The oscillating motion corresponds to so-called buoyancy waves.

### 1.4.2 Pure inertial waves

We will consider now an air parcel that is displaced horizontally in a geostrophic flow. In a geostrophic flow horizontal pressure gradients are counteracted by (Coriolis) forces due to the rotation of the earth.

Consider a background flow \(u_0\) in the \(x\)-direction on an \(f\)-plane\(^1\) that is in geostrophic equilibrium. Assume as in the previous section that any disturbance from this equilibrium does not affect the pressure. The motion is then governed by

\[
\frac{du}{dt} - fv = 0, \tag{1.12}
\]

\[
\frac{dv}{dt} - f(u - u_0) = 0, \tag{1.13}
\]

where \(u\) and \(v\) correspond to the parcels velocity in \(x\) and \(y\)-direction, respectively, and \(f\) is the Coriolis parameter. A parcel that is moving with the flow has at location \(y\) a velocity \(u_0(y)\). If the parcel is displaced across the flow over a distance \(\delta y\) then its new velocity at \(y + \delta y\) can be found from the integration of (1.12) and using the fact that \(v = dy/dt\) i.e.

\[
u(y + \delta y) = u_0(y) + f \delta y. \tag{1.14}
\]

\(^1\)The \(f\)-plane approximation assumes that meridional motions are small enough to neglect the latitudinal variation in \(f\). Furthermore the curvature of the earth is also neglected in favor of a less complicated Cartesian geometry. See e.g. Holton (1992) for details.
Some basic features of inertia-gravity waves

The velocity of the flow $u_0$ at the parcel's new location can be approximated by

$$u_0(y + \delta y) = u_0(y) + f \frac{\partial u_0}{\partial y} \delta y.$$  \hfill (1.15)

Now (1.13) can be evaluated at $y + \delta y$ and using (1.14) and (1.15) one obtains the following equation of motion

$$\frac{d^2 \delta y}{dt^2} = -\omega_I^2 \delta y,$$  \hfill (1.16)

where

$$\omega_I^2 = f \frac{\partial \Pi}{\partial y}, \quad \text{and} \quad \Pi = fy - u_0.$$  \hfill (1.17)

Note that equation (1.16) is mathematically identical to equation (1.10). The behaviour of the air parcel when it is brought out of equilibrium now depends on the sign of $f \frac{\partial \Pi}{\partial y}$ or equivalently $f(f - \partial u_0/\partial y)$. If $f(f - \partial u_0/\partial y) > 0$ the air parcel will perform an oscillation around its initial position $y$. This situation is inertially stable. An inertially unstable situation occurs whenever $f(f - \partial u_0/\partial y) < 0$. The air parcel will then move exponentially from its initial position $y$. The air parcel remains at rest at its new position $y + \delta y$ for the inertially neutral case where $f(f - \partial u_0/\partial y) = 0$. The oscillating case described above corresponds to so-called inertial waves. The air parcel moves in a circular orbit. The case in which the flow $u_0$ is inertially unstable is not very common in the atmosphere. Only in the tropical atmosphere where $f$ is small, the criterion for inertial instability is quite easily met.

1.4.3 Inertia-gravity waves

In case an parcel displacement is restored by buoyancy as well as by the effect of earth rotation, then both mechanisms described in the previous two sections play a role (Figure 1.13). If an air parcel moves a distance $\delta s$ under an angle $\alpha$ with respect to the vertical, then its restoring force in the vertical will be equal to $\rho \delta s \cos^2 \alpha$ per unit volume. The corresponding restoring force in the horizontal is equal to $\rho f^2 \sin^2 \alpha \delta s$ per unit volume. From Figure 1.13 it can be derived that the motion of an air parcel now obeys the following equation

$$\frac{d^2 \delta s}{dt^2} = -\omega_{IGW}^2 \delta s,$$  \hfill (1.18)

where

$$\omega_{IGW}^2 = f^2 \sin^2 \alpha + N^2 \cos^2 \alpha.$$  \hfill (1.19)
2.2 Introduction

\[ \theta - N \cos^2 \alpha \delta I + (d \theta / dz) \delta z \]

\[ \Phi + (d \Phi / dy) \delta y \]

\[ \rho_0 \delta \sin^2 \alpha \delta \]

\[ -\omega^2 \sin^2 \alpha \delta \]

\[ -N^2 \cos^2 \alpha \delta \]

\[ \theta_0 \]

Figure 1.13: Situation for the case an air parcel is displaced in the \( y/z \) plane over a distance \( \delta s \). The displacement induces a restoring buoyancy force \( \rho N^2 \cos^2 \alpha \delta z \) and a restoring force due to the earth rotation \( \rho \omega^2 \delta y \). Symbols are defined in the text.

In the derivation of (1.18) it was assumed that the flow \( u_0 \) had no shear. In the atmosphere \( N^2 \) is of the order of \( 10^{-4} \text{s}^{-2} \), and \( f^2 \) is of the order of \( 10^{-8} \text{s}^{-2} \), thus from (1.19) it follows that wave frequencies \( |\omega_{GW}| \) are between \( f \) and \( N \). The air parcel now moves in elliptical orbits. The oscillations that are described by (1.18) are called inertia-gravity waves, indicating that the propagation mechanism of these waves is due to the effects of both earth rotation (inertia) and buoyancy (gravity).

1.5 Central issues of this thesis

The degree of permeability of the Antarctic polar vortex edge is an important issue. As was shown in the previous sections, the Antarctic ozone hole persists over several months, which indicates that the vortex edge is a very effective barrier to transport of warm ozone-rich midlatitude air into the polar vortex. Leakage of very cold ozone-poor air from within the polar vortex towards midlatitudes might contribute to the observed downward trend in midlatitude ozone. The last 15 years or so, a large number of studies using different approaches, have provided a qualitative picture of the permeability of the Antarctic polar vortex due to large-scale processes. Only very few studies have investigated in a quantitative way the amount of exchange across the vortex edge. Also, the effect of smaller scale inertia-gravity waves has been only superficially studied. The main objective of the work presented in this thesis was to investigate the interaction of inertia-gravity waves with the Antarctic polar vortex and to quantify their effect upon the permeability of the polar vortex edge. The following scientific questions will be addressed:
Central issues of this thesis

1. How well is the polar vortex interior isolated from its surroundings?

2. How much air leaves or enters the polar vortex horizontally through the edge and how much air escapes through the vortex bottom?

3. What are the seasonal and interannual variations in the exchange?

4. What is the effect of inertia-gravity waves upon the leakage of air from the polar vortex and how does the leakage depend on the wave parameters?

5. How do inertia-gravity waves interact with idealized background flows that mimic the polar vortex edge?

The approach to tackle these questions is twofold. First of all, the permeability of the Antarctic polar vortex edge has been investigated using the trajectory model of the Royal Netherlands Meteorological Institute (KNMI). In chapter 2 the first three questions are addressed. The permeability of the vortex edge due to large-scale processes has been investigated, using high resolution wind and temperature analyses from the ECMWF. The use of such high resolution is unique in the study of the permeability of the vortex edge. Several experiments were performed of which the results shed a new light on the amount of quasi-horizontal and vertical exchange of air from the polar vortex towards midlatitudes and vice versa. Chapter 2 appeared as Öllers et al. (2002b).

Chapter 3 addresses question four. In this chapter, the effect of a prescribed inertia-gravity wave field that is superimposed upon the ECMWF wind fields is investigated, again using the trajectory model. Numerical experiments were performed for several wave parameters and for a period that (breaking) planetary waves did not contribute to leakage from the polar vortex. In this way, better estimates of inertia-gravity wave induced exchange could be calculated. Chapter 3 has been submitted as Öllers et al. (2002c).

Besides the use of a trajectory model, the effect of inertia-gravity waves upon the permeability of the vortex edge was studied by using analytical-numerical models. In chapter 4 some simple analytical linear wavemodels are presented. The model equations are based on the same set as used later in chapter 5. The simplicity of the models finds expression in the fact that the phaselines of the wave are perpendicular to the direction of the background wind. Nevertheless the models allow to be completely solved analytically. They provide some basic insight into the propagation of waves in simple horizontally sheared background flows and serve as preliminary studies for the more advanced model in chapter 5.

In this chapter 5, question 5 will be addressed. The propagation of inertia-gravity waves in a barotropic shear layer such as the polar vortex edge is studied. From
the fluid dynamical equations a wave equation will be derived that describes this propagation. The wave equation contains two singularities of a different kind and will be analysed in detail. Depending on the wave parameters a wide range of interesting wave behaviours was found. Chapter 5 appeared as Öllers et al. (2002a).
Chapter 2

A study of the leakage of the Antarctic polar vortex in late austral winter and spring using isentropic and 3-D trajectories

The permeability of the Antarctic polar vortex has been investigated in late austral winter and spring by comparing isentropic and three-dimensional (3-D) trajectories. Trajectory computations are performed with the help of the Royal Dutch Meteorological Institute (KNMI) trajectory model using data from the European Centre for Medium-range Weather Forecasts (ECMWF) from August to November 1998. Large numbers of air parcels are initially released inside and outside the polar vortex on the 350 K, 450 K and 550 K isentropic surfaces. They are integrated 4 months forward in time in an isentropic mode as well as in a 3-D mode that uses all three wind components from the ECMWF and takes into account diabatic heating and cooling effects. For the isentropic trajectory calculations very little transport (0.37%/week) is found for August and September, while October and November give somewhat higher transport rates (1.95%/week). The 3-D trajectory calculations for October give much more exchange between the vortex and midlatitudes than the isentropic ones, due to a significant number of parcels that descend inside the vortex. Descent rates are calculated for 350 K (October), 450 K (August-October) and 550 K (October). Overall the results show that 3-D trajectories will provide more accurate leakage rates than the isentropic ones. Also, despite the large scale mixing in the polar vortex or in midlatitudes, little ozone depleted air leaks from the ozone hole into the midlatitude stratosphere.
2.1 Introduction

Each austral winter and spring the ozone hole develops in the Antarctic stratosphere. The Antarctic polar vortex creates the unique dynamical environment for the ozone hole to develop and persist over the months August to November (Waugh et al., 1999; Zhou et al., 2000). The destruction of ozone within the polar vortex is the result of heterogeneous chemistry (Solomon et al., 1986; Solomon, 1999). Due to the isolation of the polar vortex, air within the vortex interior cools and temperatures can drop as low as 185 K, cold enough for the formation of polar stratospheric clouds (PSCs). Although PSCs have not yet revealed all of their mysteries concerning their composition and chemistry, a key point is that they provide a reactive surface for catalytical chemical reactions that destroy ozone (Tolbert and Toon, 2001).

The persistence of the ozone hole and the cold temperatures in the interior over several months, indicate that there is little exchange of air between midlatitudes and the vortex interior. The amount of exchange is an important issue for the following reasons. On the one hand, transport of warm and ozone-rich midlatitude air into the vortex would weaken it and slow down the heterogeneous processing of ozone. On the other hand transport of very cold ozone depleted air into midlatitudes might contribute to the observed downward trend in midlatitude ozone (Harris et al., 1997).

Nowadays the degree of isolation and thus the amount of transport and mixing between midlatitude and vortex air across the vortex edge is still under debate. The central question in this discussion is whether the air in the polar vortex is captivated on a seasonal timescale. Some previous studies (e.g. Hartmann et al., 1989a,b; Schoeberl et al., 1992) have shown that there is only little exchange of air between midlatitudes and the vortex interior due to strong gradients in potential vorticity (PV) in the horizontal and small cooling rates in the vertical (‘containment vessel’ hypothesis). Others (e.g. Proffitt et al., 1989; Tuck, 1989) have argued that there is a substantial downward flow of air through the vortex interior and then outwards into the lower polar and midlatitude stratosphere (‘flowing processor’ hypothesis).

In order to find a possible way out of this controversy various theoretical and numerical studies have been performed over the years. Juckes and McIntyre (1987) and McIntyre (1989) argue that on the basis of Ertel’s potential vorticity on isentropic surfaces the polar vortex should behave like an isolated material entity. Erosion of the polar vortex edge due to breaking planetary waves steepens its strong latitudinal gradient in PV.

Isentropic trajectory calculations (e.g. Bowman, 1993a,b; Dalhberg and Bowman, 1994; Bowman and Chen, 1994; Bowman, 1996) as well as experiments using the high resolution Contour Advection with Surgery technique (CAS) on isentropic surfaces (Waugh et al., 1994b; Chen et al., 1994; Chen, 1994) show that the polar vortex is nearly impermeable for isentropic motions above the 425 K isentropic surface.
Introduction

Below this level the polar vortex is less isolated. Below the 400 K isentropic surface a region indicated as the 'sub-vortex' region (McIntyre, 1995) can be identified where substantial isentropic mixing of vortex air with lower stratospheric midlatitude air takes place.

Although no substantial cross-edge transport of air is observed in these studies, air from the outer vortex can get torn off from the main vortex by breaking planetary waves and become organized into thin filaments. Mixing between the air in the filaments and surrounding midlatitude air occurs when the spatial scale of the filaments becomes small enough for diffusive mixing to become important. The inner vortex does not show these features and resists intrusions of air from midlatitudes (Polvani and Plumb, 1992). A drawback of the above studies is that they mostly consider adiabatic (isentropic) motions only. Also, they perform model calculations over periods of weeks to months without questioning whether these motions can still be regarded as being isentropic over such long time spans.

The STRATEOLE project (Vial et al., 1994) has been initiated to gain a better understanding of the dynamics and chemistry of the Antarctic polar vortex. In the near future isopycnic balloons will be launched inside and at the edge of the vortex in late winter and spring. Trouniday et al. (1995) performed 3-D trajectory computations using winds from a rather coarse-gridded (5° × 5°) 3-D stratosphere-mesosphere model. For an approximately 10 week period they recorded the number of crossings of air parcels and isopycnic balloons across the vortex edge which were initially released in a zone between 50° and 75°S. They demonstrated the resemblance between the horizontal mixing of air parcels and balloons.

Wauben et al. (1997a) performed exchange calculations using a 3-D Eulerian tracer transport model driven by ECMWF winds. For tracers initially released at 72.5 hPa inside the polar vortex they calculated exchange rates averaged over the four year period 1990-1993. They arrived at a quasi-horizontal cross-edge transport of 0.24%/day while 0.83%/day of the vortex mass descended into the troposphere. They stated that 65% of the total tracer mass is flushed out during August-October. A shortcoming of their model is that the coarse resolution (5.0 × 3.75°) may lead to additional numerical leakage out of the vortex.

In this study a comparison is made between isentropic and 3-D trajectory computations in the polar vortex region, which has not been presented before. Also, new quantitative estimates of transport across the polar vortex edge will be obtained from high resolution (1° × 2°) trajectory calculations. Such a resolution is much higher than what was used in studies of e.g. Trouniday et al. (1995) and Wauben et al. (1997a). The validity of the isentropic approximation in exchange studies of the Antarctic polar vortex will be discussed. Also, average descent rates will be determined from the
3-D trajectory computations. These values will be compared to average descent rates of tracer isopleths and values reported in the literature.

2.2 Trajectory model

Trajectories are calculated with the KNMI trajectory model (Scheele et al., 1996). The model computes the three dimensional displacement of air parcels with the iterative scheme of Petterssen (1940) and a time step of $\delta t = 10$ min. The trajectory model uses ECMWF 6 hour forecasts (first-guess fields) of the three dimensional wind and temperature. First-guess data are preferred to analysis data because they show better physical balance of the wind and mass density fields. Analysed quantities are slightly out of balance due to the recent addition of new observations in the analysis step. The input data are interpolated linearly in the horizontal and with $\log(p)$ in the vertical to the instantaneous locations of the trajectories. The ECMWF data are available on a horizontal $1^\circ \times 1^\circ$ grid at 31 hybrid pressure-sigma model levels of which 10 are in the Antarctic stratosphere. The highest model level in the stratosphere is 10 hPa. In this study we use these data for the period of August 1 1998, 12 GMT to November 30 1998, 12 GMT in order to allow comparison to Wauben et al. (1997a).

The accuracy of the trajectory calculations is influenced by errors in the numerical trajectory integration scheme and errors in the wind fields. The errors due to the numerical scheme are of the order of 1% or less (Stohl et al., 2001). However, the study of Stohl et al. (2001) focusses on 3-D trajectories in the troposphere. In this study we also deal with isentropic as well as with 3-D trajectories in the stratosphere. The relative error in the final position of air parcels for these types of trajectories in the stratosphere due to errors in the wind field is determined as follows. An air parcel is integrated forward in time from its initial position. From its final position it is then integrated backward in time. It is expected that an air parcel that is integrated backward in time from its final position will not exactly return to its initial position. It turns out that the positioning error relative to the travelled distance error is 5% or less for isentropic trajectories and 2% or less for 3-D trajectories (Rinus Scheele personal communication, 2001). Errors in the trajectories will therefore be mainly determined by errors in the wind field as has already been suggested by Bowman (1993a).

2.3 Experimental setup

Air parcel trajectories are calculated both in the isentropic mode and in the 3-D mode. The isentropic mode conserves potential temperature ($\theta$) along the trajectories. Air parcels that are initially on an isentropic surface are tied to that surface explicitly.
Experimental setup

by the model. The 3-D mode takes into account vertical air motions caused by diabatic heating and cooling effects. These effects become especially important in early Southern Hemisphere spring (October-December), when the sun rises and the Antarctic atmosphere starts to warm. The air parcels are started on the 350, 450 and 550 K isentropic surfaces, because these levels represent the lower, middle and upper vortex (~13-25 km) levels, respectively. This region also covers the altitudes at which ozone is destroyed in early winter and spring (~15-20 km) (Solomon, 1999). In all cases discussed below trajectories are started on a $\Gamma \times 2^\circ$ latitude-longitude grid and integrated four weeks forward in time.

2.3.1 Initial position of air parcels inside/outside the vortex

The initial position of the air parcels inside or outside the vortex is determined with the help of PV maps on isentropic surfaces based on ECMWF data using a method similar to the one described in Teitelbaum et al. (1998). In order to prevent the initial position of air parcels being in the vortex edge PV contours are determined for both the inner and outer vortex edges. Going from low to high PV values the gradient in PV will change abruptly twice i.e. once at the transition from the midlatitude surf zone and the (outer) vortex edge and once at the transition from the (inner) vortex edge and the vortex core. At these locations PV contours are determined, say PV\text{out} and PV\text{in}, respectively. The initial position of air parcels starting inside/outside the polar vortex is therefore restricted by the area inside/outside the critical PV\text{in}/PV\text{out} contour corresponding to the inner/outer vortex edge. For air parcels that initially start inside the polar vortex the PV\text{in} contour will be denoted as the critical PV contour. Similarly, for air parcels that initially start outside the polar vortex the PV\text{out} contour will be denoted as the critical PV contour.

2.3.2 Exchange diagnostics

An air parcel is assumed to have left or entered the polar vortex if the air parcel crosses the critical PV contour and does not cross this contour again for the next 120 hours. The time span of 120 will be denoted as the threshold residence time.

The threshold residence time is introduced to eliminate positioning errors and is determined by preliminary analyses of the trajectory runs. At regular time intervals area (isentropic runs) and mass-weighted fluxes (3-D runs) are determined for air parcels that cross a given critical PV contour. Following Dalhberg and Bowman (1994) and Seo and Bowman (2001), area and mass exchange rates are calculated by determining the initial area and volume (mass) of air parcels and the area and mass of parcels that leave or enter the vortex. For several different threshold residence times, exchange rates can be determined for intruding or extruding air parcels. In general
we expect that the area and mass exchange rates represented by parcels leaving or entering the vortex should increase with time. However for threshold times less than 96 hours, we have regularly observed a decrease in the exchange rates represented by parcels that leave the vortex. Further inspection has shown that this is due to frequent reversible parcel displacements i.e. air parcels leave or enter the vortex for a couple of hours but then return. In this context it is worth mentioning that the computation of PV along the trajectories may not be very accurate due to the dependency of PV on first order wind and temperature derivatives. There is also an uncertainty in the PV value of the critical PV contour due to the vertical interpolation of wind, temperature and pressure in the trajectory model. By increasing the threshold to 120 hours it is found that an increasing number of trajectories leave or enter the polar vortex. The advantage of our analysis method is that it combines the methods used in Dalhberg and Bowman (1994) and Seo and Bowman (2001) (area/mass-exchange rates and exchange across PV contour) and Trounday et al. (1995) (determination of the vortex edge and threshold time).

2.3.3 Isentropic mode

In the isentropic mode 12 model runs are performed in which trajectories are started separately on the 350 K, 450 K and 550 K isentropic surfaces on August 1 1998, 12 GMT, August 30 1998, 12 GMT, September 30 1998, 12 GMT and October 30 1998, 12 GMT. Different initial times are used in order to get unambiguous information about the time evolution of the exchange across the vortex edge. The month of August is typically the period of vortex buildup and development of the ozone hole. In September the polar vortex is at its strongest with decreasing ozone values and reaching minimum values in October. In the month November usually vortex breakdown takes place. Henceforth these runs will be denoted as the isentropic 'interior' runs.

In order to study the permeability of the polar vortex edge to intruding midlatitude air, two additional isentropic runs are started in which air parcels are initially outside the polar vortex on the 450 K isentropic surface on August 1 1998, 12 GMT and September 30 1998, 12 GMT, respectively. On the 350 K and 550 K isentropic surfaces the polar vortex is less well defined and it is therefore less favourable to perform the calculations on these surfaces. The choice to represent a late winter (August) and spring (October) situation is motivated by the observation that these periods showed, respectively, few and large intrusions of air from the polar vortex to midlatitudes (section 2.4.1). Henceforth these runs will be denoted as the isentropic 'exterior' runs.

The critical PV contours for different starting dates and different isentropic surfaces used in the isentropic interior runs are listed in Table 2.1. For the two isentropic
Experimental setup

Table 2.1: The critical PV contours for different starting dates and isentropic surfaces in the isentropic interior runs. The critical PV contours are determined with help of PV maps on isentropic surfaces. PV is given in potential vorticity units (PVU) (1 PVU=10^6 Km^2kg^-1s^-1).

<table>
<thead>
<tr>
<th>Date</th>
<th>350 K</th>
<th>450 K</th>
<th>550 K</th>
</tr>
</thead>
<tbody>
<tr>
<td>98080112</td>
<td>-6.0</td>
<td>-19</td>
<td>-50</td>
</tr>
<tr>
<td>98083012</td>
<td>-6.0</td>
<td>-20</td>
<td>-43</td>
</tr>
<tr>
<td>98093012</td>
<td>-6.5</td>
<td>-20</td>
<td>-50</td>
</tr>
<tr>
<td>98103012</td>
<td>-6.5</td>
<td>-21</td>
<td>-52</td>
</tr>
</tbody>
</table>

Table 2.2: Numbers of air parcels starting at different dates and isentropic surfaces in the isentropic interior runs. Example: On 98083012, 4630 parcels start on the 450 K isentropic surface.

<table>
<thead>
<tr>
<th>Date</th>
<th>350 K</th>
<th>450 K</th>
<th>550 K</th>
</tr>
</thead>
<tbody>
<tr>
<td>98080112</td>
<td>2014</td>
<td>4698</td>
<td>4128</td>
</tr>
<tr>
<td>98083012</td>
<td>2862</td>
<td>4630</td>
<td>4468</td>
</tr>
<tr>
<td>98093012</td>
<td>2261</td>
<td>4098</td>
<td>4080</td>
</tr>
<tr>
<td>98103012</td>
<td>2575</td>
<td>4366</td>
<td>3834</td>
</tr>
</tbody>
</table>

Trajectories are calculated for several thousands of parcels. In Table 2.2 the number of parcels that start in the isentropic interior runs for four subsequent starting dates and isentropic surfaces are summarized. Note that in each run a different number of parcels starts. This is because the shape and area of the polar vortex changes in time as well as with height. For the isentropic exterior run 4978 parcels start on 98080112 and 5546 air parcels start on 98093012 at 450 K.

2.3.4 3-D mode

From the results of the isentropic interior runs it turns out that the highest leakage rates are found in October. In order to see if leakage is still larger in the 3-D mode it is decided to start a run firstly on 98093012. Again air parcel trajectories are started inside the vortex on the 350 K, 450 K and 550 K isentropic surfaces.

1Henceforth dates are denoted as a year/month/day/hour combination. For example: 98080112 corresponds to August 1 1998, 12 GMT.
A study of the leakage of the Antarctic polar vortex

Table 2.3: Exchange rates in percent per week for the isentropic interior runs computed for August-November 1998. Exchange rates are calculated after a four week trajectory integration. Starting dates are listed above for the 350, 450 and 550 K level.

<table>
<thead>
<tr>
<th>Date</th>
<th>350 K</th>
<th>450 K</th>
<th>550 K</th>
</tr>
</thead>
<tbody>
<tr>
<td>98080112</td>
<td>0.00</td>
<td>0.00</td>
<td>0.37</td>
</tr>
<tr>
<td>98083012</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>98093012</td>
<td>0.33</td>
<td>1.34</td>
<td>1.95</td>
</tr>
<tr>
<td>98103012</td>
<td>0.40</td>
<td>0.08</td>
<td>1.16</td>
</tr>
</tbody>
</table>

For the 3-D run that starts on 98093012, the critical PV contours on the 350, 450 and 550 K level were at -6.5 PVU, -20 PVU and -50 PVU, respectively. Additionally the critical PV contours are determined at 340 K, 400 K, 500 K, 600 K, 650 K, 700 K, 750 K, 800 K and 850 K for 98093012, 98101012, 98102012 and 98103012. At intermediate levels and times a linear interpolation scheme is applied.

For the 3-D trajectory calculation that starts on 98093012, the same air parcels are started on the 350, 450 and 550 K isentropic surface on 98093012 as in the isentropic case (Table 2.1). For the additional runs that start on 93080112, 93083012, 93093012, respectively, 3603, 4253 and 4481 air parcels are initially started inside the vortex at 450 K. For the trajectory runs that start on 98080112 and 98083012, respectively 4680 and 4661 air parcels are initially started inside the vortex at 450 K.

2.4 Isentropic trajectory results

2.4.1 Isentropic interior runs

For each date and level and at regular time intervals, the number of air parcels that have crossed the critical PV contour and obey the critical residence time is calculated. From this, average transport rates expressed in percent per week can be determined. The results are summarized in Table 2.3. It can be seen that there is no or only very little transport out of the polar vortex in August and September on all three isentropic surfaces, indicating a nearly impermeable polar vortex edge. This is confirmed by the strong PV gradients marking the edge of the polar vortex throughout August and September (Figure 2.1). The strongest PV gradients are found at 450 K. At 550 K a clear vortex edge can be identified although features of weak filamentation are also
Figure 2.1: Potential vorticity (0.1 PVU) at 350 K on 98080112 (upper left) and on 98083012 (upper right) and at 450 K on 98080112 (middle left) and on 98083012 (middle right) and at 550 K on 98080112 (lower left) and on 98083012 (lower right) based on ECMWF fields. The shaded contours mark the vortex edge.
Figure 2.2: (a) PV at the start (t=0) plotted against the PV after three weeks (t=504) for the isentropic trajectories that start on 98080112 at 350 K. (b) The same as in (a) but for the run that starts on 98083012 at 350 K. The dashed lines mark the location of the critical PV contour.

present. This may explain the somewhat higher exchange rate in August at 550 K. It is remarkable that virtually no air parcels have crossed the critical PV contour in August and September at 350 K. Almost all air parcels released at 350 K on 98080112 and 98083012 stay within the region marked by the critical PV contour, as can be seen from Figure 2.2. The parcels are well mixed inside the area confined by the critical PV contour. Note however from Figure 2.1, that at 350 K on 98080112 and 98083012 the vortex edge is not so well defined. There is more than one closed contour for the critical PV and the air parcels confined by it may well reach into midlatitudes.

These results mostly confirm those of previous studies (Bowman, 1993a; Chen, 1994; Chen et al., 1994) using isentropic trajectory calculations and contour advection techniques. In those studies nearly complete isolation of the polar vortex is found above approximately 425 K. Chen (1994) concluded from his contour advection calculations at 350 K and 375 K that the air inside the vortex is very well mixed on August 31, 1993 after 40 days of integration. Most of this air is still confined within the 'line of separation' separating the inner and outer vortex. If we tentatively identify Chen’s ‘line of separation’ with our critical PV contour, then our findings are in qualitative correspondence with those of Chen at 350 K. He suggested that the low exchange rates at 350 K are related to the strong PV gradients marking the tropopause obstructing air parcels to intrude into the troposphere. Slightly larger leakage rates are found for October and November, especially at 550 K. As noted above during August and September and also in October and November there is an
increased intensity of filamentation at 550 K. Several filaments are wrapped around the vortex. Subsequent mixing with midlatitude air may transfer vortex air into mid-latitudes. The filaments develop at locations where the PV gradient is weak. This facilitates the exchange of air between the polar vortex and midlatitudes. Figure 2.3 gives an example of a situation where air parcels are organized in a filament. It shows the distribution of air parcels that have left the polar vortex after 240 hours since the start on 98093012 on the 550 K isentropic surface. The corresponding PV distribution on 98101012 at 550 K is shown in Figure 2.4. From this figure two filaments wrapped around the main vortex can clearly be identified. The locations of a major part of the air parcels shown in Figure 2.3 coincides with the filament in Figure 2.4 located between approximately 60° E-120° E and 40° S-50° S. This filament develops soon after the start on 98093012 near a location with relatively weak PV gradients and most air parcels shown in Figure 2.3 leave the vortex here. The filament persists for most of the time between the start and 98101012. The other filament located between approximately 90° W-180° W and 30° S-40° S develops at a later time making it less likely for air parcels to have left the vortex through this filament on 98101012.
Note that the PV gradients near the location where the filaments develop are weaker than elsewhere near the vortex edge. The large filaments observed in October at 550 K are less pronounced in November at 550 K and in October and November at 450 K. On the 350 K isentropic surface the polar vortex is not so well-defined from August to November and therefore it is hard to distinguish any filaments.

Figure 2.4: PV on the 550 K isentropic surface on October 10, 1998

2.4.2 Isentropic exterior runs

To investigate possible intrusions of air parcels from midlatitudes into the polar vortex, two trajectory runs are performed in which the air parcels are initially located outside the polar vortex on the 450 K isentropic surface (see section 2.3.3). For the run starting on 98080112 we do not find a single air parcel that enters the polar vortex, which would correspond to a perfectly impermeable vortex edge for exterior air parcels. In the previous section it is noted that in August at 450 K there is no air parcel that crosses the vortex edge starting from the interior. This indicates that in August at 450 K the polar vortex edge acts as a perfect barrier to quasi-horizontal mixing.
For the trajectory run starting on 980930 at 450 K, an exchange rate of 0.20% per week is found. Figure 2.5 shows the locations of air parcels after three weeks (504 hours) that have crossed the vortex edge, as well as those that have stayed outside the vortex. Parcels that have entered the vortex are initially quite close to the outer edge of the vortex. Obviously, these parcels would have the highest probability of entering the vortex core. Parcels that start well outside the polar vortex are well-mixed at midlatitudes. Note that the exchange rates in October at 450 K for parcels entering the polar vortex from midlatitudes are even lower than those for parcels moving from the vortex core to midlatitudes (Table 2.3). This could be a confirmation of previous suggestions made by Juckes and McIntyre (1987) and Polvani and Plumb (1992), concerning the ‘one-sidedness’ of air mass intrusions. They stated that the outer vortex edge is more robust to exchange of air from midlatitudes into the vortex core, than the inner vortex edge is to exchange of air from the vortex core into midlatitudes. They attribute this to the absence of breaking planetary waves near the inner vortex. At this point it is interesting to note that our results suggest that although there is very little cross-edge mass exchange in August and September on the 350, 450 and 550 K levels, mixing inside the vortex core itself is significant. A clear example of this is shown in Figure 2.2 for 350 K and Figure 2.6 for 450 and 550 K. A region of strong mixing inside the vortex core and at midlatitudes, separated by a zone of weak mixing and little mass exchange corresponding to the vortex edge has also been found in a recent study of Lee et al. (2001).

2.5 3-D trajectory results

In section 2.4 the largest exchange rates are found for isentropic trajectories starting inside the vortex on 980930 at 350 K, 450 K and 550 K (Table 2.3). In order to investigate whether diabatic effects change the leakage, 3-D trajectories are calculated starting from the same locations on 980930 at 350 K, 450 K and 550 K. In this 3-D mode the air parcels can traverse isentropic surfaces due to vertical movements related to diabatic cooling or heating processes. As a result, we can not only quantify quasi-horizontal cross-edge transport (as in the isentropic case) but also vertical exchange through the vortex ‘bottom’ and ‘top’. The top of the vortex in our model runs is restricted by the upper level of the ECMWF model corresponding to the 10 hPa isobaric level (θ ≈ 850 K). No air parcels can leave the vortex through the top level. The bottom of the vortex is determined using PV maps on isentropic surfaces. We search for an isentropic surface where there is still a PV gradient present marking the edge of the vortex. It turns out that below the 340 K isentropic surface no clear vortex edge structure is visible anymore. Therefore the bottom of the vortex is taken to be at the 340 K isentropic surface. Air parcels descending below 340 K might eventually
Figure 2.5: Location of air parcels that have intruded into the polar vortex after 504 hours since the start (top) and the location of parcels that do not enter the vortex core (bottom). Parcels that have entered the polar vortex after 504 hours have passed there already at least 120 hours in correspondence with the threshold residence time.
enter the troposphere. The troposphere can be assumed to start below the -3.5 PVU contour (approximately 307 K). This altitude has been derived from ECMWF PV/$\theta$ cross-sections for October 1998. The quasi-horizontal and vertical exchange rates are summarized in Table 2.4. One should be cautious in comparing directly the quasi-horizontal transport of the 3-D trajectories and the isentropic trajectories (Table 2.3). The leakage rates in Table 2.4 are mass leakage rates whereas the leakage rates in Table 2.3 correspond to area leakage rates (see also section 2.3.4). Tentatively, we might state that cross-edge leakage rates of the 3-D trajectories are somewhat larger

Figure 2.6: PV at the start (t=0) plotted against the PV after three weeks (t=504) for the isentropic trajectories that start on 98080112 and 98083012 at 450 K and 550 K. The dashed lines mark the location of the critical PV contour.
Table 2.4: Exchange rates in percent per week for the 3-D run after a four week integration starting on 98093012 on the 350 K, 450 K and 550 K isentropic surfaces.

<table>
<thead>
<tr>
<th></th>
<th>350 K</th>
<th>450 K</th>
<th>550 K</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-edge transport</td>
<td>0.62</td>
<td>1.27</td>
<td>2.42</td>
<td>4.31</td>
</tr>
<tr>
<td>Descending below 340 K</td>
<td>4.11</td>
<td>0.02</td>
<td>0.00</td>
<td>4.13</td>
</tr>
</tbody>
</table>

compared to the purely isentropic mode, except at 450 K where the leakage rate in the 3-D mode (1.27 %/week) is slightly smaller than for the isentropic mode (1.34 %/week).

The greater exchange in the 3-D mode is not unreasonable, as air parcels are no longer restricted to conserve potential temperature and PV. Parcels may now descend to a height where there is an increased chance of crossing the critical PV contour and may move to midlatitudes. At higher levels (450 K and 550 K) the polar vortex still acts as an almost impermeable barrier to quasi-horizontal cross-edge transport. Vertical exchange rates are largest for the trajectories starting at 350 K and very low for 450 K and 550 K. For all three levels, there are no air parcels that enter the troposphere. It is interesting to note that although on average air parcels descend within the vortex, a significant number of parcels move to higher levels than where they have started, as is shown in Figure 2.7. Rising air parcels are mainly found near the edge of the polar vortex. A possible explanation for this is that in spring (October) air parcels near the edge are likely to be subjected to heating.

2.6 Validity of the isentropic approximation

The results in section 2.5 show that the 3-D trajectory computations give rise to additional transport across the vortex edge in comparison to isentropic exchange rates (section 2.4). In other studies of the polar vortex using isentropic trajectories (Bowman, 1993a; Bowman and Chen, 1994) and the CAS-technique (Chen et al., 1994; Chen, 1994) mixing and mass transport is determined on isentropic surfaces after several weeks to months of model integration. In these studies it is not questioned whether motions can still be regarded as being adiabatic after such long periods. In order to test the validity of the isentropic approximation after long periods of integration, a number of 3-D trajectories have been selected that are initially located well inside the vortex at -90°W and 88°S, 84°S, 80°S, 78°S, 72°S, 68°S, and 64°S at 350 K, 450 K and 550 K. The trajectories are integrated four weeks forward in time starting on 98093012. The change in potential temperature θ along the trajectories as a function of time is shown in Figure 2.8. On all three isentropic surfaces trajec-
Validity of the isentropic approximation

Figure 2.7: Distribution of air parcels after 4 weeks since the start on 98093012 at 350, 450 and 550 K.

...start to deviate significantly (5-10 K) from their initial isentropes in less than 48 hours after the start. Qualitatively the same behaviour is found for the months of August, September and November. Note that the deviation from the initial isentropes is largest for air parcels that start on the 450 and 550 K isentropic levels. Also, in most cases the air parcels move to lower isentropic levels but as time increases their rate of descent decreases. A possible explanation for this is that the descent rates within the polar vortex are not the same on different isentropic levels. Manney et al. (1994) have shown that vertical cooling rates within the polar vortex are larger at higher isentropic levels (larger potential temperature). Motivated by their results, average cooling rates are determined on the 350, 450 and 550 K isentropic levels for air parcel trajectories that start on 98093012 (see section 2.5). The largest descent rate is found for air parcels that start at 550 K (2.2 K/day) and smaller rates for those that start at 450 K (0.6 K/day) and 350 K (0.3 K/day). Therefore, air parcels that start
Figure 2.8: Potential temperature $\theta$ as a function of time since the start ($t=0$). Trajectories are started on 98093012 at 350, 450 and 550 K.
initially at 550 K will experience the largest descent but this will gradually decrease as they move to lower isentropic levels were the descent rates are smaller. This effect becomes less pronounced for air parcels that start at 450 K and 350 K. It is interesting to note from Figure 2.8, that for individual air parcels descent rates in $\theta$ can be of the order 5 K/day or less. Obviously diabatic processes do play an important role in the polar vortex even on short timescales.

2.7 Comparison to Eulerian exchange estimates

Our results form an interesting contrast with the findings of Wauben et al. (1997a) (W97a hereafter). They performed calculations with a global tracer transport model using the same ECMWF data as in the present study but for August-October of the years 1990-1993. Tracers were released on August 1, 1990-1993 on the 72.5 hPa isobaric level ($\theta \approx 420$ K) inside as well as outside the polar vortex. Exchange rates were then calculated for the second half of the integration period i.e. from the second half of September to the end of October. W97a calculated quasi-horizontal exchange using a fictitious vortex boundary at 61°S$^2$. They determined vertical exchange rates for tracers that entered the atmosphere below the 275 hPa isobaric level, which they designated as the 'troposphere'.

The quasi-horizontal and vertical exchange rates of W97a (their table 1) can be compared with our results at 450 K. To allow such a comparison additional 3-D trajectories are calculated for August-October 1993 and for August and September 1998 for air parcels that are initially located inside the vortex.

The results obtained for August-October 1993 at 450 K allow a direct comparison with the results of W97a. In our case very little cross-edge transport is observed for August (0.03 \%/week) and September (0.03 \%/week) and somewhat more for October (0.44 \%/week). The total contribution to cross-edge transport for the period August-October 1993 (0.50 \%/week) is much smaller than what was found by W97a (1.19 \%/week) for the same period and year. In our case there are no air parcels that descended below 340 K for August-October compared to 5.18 \%/week in W97a for tracers descending below 275 hPa ($\approx 300$ K).

For the months of August and September 1998 at 450 K, we find no quasi-horizontal transport across the vortex edge at all. Hence, in our case, the only contribution to cross-edge transport at 450 K stems from October 1998. Since there is now no direct comparison possible with W97a, our results will be compared with their mean exchange rates. The mean quasi-horizontal exchange rate of W97a (1.68 \%/week) is only slightly larger than ours (1.27 \%/week) at 450 K in October. In our

$^2$In W97a the vortex boundary was incorrectly stated to have been at 51°S (Tuck and Proffitt, 1997; Wauben et al., 1997b)
A study of the leakage of the Antarctic polar vortex

runs for August-October 1998 very few parcels that started at 450 K descended below 340 K in August (0.00%/week), September (0.01%/week) and October (0.02%/week). W97a found a mean rate of 5.81%/week for tracers descending below 275 hPa (≈ 300 K).

W97a found an average radiative cooling rate for August-October 1993 of about 0.3 K/day in θ and stated that their values are in good agreement with cooling rates calculated by detailed radiative transfer models (Hartmann et al., 1989b; Rosenfield et al., 1994). Our calculations at 450 K show average cooling rates of 0.48 K/day in θ for August-October 1993 and 0.67 K/day for August-October 1998 which is in closer agreement with the values found by Manney et al. (1994) and recently Hicke et al. (1998). Kawamoto and Shiotani (2000) calculated vertical descent rates in the Antarctic polar vortex from long-lived trace gas data provided by the Halogen Occlusion Experiment (HALOE) and come to values of 1.1-1.9 km/month which is in correspondence with our results (0.48 and 0.67 K/day ≈ 1.2 and 1.3 km/month, respectively).

Much larger descent rates (1.75 K/day) are found by Proffitt et al. (1989) derived from aircraft measurements which are necessarily only available for limited domains in space and time. In this context it is interesting to note that in section 2.6 we also found large descent rates of 5 K/day or less for individual air parcels. W97a may have underestimated vertical motions due to their use of a relatively coarse horizontal resolution.

From the above context it is also interesting to note that, despite the overestimation of the cross-edge transport in W97a, they have found a clear year-to-year variability in the quasi-horizontal exchange rates, with larger exchange in so-called even years (1990 and 1992) and smaller rates in odd years (1991 and 1993). They attributed this to fluctuations in planetary wave activity. Shindell et al. (1997) have shown that the interannual variations in the severity of the Antarctic ozone hole in winter are driven by variations in tropospheric wave activity. In our case smaller cross-edge exchange rates are also found for the odd year 1993 and larger rates for the even year 1998. Also, our results have shown that the average cooling rate for 1993 is smaller than that for 1998, which is in agreement with recent observations of Kawamoto and Shiotani (2000), who identified an interannual oscillation in the vertical descent rates with smaller rates in the odd years (1993, 1995 and 1997) and larger rates in the even years (1992, 1994 and 1996). They have given dynamical evidence for the role of planetary waves in the interannual variability of vertical descent rates and show that planetary wave activity is more severe in even years than in odd years.


2.8 Conclusions

In this paper the permeability of the Antarctic polar vortex edge in late austral winter and spring has been investigated in a quantitative sense. We have provided new quantitative estimates on the amount of transport across the Antarctic polar vortex edge. Also, isentropic and 3-D trajectory results have been compared with each other. The validity of the isentropic approximation for trajectories in the Antarctic vortex region has been investigated.

The results can be summarized as follows:

- For isentropic trajectories very little quasi-horizontal exchange across the vortex edge is observed in August and September 1998 (less than 0.37%/week). The months October and November 1998 give somewhat higher leakage rates with a maximum value in October at 550 K of 1.95%/week. The latter was due to the more frequent occurrence of filamentation in this month. Two isentropic runs for August and October 1998 at 450 K show, respectively, no or only little transport (0.20%/week) of midlatitude air across the vortex edge into the vortex core.

- The 3-D trajectory results for October 1998 show maximum quasi-horizontal mass leakage rates at 550 K of 2.42%/week. Somewhat lower rates are found at 350 K (0.62%/week) and 450 K (1.27%/week).

- Both isentropic and 3-D trajectory results show little quasi-horizontal exchange. However, the 3-D trajectory results show that diabatic effects are already important on a timescale of a couple of days. Therefore, 3-D trajectory calculations will provide more accurate exchange rates than the isentropic ones.

- A comparison has been made between our 3-D trajectory results with the Eulerian model results of W97a. The quasi-horizontal leakage rate of W97a for August-October 1993 (1.19%/week) is significantly larger than our mass leakage rate (0.50%/week) at 450 K for the same period and year. The mean quasi-horizontal leakage rate of W97a (1.68%/week) is only slightly larger than our value at 450 K (1.27%/week) for August-October 1998. We have found average diabatic cooling rates of 0.48 and 0.67 K/day in $\theta$ for, respectively, August-October 1993 and 1998, which are in good agreement with recently reported values in the literature. The average value of 0.3 K/day of W97a for the same months is probably too low as a consequence of their coarse model resolution.
Larger quasi-horizontal exchange rates and diabatic cooling rates are found for the even year 1998 than for the odd year 1993. These results support observations in recent literature concerning the interannual variability of the permeability of the polar vortex.

Overall our results confirm that the Antarctic polar vortex edge is highly impermeable in late austral winter and spring. This is also the season when major depletion of ozone occurs within the polar vortex. The modelled and observed ozone chemical loss rates in the Antarctic ozone hole reported by other authors (Mackenzie et al., 1996; Ricaud et al., 1998; Solomon, 1999) are on average of the order 1.0-2.0 \%/day, which is much higher than the exchange rates in our study. Hence our results support the ‘containment vessel hypothesis’ in contrast to the ‘flowing processor’ hypothesis which assumes that dynamical and chemical time scales are approximately of the same order of magnitude.
Chapter 3

A study of inertia-gravity wave induced exchange across the Antarctic polar vortex edge in late austral winter using 3-D trajectories

The degree of permeability of the Antarctic polar vortex edge to small-scale inertia-gravity waves (IGWs) has not been studied very much. The Royal Netherlands Meteorological Institute (KNMI) trajectory model has been used to investigate the effect of IGWs upon the leakage of the Antarctic polar vortex edge in late austral winter. The trajectory computations are performed using wind data from the European Centre for Medium-range Weather Forecasts (ECMWF) for August 1998 on the 450 K isentropic level. A prescribed inertia-gravity wave wind field is superimposed on the ECMWF wind fields. Thousands of 3-D air parcel trajectories are integrated one month forward in time. In the absence of a prescribed IGW-field no mass exchange is observed as was found from earlier experiments. For a range of realistic values of the wave parameters such as the horizontal and vertical wavelength and wave amplitude, mass exchange rates are computed. The orientation of the horizontal wavevector \( \mathbf{k} \) is found to significantly influence the amount of exchange from the vortex as well as the final horizontal and vertical distributions of air parcels in the polar vortex. Exchange rates of 0.11%/month or less are found. In this case most exchange occurs in the lower part of the polar vortex below 400 K. The presence of a prescribed IGW-field has an effect on the diabatic descent of air parcels in the vortex. Average
IGW induced exchange across the Antarctic polar vortex edge

diabatic cooling rates proportionally increase with decreasing horizontal wavelength from 1.47 K/day to 1.85 K/day. The explanation for this is twofold. First, the results show that in the presence of IGWs air parcels shift slightly poleward where they experience larger subsidence. On the other hand, the polarization relations for IGWs show that the shortest waves have the largest wave frequencies and consequently the largest vertical wave amplitudes. Therefore shorter waves contribute to a larger degree to the downward motion in the vortex than longer waves do. Our results show that small-scale displacements due to IGWs, can give rise to increased leakage of ozone depleted air from the ozone hole into the midlatitude stratosphere.

3.1 Introduction

The degree of permeability of the Antarctic polar vortex has been much debated since the discovery of the ozone hole in the mid 1980s (see Solomon (e.g. 1999) for an overview). Studies by Juckes and McIntyre (1987) and McIntyre (1989) have shown that, based on theoretical arguments and numerical experiments, the polar vortex edge should act as a barrier to cross-edge transport of trace constituents such as ozone. Other studies have confirmed their hypotheses and ideas and have shown that the edge of the Antarctic polar vortex is quite resistant to transports induced by large-scale atmospheric motions (e.g. Bowman, 1993a; Chen, 1994; Chen et al., 1994; Öllers et al., 2002b).

On the other hand, the degree of permeability of the vortex edge or other PV gradients to smaller scale disturbances such as IGWs has been much less studied. Danielsen et al. (1991) showed that low-frequency IGWs can induce cross-jet stratosphere-troposphere exchange of minor trace constituents at midlatitudes. Tuck (1989) and Tuck et al. (1992) observed blobs of vortex air in the lower midlatitude stratosphere and argued that these features might reflect cross-edge exchange. Later, Pierce and Fairlie (1993) argued that the observation of such blobs in the midlatitude stratosphere could be a consequence of cross-edge exchange induced by small-scale motions like IGWs. However, the physical relevance of some of these blobs is subject to uncertainties. Part of them are artefacts due to numerical aliasing and assimilation procedures. Pierce et al. (1994) showed, using Lagrangian material line simulations, that IGWs can indeed have a significant impact on the mixing in the vortex edge region.

Based on data obtained during various measurement campaigns Reid et al. (1993, 1994); Teitelbaum et al. (1994) and Teitelbaum et al. (1996) have shown that the occurrence of part of the ozone laminae close to the polar vortex edge can be related to propagating IGWs.

Recently, Pfenninger et al. (1999) and Yoshiki and Sato (2000) studied charac-
Trajectory model and data characteristics of gravity waves using radiosonde observations above the South Pole. Guest et al. (2000) examined the properties of IGWs in the Southern Hemisphere lower stratosphere. These studies provide values for a wide range of parameters that apply to IGWs and the wintertime polar vortex. Some of these values will be used in this study.

In this study the impact of (idealized) IGWs on the permeability of the Antarctic polar vortex edge will be investigated using high resolution (1° × 1°) 3-D trajectory calculations. Earlier calculations with this trajectory model (Öllers et al., 2002b) have shown negligible leakage through the vortex edge in the absence of IGWs. The paper is organized as follows. In section 2 the trajectory model and data used are briefly described. In section 3 the polarization relations for IGWs are given and used to assign realistic values to the wave parameters. Section 4 describes the numerical experiments that have been performed, and the analysis method used for determining the amount of exchange across the vortex edge. The results are discussed in section 5. The last section summarizes our most important results.

3.2 Trajectory model and data

The Royal Netherlands Meteorological Institute (KNMI) trajectory model (Scheele et al., 1996) has been used to calculate the 3-D trajectories. The trajectory model uses 6 hour forecasts (first-guess) of the three dimensional wind and temperature from the ECMWF to compute the displacement of air parcels over a model time step of \( \Delta t = 10 \) min. The advantage of using forecast instead of analysis data is that the wind and mass density fields are in better physical balance. Analysis data exhibit imbalances due to the inclusion of new observations. In order to obtain the wind and temperature at the trajectory locations, a linear interpolation scheme in longitude and latitude is applied. In the vertical an interpolation with log \( (p) \) is applied to the data. The time interpolation is quadratic. The ECMWF data are available on a horizontal grid of 1° × 1° resolution and 31 hybrid sigma-\( p \) model levels. The Antarctic stratosphere is represented by approximately 10 model levels, with the uppermost level at 10 hPa. The trajectory calculations are performed using data for the period of August 1 1998, 12 GMT to August 31 1998, 12 GMT. The accuracy of the 3-D trajectories is of the order of 2% or less and is mainly determined by inaccuracies in the background wind fields (Bowman, 1993a) and the numerical trajectory integration scheme. For more details concerning the accuracy of the trajectories the reader is referred to Öllers et al. (2002b).
3.3 Imposed inertia-gravity wave field

Although the ECMWF wind and temperature fields are available at relatively high resolution, they do not properly resolve the small-scale disturbances induced by IGWs. In order to account for the effect on the circulation of unresolved small-scale waves like IGWs the ECMWF model and other general circulation models contain subgrid scale gravity wave parameterizations usually formulated in terms of wave drag (e.g. McLandress, 1998; Manzini and McFarlane, 1998). In this way, the effects of (inertia-)gravity waves are to a certain degree represented in a physically realistic manner. For example, the ECMWF model takes into account momentum transport induced by breaking subgrid-scale mountain waves (e.g. Miller et al., 1989; Lott, 1994). These parameterisations do not allow evaluation of the effect of the small-scale motions themselves on the displacement of air parcels. They only evaluate the effect on the background wind. Pierce et al. (1994) noted, based on their Lagrangian material line simulations, that the inclusion of an explicitly prescribed IGW wind field in the trajectory model ensures that the air parcel trajectories can actually feel the impact of subgrid scale disturbances. Here we largely follow their approach, i.e. we do not consider dynamical interactions between IGWs and the background flow or wave-wave interactions. Here we will follow a kinematic approach by simply adding the wave velocity components to the background wind field.

The wind field components of the background flow are defined in spherical \( \lambda \phi z \) coordinates. Because the prescribed IGW wind field will only act in a limited region near the polar vortex edge (\( \Delta \phi \simeq 15^\circ \) latitude, section 4) it is assumed that the IGW velocity components can be defined in a local Cartesian \( xyz \) reference frame. This is only valid if the \( x \)-coordinate is locally perpendicular to meridians and directed eastward. The \( y \)-coordinate is locally perpendicular to parallels and directed northward (towards the equator). The \( z \)-coordinate corresponds to altitude.

The perturbations of the wind components \( u' \), \( v' \) and \( w' \) in the \( x \), \( y \) and \( z \)-direction, respectively, due to IGWs can be obtained from the linearized inviscid Navier-Stokes equations in log-pressure coordinates for small-amplitude hydrostatic perturbations, i.e.

\[
\begin{align*}
    u' &= \frac{1}{(k^2 + l^2) \Lambda^2} (\omega k \cos \phi + lf \sin \phi)RT', \\
    v' &= \frac{1}{(k^2 + l^2) \Lambda^2} (\omega l \cos \phi - kf \sin \phi)RT', \\
    w' &= -\frac{\omega m}{N^2} \cos \phi RT',
\end{align*}
\]

which are the polarization relations for IGWs (see appendix A for details). In (3.1) through (3.3) \( \Lambda^2 = N^2/m^2 \). \( N \) is the Brunt-Väisälä frequency, \( f \) the Coriolis fre-
Imposed inertia-gravity wave field

Figure 3.1: Potential vorticity (0.1 PVU) on the 450 K isentropic level on August 1 1998, 12 GMT derived from ECMWF analysis data.

frequency, and \( m \) the vertical wavenumber in the \( z \)-direction. \( k \) and \( l \) are the horizontal wavenumbers in the \( x \) and \( y \)-directions, respectively. \( T' \) is the amplitude of the temperature perturbation and \( R \) the gas constant for dry air \( (R = 287 \text{ JK}^{-1}\text{kg}^{-1}) \). The phase of the waves is given by \( \varphi = \omega t - kx - ly - mz \).

In every integration step of the trajectory model, the horizontal and vertical wave velocity components are determined from the polarization relations given above and added to the background wind given by the ECMWF model at the air parcel location.

Pfenninger et al. (1999) and Yoshiki and Sato (2000) show that typical values for the Brunt-Väisälä frequency \( N \) in the lower stratosphere (15-20 km) in August above the South Pole are of the order \( 2\cdot10^{-2} \text{ s}^{-1} \). The scale height \( H \) is chosen to be 7000 m. The value of the Coriolis parameter \( f \) is calculated at 65 S yielding \( -1.3 \cdot 10^{-4} \text{ s}^{-1} \). Typical realistic values for the wave parameters are chosen from the relatively sparse studies that report on observations of IGWs in the polar stratosphere and analyse their characteristics. Typical values for the amplitude of \( T' \) are 1.5 K
(Pfenninger et al., 1999). Guest et al. (2000) report on the properties of IGWs in the Southern Hemisphere lower stratosphere and state that typical horizontal wavelengths are between 100 and 1000 km and vertical wavelengths between approximately 1 and 7 km. Wave frequencies are usually around $5/f$ or less. Here our primary interest is in small-scale displacements, so in order to fit a couple of wavelengths within the width of the vortex edge, it is assumed that $\lambda_H \ll L_{edge}$. $\lambda_H$ corresponds to the horizontal wavelength of the IGW and $L_{edge}$ corresponds to the width of the polar vortex edge.

A typical value for $L_{edge}$ for the period under consideration (August 1998) is about 1000 km (Figure 3.1). We choose in this study horizontal wavelengths to be between 150 and 350 km and vertical wavelength between 2.5 and 10 km. The value of the wave frequency is calculated from the dispersion relation (identity (3.11) in appendix A) and ranged from 1.5 to 5.4 times the Coriolis frequency $f$.

### 3.4 Experimental design and exchange diagnostics

In every trajectory calculation, air parcel trajectories are started on a $1^\circ \times 1^\circ$ latitude-longitude parcel grid in a 3-D mode. The 3-D mode takes into account cross-diabatic air motions. As is shown in Öllers et al. (2002b), for integration times longer than a few days, the 3-D mode gives a more realistic representation of the air motion in the polar vortex than trajectory calculations performed on isentropic surfaces. The trajectories are integrated 1 month forward in time starting on the 450 K isentropic level on August 1, 1998, 12 GMT (Figure 3.1). The month August corresponds to the period just after vortex build-up. At that time the polar vortex edge is well-defined and relatively undisturbed by large-scale planetary waves. The 450 K isentropic level corresponds to altitudes where most ozone is destroyed in late winter and spring (approximately 20 km) (e.g. Solomon, 1999). Also, in Öllers et al. (2002b) it is shown that for this period and initial potential temperature, there is negligible mass exchange between the polar vortex and midlatitudes in the absence of a prescribed IGW-field. Therefore, this period represents a good case for studying the effect of small-scale disturbances like IGWs upon the exchange of air across the polar vortex edge.

The method to determine whether air parcels are initially inside or outside the polar vortex is described in Öllers et al. (2002b). In short, it consists of the use of maps of potential vorticity (PV) on isentropic surfaces derived from ECMWF data, to determine typical PV contours for the inner and outer vortex edge. Air parcels that are ‘inside’ the polar vortex are within the area restricted by the inner edge critical PV contour of the outer edge. At regular time intervals (96 hours) mass-weighted fluxes are evaluated by determining which air parcels cross the critical PV contour and do not cross this contour again in the opposite sense the next 120 hours (the threshold for the residence time).
Critical PV contours are determined every 50 K between 350 K and 750 K and additionally at 340 K for August 1 1998, 12 GMT, August 10 1998, 12 GMT, August 20 1998, 12 GMT and August 30 1998, 12 GMT. The critical PV contour at intermediate isentropic levels and times is determined by linear interpolation in $\log(p)$ and time. This is sufficient to describe the mean time evolution of the PV at the vortex edge for our purposes. The 10 hPa level ($\approx 750$ K) is the uppermost level of the ECMWF model and will be taken here as the top of the polar vortex. The vortex bottom is also determined using PV maps. It is found to be at 340 K throughout the studied period. Below the bottom there is no well-defined polar vortex structure anymore.

In every trajectory calculation, initially approximately 9000 air parcels are started inside the polar vortex on the 450 K isentropic level corresponding with a latitude-longitude resolution of $1^\circ \times 1^\circ$.

Since we are primarily interested in mass exchange across the vortex edge, we superimpose the IGW-field (3.1) through (3.3) on the ECMWF background field between 50$^\circ$ and 65$^\circ$ S and between 10 and 100 hPa (approximately between 380 K and 650 K). The region between 50$^\circ$ and 65$^\circ$ S for isentropic levels above 400 K, covers the average position of the vortex edge. Below approximately 400 K the vortex edge is already less well defined.

In each numerical experiment the horizontal angle of incidence of the IGW, which is defined as $\alpha_H = \arctan(l/k)$ has been chosen equal to $\alpha_H = 45^\circ$. This means that if the IGW-field is locally superimposed on a westerly background flow, the horizontal wavevector $\vec{k} = (k,l)$ is oriented north-east (note that ‘north’ is towards the equator here).

The numerical experiments are subdivided into three other experiments, to be referred to as experiment 1, 2, and 3. Experiments 1 and 2 study the effect of a varying horizontal and vertical wavelength, respectively. In experiment 3 the wave amplitude has been varied.

In experiment 1 the effect of a changing horizontal wavelength $\lambda_H$, determined by $(k^2 + l^2)^{1/2}$, is investigated. The vertical wavelength $\lambda_V$ is kept constant. The horizontal wavelength $\lambda_H$ takes the values 150, 200, 250, 300 and 350 km and $\lambda_V = 5$ km. The corresponding wave frequencies expressed as a fraction of $f$ take the values 5.4, 4.1, 3.3, 2.8 and 2.5, respectively. All other parameters are kept constant and their values are already defined in the previous section. Horizontal perturbation wave amplitudes $u'$ and $v'$ are of the order 6-7 ms$^{-1}$. From equation (3.3), it follows directly that vertical perturbation wave amplitudes $u'$ increase with increasing wave frequency. Therefore, the largest vertical velocity amplitudes are found for $\lambda_H = 150$ km and are of the order 0.03 ms$^{-1}$.

In experiment 2 $\lambda_V$ is varied and takes the values 2.5, 5 and 10 km, while $\lambda_H$
Table 3.1: Overview of the varied parameters and its corresponding values in each experiment.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_H$ (km)</td>
<td>variable</td>
<td>150</td>
<td>150</td>
</tr>
<tr>
<td>$\lambda_V$ (km)</td>
<td>5.0</td>
<td>variable</td>
<td>5.0</td>
</tr>
<tr>
<td>$\omega/f$ ($s^{-1}$)</td>
<td>variable</td>
<td>variable</td>
<td>5.4</td>
</tr>
<tr>
<td>$T'$ (K)</td>
<td>0.5</td>
<td>0.5</td>
<td>variable</td>
</tr>
</tbody>
</table>

is kept constant at 150 km. The corresponding wave frequencies as a fraction of $f$ takes the values 1.5, 5.4 and 4.6, respectively. All other parameters are kept constant (see end of section 3.3). In case $\lambda_V = 2.5$ km, $u'$ and $v'$ are of the order 10-12 ms$^{-1}$ and vertical perturbation wave amplitudes $w'$ are of the order 0.05 ms$^{-1}$. In the case where $\lambda_V = 10$ km, $u'$ and $v'$ are of the order 2-3 ms$^{-1}$ and $w'$ is of the order 0.02 ms$^{-1}$.

In experiment 3 we keep all parameters in (3.1) through (3.3) constant with the exception of $T'$. $T'$ is the amplitude of the wave in temperature. Here $\lambda_H$ and $\lambda_V$ are kept constant at 150 km and 5 km, respectively, $\omega/f$ is equal to 5.4 and $T'$ takes the values 0.1, 0.2, 0.3, 0.4, and 0.5 K. In these cases, the horizontal wave amplitudes are of the order 6-7 ms$^{-1}$ or less and vertical perturbation wave amplitudes are of the order 0.03 ms$^{-1}$ or less.

In Table 3.1 the experiments performed and the varied parameters are summarized.

3.5 Trajectory results

3.5.1 Mass exchange rates

Mass exchange rates are determined for experiments 1 through 3. The results of experiment 1 are presented in Table 3.2. Clearly, the contribution of IGWs to the cross-edge transport is small. For large wavelengths the cross-edge transport is numerically insignificant. The exchange rates for air parcels descending below 340 K is only slightly different from zero. The fact that the cross-bottom exchange shows small irregularities is due to the fact that below approximately 390 K the polar vortex edge is generally not well defined. Maximum exchange rates for transport across the vortex edge and the vortex bottom at 340 K are found for $\lambda_H = 150$ km. This is not unreasonable since the smallest wavelengths have also the highest frequencies. This means that air parcels will experience more frequent displacements than they would
Results

experience for longer waves with lower frequencies, which increases the chance for cross-edge or cross-bottom exchange.

In Figure 3.2 and Figure 3.3 the distribution after 576 hours of integration of air parcels that have left the vortex (cross-edge and below 340 K) and those that stay inside the vortex, respectively, are shown for the background flow with a superimposed IGW-field with $\lambda_H = 150$ km. For comparison the distribution of air parcels that stay inside the vortex for the background wind only is shown in Figure 3.4. Note that none of these air parcels have left the vortex after 576 hours. Figure 3.2 indicates that most mass exchange occurs in the lower polar vortex (below 380 K). Comparing Figure 3.3 and Figure 3.4 shows that in the presence of the prescribed IGW-field more air parcels have descended below approximately 400 K after 576 hours. This result may seem rather puzzling at first instance. However, the air parcels are undergoing stronger subsidence due to the wave motions, as we will see below.

Increasing the vertical wavelength $\lambda_V$ decreases cross-edge and cross-bottom exchange which can be seen from Table 3.3. The decreased cross-edge exchange can be explained from Eqs. (3.1) through (3.2), i.e. the horizontal wave amplitude decreases with increasing vertical wavelength. The increasing amount of cross-bottom exchange can be explained in a similar way from Eq. (3.3), which shows that vertical wave amplitudes increase for decreasing vertical wavelength. This effect is also reflected in the average descent rates which are calculated for the three cases. The results are given in Table 3.3. The average descent rate for $\lambda_V = 2.5$ km is somewhat higher (2.21 K/day) than the descent rates for $\lambda_V$ equal to 5 km (1.85 K/day) and 10 km (1.60 K/day). These results clearly show why most cross-bottom exchange is found for $\lambda_V = 2.5$ km. A larger descent of air parcels inside the polar vortex increases the chance that after 1 month more air parcels are found at isentropic levels below 340 K.

Quasi-horizontal and vertical exchange rates decrease with decreasing amplitude of the IGWs ($\sim T'$), which can be observed from Table 3.4. Clearly, an increase in wave amplitude has a significant effect on the amount of exchange from the polar vortex.

The presence of the IGW-field between 50-65° S creates a rather abrupt transition between the region with background wind and the IGW-field superimposed and the region where there is the background wind only. However, the results described above do not change when a more smooth transition between these two regions is chosen. Furthermore, exchange rates are also calculated for $\alpha_H$ equal to 30° and 60° and -30°, -45° and -60°. In case $\alpha_H$ is negative, the horizontal wavevector $\hat{k}$ is oriented north-west with respect to the westerly background flow. It is found that for $\alpha_H$ equal to 30° and 60°, cross-edge exchange rates are 0.02%/month and 0.04%/month,
Figure 3.2: Distribution of the total number of air parcels that have left the vortex either quasi-horizontally or vertically through the vortex bottom at 340 K after 576 hours. The wind field consists of the ECMWF background wind plus a prescribed IGW-field ($\lambda_H = 150$ km, $\lambda_V = 5$ km, $T' = 0.5$ K) added. The trajectory calculation started on August 1 1998, 12 GMT on a regular $1^\circ \times 1^\circ$ grid.

Figure 3.3: Same as in Figure 3.2 but now the distribution of air parcels that stay inside the polar vortex.
Results

Table 3.2: Mass exchange rates in percent per month for the trajectory calculation performed in experiment 1, i.e. various different horizontal wavelengths $\lambda_H$ and $\lambda_V = 5$ km, $T' = 0.5$ K and $\alpha_H = 45^\circ$.

<table>
<thead>
<tr>
<th>$\lambda_H$ (km)</th>
<th>Cross-edge</th>
<th>Cross-bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>200</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td>250</td>
<td>0.05</td>
<td>0.04</td>
</tr>
<tr>
<td>275</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>300</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>350</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3.3: Mass exchange rates in percent per month for the trajectory calculation performed in experiment 2, i.e. for various vertical wavelengths $\lambda_V$ and $\lambda_H = 150$ km, $T' = 0.5$ K and $\alpha_H = 45^\circ$.

<table>
<thead>
<tr>
<th>$\lambda_V$ (km)</th>
<th>Cross-edge</th>
<th>Cross-bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>0.21</td>
<td>1.06</td>
</tr>
<tr>
<td>5.0</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>10</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 3.4: Mass exchange rates in percent per month for the trajectory calculation performed in experiment 3 for various values of the temperature perturbation $T'$. $\lambda_H = 150$ km, $\lambda_V = 5$ km and $\alpha_H = 45^\circ$.

<table>
<thead>
<tr>
<th>$T'$ (K)</th>
<th>Cross-edge</th>
<th>Cross-bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.2</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.3</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.4</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>0.5</td>
<td>0.11</td>
<td>0.02</td>
</tr>
</tbody>
</table>

respectively. The cross-bottom exchange is 0.00%/month and 0.06%/month, respectively. For negative $\alpha_H$ no significant exchange has been found.
3.5.2 Features of air mass distributions in the presence of IGWs

From comparison of Figure 3.3 and Figure 3.4 it can be observed that the distribution of air parcels is horizontally more homogeneous in case they are advected by the background wind only than in case additionally a prescribed IGW-field is present.

The horizontal distribution of air parcels after 1 month for the background wind only and for the background wind with IGWs of different horizontal wavelengths superimposed are shown in Figure 3.5. From this figure it becomes clear that in the presence of an IGW-field, the air parcels are displaced more towards higher latitudes. This is even more obvious from Figure 3.6, where the potential temperature of the air parcels is plotted as a function of latitude, again for the background wind only and for the background wind with IGWs of different horizontal wavelengths superimposed.
Results

Figure 3.5: Horizontal distribution of air parcels after 720 hours, in case they are advected by the background wind only, and in case IGW-fields with different horizontal wavelengths are superimposed on the background wind. In case IGWs are prescribed, $\lambda_V = 5$ km and $T^0 = 0.5$ K. The horizontal angle of incidence $\alpha_H = 45^\circ$.

At first instance this parcel drift may appear a bit peculiar. However, in general a parcel that is moving in a wave of finite amplitude will start to drift viewed in a frame moving with the wave. This drift becomes larger as the ratio between the wave amplitude and the phase velocity increases. We verified this by solving equation of motion for an parcel in a wave of finite amplitude i.e. $d\vec{r}/dt = \vec{u}_w(\vec{r}, t)$, with $\vec{u}_w$ the velocity of the wave as a function of the position $\vec{r}$ and time $t$ (see appendix B for details). Using the wave parameter values from section 3.4, the drift increases, as expected, as we increase the ratio between the wave amplitude and the phase velocity. However,
Figure 3.6: Potential temperature (K) as a function of latitude after 720 hours, for air parcels that are either advected only by the background wind or by the background wind plus an IGW-field. Results for IGWs with different horizontal wavelengths are shown. In case IGWs are prescribed, $\lambda_V = 5$ km and $T' = 0.5$ K. The horizontal angle of incidence $\alpha_H = 45^\circ$.

one cannot state that the poleward shift of air parcels seen in Figure 3.6 is merely due to the drift caused by the wave field. The effect of the background wind should also be taken into account and in that case one should solve the equation of motion $dr^2/dt = \vec{u}_w(r,t) + U_0(r,t)$, $U_0$ representing the background wind. This inevitably
complicates the problem. Note also, that in the type of waves described above (appendix B) the parcel movement is in a straight line and is parallel (longitudinal) to the direction of phase propagation. This is not exactly the case for IGWs, whose parcel orbits are elliptic. It is therefore difficult to explain in detail why the air parcels in Figure 3.6 move poleward and not in some other direction. Nevertheless, it is clear that the wave field does play an important role in the poleward shift of the air parcels.

To illustrate this in another way, additional trajectory calculations are performed using the ECMWF monthly averaged background wind field for the month August 1998. The advantage of using a monthly averaged wind field is that it is very smooth and regular and it is therefore very well suited for investigating the role of IGWs in the poleward shift of air parcels (Figure 3.7). Figure 3.7 shows that in the monthly averaged case, PV contours at 450 K are nearly coincident with latitude circles. The corresponding wind field is nearly zonal. A total of 11 trajectories is initially located at 0°E and 70°S, 66°S, 62°S, 58°S, 54°S and 50°S at 450 K and are integrated one month forward in time. The IGW-field, with $\lambda_H = 150$ km, $\lambda_V = 5$ km and $T' = 0.5$ K, is located between 50° and 65°S. The trajectories are calculated for the
IGW induced exchange across the Antarctic polar vortex edge

background flow only, and for the background flow with the IGW-field superimposed for $\alpha_H$ equal to $30^\circ$, $45^\circ$ and $60^\circ$. Note that the background wind field, in this simplified case, is ‘frozen’. At every integration step in time the air parcels ‘experience’ the same large-scale wind field but at every spatial integration step the air parcels move to a different location where the wind field is (slightly) different. Overall, the air parcels will move mainly in the zonal direction.

In Figure 3.8 the latitude is plotted as a function of time for the case the 11 air parcels are advected by the monthly mean wind field only (a) and, in case the IGW-field is added to the background wind for different values of $\alpha_H$ (b)-(d). Most air parcels move mainly in the zonal direction in the case where they are advected by the background wind field only. However, when the IGW wind field is added, most parcels drift from their initial latitudes and move towards higher latitudes. Eventually after 1 month (720 hours), the air parcels have drifted approximately $10-15^\circ$ in latitude poleward. Also, note that the air parcels that start initially at $70^\circ$S and $66^\circ$S are unaffected by the IGW-field (Figure 3.8(b)-(d)). Note that the drift towards the pole slightly increases as $\alpha_H$ increases. This is in agreement with the results found in section 3.5.1 for the cross-bottom exchange. At the end of section 3.5.1. and in Table 3.2, cross-bottom exchange rates are found for IGWs with $\lambda_H = 150$ km, $\lambda_V = 5$ km and $T' = 0.5$ K of $0.00 (\alpha_H = 30^\circ)$, $0.02 (\alpha_H = 45^\circ)$, and $0.06/\text{month} (\alpha_H = 60^\circ)$. As air parcels move towards higher latitudes for larger $\alpha_H$, they enter a region where diabatic cooling is more intense in late winter than near the vortex edge (Manney et al., 1994). Therefore, in the presence of IGWs and for increasing $\alpha_H$ more air parcels will experience a larger vertical descent than in the absence of IGWs. In the latter case more air parcels remain near the edge of the polar vortex where descent rates are smaller (Figure 3.6). Eventually, in the presence of IGWs, more air parcels are found at low potential temperatures. At low potential temperatures the polar vortex edge is less well-defined and mass exchange can more easily occur.

3.5.3 Effect of IGWs on diabatic descent

In the previous section it is shown that the addition of IGWs onto the background wind leads to a small shift of air parcel trajectories towards higher (poleward) latitudes. Air parcels that move towards higher latitudes inside the polar vortex, enter a region where they experience a larger subsidence than near the vortex edge (Manney et al., 1994). Therefore, in the presence of IGWs more air parcels will experience a significant vertical descent than in the absence of IGWs. As a consequence, the average descent rate may differ depending on whether a prescribed IGW-field is present or not. Average descent rates are calculated for the trajectories that start at $450$ K on August 1 1998, 12 GMT. The results are summarized in Table 3.5. Clearly, the
**Figure 3.8:** Latitude (degrees S) as a function of time (hours) in case the air parcels are advected by the monthly averaged background wind (a) and for the case the air parcels are advected by the monthly averaged background wind plus an IGW-perturbation ($\lambda_H = 150$ km, $\lambda_V = 5$ km and $T' = 0.5$ K) for $\alpha_H$ equal to 30°, 45° and 60° (b)-(d).

Calculated average descent rates are all larger in case IGWs are present. In the absence of a prescribed IGW-field, the calculated average descent rate is 1.20 K/day.
Table 3.5: Average descent rates in K/day for the trajectory calculations in which a prescribed IGW-field is superimposed upon the ECMWF background wind for $\alpha_H = 45^\circ$ and for different horizontal wavelengths. $\lambda_V = 5$ km and $T' = 0.5$ K. The average descent rate for the case that no prescribed IGW-field is present is also given. The trajectory calculations are started on August 1 1998, 12 GMT at 450 K.

<table>
<thead>
<tr>
<th>Average descent rate (K/day)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_H = 150$ km</td>
<td>1.85</td>
</tr>
<tr>
<td>$\lambda_H = 250$ km</td>
<td>1.73</td>
</tr>
<tr>
<td>$\lambda_H = 350$ km</td>
<td>1.47</td>
</tr>
<tr>
<td>Background wind only</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Also, average descent rates increase for decreasing horizontal wavelength. This can be explained from the polarization relations and the dispersion relation for IGWs (see appendix A) which show that for decreasing horizontal wavelength, frequencies increase. This results, for constant vertical wavelength, in an increase of the vertical wave perturbation amplitude $w'$ (see end of section 4). Therefore shorter waves will contribute to a relatively larger degree to the downward motion inside the polar vortex than longer waves do. The larger average descent rates in the case where IGWs are present also explains why more air parcels are found at low potential temperatures in Figure 3.6. At low potential temperatures the air parcels are no longer confined by a well-defined vortex edge (Öllers et al., 2002b). Since the background wind field is generally small in the lower polar vortex, the IGW field induces more easily exchange of air parcels from the polar vortex to midlatitudes.

3.6 Conclusions

In the context of the permeability of the Antarctic polar vortex edge, the effect of a prescribed IGW-field that is superimposed upon the ECMWF background wind field has been investigated in late austral winter using high resolution 3-D trajectories. For the first time, quantitative exchange rates across the vortex edge and through the bottom at 340 K have been determined for a case with IGWs superimposed upon the background wind. Three different numerical experiments have been performed in which the effect of horizontal and vertical wavelength and the wave amplitude upon the leakage of air parcels from the polar vortex into midlatitudes has been investigated. The angle of incidence of the wave on the background wind is equal to $45^\circ$.

The main results are:
In general it can be concluded that IGWs induce a small but significant increase in the mass leakage from the polar vortex into midlatitudes. Typical cross-edge and cross-bottom exchange rates are still of the order $0.11\%$/month and $0.04\%$/month of the air in the vortex, respectively. The amount of cross-edge exchange is largest for shortest horizontal wavelengths. This can be explained from the fact that shorter waves have higher frequencies. As a consequence, air parcels will experience more frequent displacements, which increases the chance for cross-edge exchange. The amount of cross-bottom exchange is generally small and occurs for horizontal wavelengths shorter than 300 km. An increase in vertical wavelength shows a decrease in cross-edge and cross-bottom exchange. This result can be explained from the fact that the horizontal and vertical wave amplitudes increase for decreasing vertical wavelength. Larger vertical wave amplitudes increase the amount of cross-bottom exchange. This is confirmed by average descent rates which have been calculated. The average descent rates increase for decreasing vertical wavelength. From the results above it can therefore be concluded that both wave frequency and wave amplitude play an important role in the amount of exchange induced by IGWs from the polar vortex.

The distribution of air parcels in the presence of a prescribed IGW-field is less homogeneous than in the absence of IGWs. In general, it has been shown that parcels moving in a finite amplitude wave will start to drift depending on the ratio between the wave amplitude and the phase velocity. Although the parcel orbits in IGWs are generally elliptic, the simple wave model may nevertheless explain to some degree the observed drift in our numerical experiments. Furthermore, the presence of the background wind complicates the problem a lot and does no longer allow to investigate in detail why the air parcels move poleward in our trajectory calculations. Also, the poleward drift of the air parcels increases as the horizontal angle of incidence $\alpha_H$ is increased. At higher latitudes air parcels experience a larger vertical descent than near the edge of the vortex. Eventually, more air parcels subside to low potential temperatures as $\alpha_H$ increases. There, the polar vortex is less well defined and mass exchange will take place more easily. This is confirmed by the trajectory calculations; more cross-bottom exchange occurred for increasing values of $\alpha_H$.

In the presence of a prescribed IGW-field, more air parcels are moved to high latitudes. Diabatic cooling is larger in the vortex core than near the edge in winter. Therefore more air parcels will experience a larger diabatic cooling in
the vortex core than in the absence of IGWs. The number of air parcels that arrive below 340 K increases with decreasing horizontal wavelength. Decreasing the horizontal wavelength for constant vertical wavelength, increases the wave frequency and thus the vertical wave amplitude. Waves of short wavelength will therefore contribute to a larger degree to the downward motion inside the polar vortex. This is reflected in the average diabatic cooling rates that have been calculated. In the presence of IGWs the values for the average descent range from 1.47 K/day ($\lambda_H = 350$ km) to 1.85 K/day ($\lambda_H = 150$ km), whereas in the absence of IGWs a value of 1.20 K/day has been found.

Summarizing, four major effects can be distinguished that determine the amount of exchange from the polar vortex and the distribution of air parcels in the polar vortex due to IGWs. First, IGWs will induce a poleward shift of air parcels inside the polar vortex. Also, the shift can increase or decrease depending on the orientation of the IGW-field with respect to the background flow. A larger poleward shift, eventually leads to more cross-bottom exchange. Furthermore, waves of short wavelength contribute to a larger degree to the amount of cross-bottom exchange, due to their larger vertical wave amplitude. Finally, increasing the horizontal and vertical wave amplitudes increases the amount of cross-edge and cross-bottom exchange.

Although our results have been obtained using a systematic analysis method, there are a couple of drawbacks that should be mentioned. The IGW-field used is highly idealized. It is not likely that in the real atmosphere monochromatic IGWs are present for over a month in a broad latitude band between 50° and 65° S and between 380 and 650 K, as assumed in our study. In most cases internal gravity waves propagating into the middle atmosphere originate from localized sources, such as tropospheric frontal and jet systems, surface orography and convection (e.g. Guest et al., 2000), and are thus clearly distinct from plane waves. We have performed many trajectory calculations in which the prescribed IGW-field is imposed over a much smaller region, for example between 50° and 65° S, between 270° and 300° E and between 380 and 500 K. However, in none of these cases any significant mass exchange between the polar vortex and midlatitudes is found. Therefore, our results suggest that IGWs cannot be considered as an important contribution to the leakage of air from the Antarctic polar vortex.

Appendix A

The IGW-induced velocity perturbations are derived from the linearized dynamical equations for small hydrostatic perturbations defined in a local Cartesian $xyz$ refer-
Appendix

ence frame (e.g. Holton, 1992). The $x$ and $y$-direction are locally parallel to longitudes and latitudes, respectively. The equations take the form

\[
\frac{\partial u}{\partial t} - fu = -\frac{\partial \phi}{\partial x}, \tag{3.4}
\]
\[
\frac{\partial \phi}{\partial t} + fu = -\frac{\partial \phi}{\partial y}, \tag{3.5}
\]
\[
\frac{\partial^2 \phi}{\partial t \partial z} + N^2 w = 0, \tag{3.6}
\]
\[
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{3.7}
\]

where $u$, $v$ and $w$ are horizontal and vertical velocity perturbations in $x$, $y$ and $z$ directions, respectively, $\phi$ corresponds to the geopotential height, $f$ is the Coriolis frequency, $p$ is the pressure and $g$ is the gravitational acceleration. The Brunt-Väisälä frequency $N$ in Eq. (3.6) is defined by $N = [(g/\theta_0)\partial \theta_0/\partial z]^{1/2}$ with $\theta_0$ the background potential temperature. It is assumed that $N$ is constant (Andrews et al., 1987; Pfenninger et al., 1999). For low frequency IGWs the assumption of hydrostatic equilibrium is justified. We assume wavelike disturbances $\zeta(x, y, z, t) = \zeta' e^{i(\omega t - kx - ly - mz)}$, where $\zeta'$ is a general notation for $u$, $v$, $w$ and $\phi$. Substitution in Eqs. (3.4) through (3.7) yields eventually the following expressions for $u'$, $v'$ and $w'$ in terms of $\psi'$

\[
u' = \frac{\omega k - il f}{\omega^2 - l^2} \psi', \tag{3.8}
\]
\[
u' = \frac{\omega l + ik f}{\omega^2 - l^2} \psi', \tag{3.9}
\]
\[
u' = -\frac{\omega m N^2}{\omega^2} \psi', \tag{3.10}
\]

$k$, $l$ and $m$ being the horizontal and vertical wavenumbers, respectively, and $\omega$ is the wave frequency. From the equations above the following dispersion relation can be derived

\[
\omega^2 - f^2 = (k^2 + l^2)\Lambda^2, \tag{3.11}
\]

with $\Lambda^2 = N^2/m^2$.

Finally, from (3.8) through (3.10), the polarization relations can be derived

\[
u' = \frac{1}{(k^2 + l^2)\Lambda^2} (\omega k \cos \varphi + l f \sin \varphi)RT', \tag{3.12}
\]
\[
u' = \frac{1}{(k^2 + l^2)\Lambda^2} (\omega l \cos \varphi - k f \sin \varphi)RT', \tag{3.13}
\]
\[
u' = -\frac{\omega m}{N^2} \cos \varphi RT', \tag{3.14}
\]
Consider the following general equation describing the motion of a parcel in a wave of finite amplitude,

$$\frac{d\vec{r}}{dt} = \vec{u}_w(\vec{r}, t),$$

(3.15)

with $\vec{u}_w = (u, v, w)$ the velocity of the wave as a function of the position $\vec{r} = (x, y, z)$ and time $t$.

Consider now the equation of motion for a parcel in a one-dimensional wave of finite amplitude in the $y$-direction i.e.

$$v(y, t) \equiv \frac{dy}{dt} = -v_0 \sin(\omega t - ly),$$

(3.16)
with \( v \) the velocity of the wave, \( u_0 \) the amplitude of the wave (for example 6 m\( \text{s}^{-1} \)), \( \omega \) the wave frequency and \( l \) the wavenumber in the \( y \)-direction. Note that from Eq. (3.16) it follows that the parcel movement is in the same direction as the propagation of the phase lines (longitudinal). If we introduce, \( \xi = \omega t \) and \( y^* = ly \) and substitute this in Eq. (3.16) we obtain

\[
\frac{dy^*}{dt^*} = \frac{u_0}{c} \sin(y^* - t^*),
\]

(3.17)

with \( c = \omega/l \) the phase velocity of the wave in the \( y \)-direction. For a wave with a wavelength of 150 km and \( \omega = 5.4f \) (section 3.4), the phase velocity is approximately 16 m\( \text{s}^{-1} \), leading to \( u_0/c = 0.375 \). Consider the frame of reference moving with the wave, i.e. \( \xi = y^* - t^* \) and rewrite Eq. (3.17) as

\[
\frac{d\xi}{dt^*} = \frac{u_0}{c} \sin(\xi) - 1.
\]

(3.18)

Eq. (3.18) can be solved for \( \xi \) using standard mathematical software packages. In Figure 3.9 \( \xi \) is shown as a function of time \( t^* \) for \( u_0/c = 0 \) and \( u_0/c = 0.375 \). The case \( u_0/c = 0 \) corresponds to waves of infinitesimally small amplitude. In this case the parcel does not drift. For finite amplitude waves with \( u_0/c = 0.375 \) a clear drift is visible.
Chapter 4

Simple analytical models of propagating inertia-gravity waves in barotropic shear layers

In this chapter some simple analytical models are presented, that describe the propagation properties of inertia-gravity waves (IGWs) in barotropic shear flows. These studies are meant as a first approach towards the more advanced model in chapter 5. The imposed simplification is that the direction of wave incidence is taken to be perpendicular to the flow direction of the background flow. In this configuration no critical layers occur in the flow. In each case, the profile of the background flow is chosen such that the model can be solved analytically. Analytical expressions for the reflection and transmission coefficients are derived. Although the models are highly idealized, they give some basic insight into the propagation of inertia-gravity waves in simple background flows. Moreover, they can be used as benchmarks to test the performance of numerical codes with which more complicated models, as presented in chapter 5 can be solved.

4.1 Inertia-gravity waves incident upon a piecewise continuous wind profile

Here analytical expressions for the reflection and transmission coefficients will be derived for a piece-wise continuous background wind profile $U(y)$. Here the special case of perpendicular wave incidence with $k = 0$ is considered. For this case the IGW propagates in the $yz$-plane only.

The linearized equations in log-pressure coordinates for hydrostatic perturbations
in a horizontally-sheared background flow $U(y)$ with $k = 0$ are,

$$\frac{\partial u}{\partial t} + v \frac{dU}{dy} - fv = 0,$$  \hspace{1cm} (4.1)

$$\frac{\partial v}{\partial t} + fu = -\frac{\partial \phi}{\partial y},$$  \hspace{1cm} (4.2)

$$\frac{\partial^2 \phi}{\partial t^2} + N^2 w = 0,$$  \hspace{1cm} (4.3)

$$\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} - \frac{w}{H} = 0,$$  \hspace{1cm} (4.4)

where $u, v$ and $w$ are horizontal and vertical velocity perturbations, respectively, $\phi$ corresponds to the geopotential, $f$ is the Coriolis parameter and $H$ is the vertical scale height. In the following $H$ is constant. In Eq. (4.3) $N^2 = (g/\theta_0) d\theta_0/dz$ is the Brunt-Väisälä frequency, where $g$ is the gravitational acceleration and $\theta_0$ is the background potential temperature. Hereafter it is assumed that $N^2$ is constant. The assumption of hydrostatic equilibrium is justified since we are primarily interested in low-frequency IGWs with horizontal scales much larger than the vertical scales.

Assume now wave-like disturbances $\zeta(x,y,z,t) = \zeta(y)e^{i(k-yx+\gamma t)}$, where $\zeta$ is a general notation for $u, v, w$ and $\phi$. After substitution in Eqs. (4.1) through (4.4), the following wave equation for $\zeta$ can be derived

$$\frac{d^2 \zeta}{dy^2} - \frac{1}{A} \frac{dA}{dy} \frac{d\zeta}{dy} + A \frac{\zeta}{A^2} = 0,$$  \hspace{1cm} (4.5)

with $A$ equal to

$$A = \omega^2 + f(dU/dy - f).$$  \hspace{1cm} (4.6)

Divide Eq. (4.5) by $A$ and rewrite it as

$$\frac{d}{dy} \left( \frac{1}{A} \frac{d\zeta}{dy} \right) + \frac{A}{A^2} \frac{\zeta}{A} = 0.$$  \hspace{1cm} (4.7)

Setting $\hat{\phi} = dh/dy$ and integrating this equation with respect to $y$ and multiplying by $A$ yields the differential equation

$$\frac{d^2 h}{dy^2} + \frac{A}{A^2} h = CA.$$  \hspace{1cm} (4.8)

Eq. (4.8) has a particular solution $h_p(y) = CA^2$. The homogeneous equation the reads,

$$\frac{d^2 h}{dy^2} + \frac{A}{A^2} h = 0.$$  \hspace{1cm} (4.9)
Inertia-gravity waves incident upon a tophat jet wind profile

with $A = \omega^2 + f (dU/dy - f)$ and $\Lambda^2 = N^2 / (1/4H^2 + m^2)$.

If one introduces the following scales:

\[
(x, y) = L(\tilde{x}, \tilde{y}), \quad k = \tilde{k}/L, \quad m = (N/fL)\tilde{m},
\]

\[
\omega = f\tilde{\omega}, \quad \Omega = f\tilde{\Omega}, \quad U = U_0\tilde{U}.
\] (4.10)

Substitution of (4.10) into (4.9) and omitting tildes yields

\[
\frac{d^2h}{dy^2} + A(m^2 + \chi^2)h = 0,
\] (4.11)

with

\[
A = \omega^2 + (Ro dU/dy - 1), \quad \chi = Lf/2HN,
\] (4.12)

and the Rossby number

\[
Ro = \frac{U_0}{fL}.
\] (4.13)

For the background flow we choose the following piece-wise continuous velocity profile (the ‘tophat’ jet) (Fig. 4.1)

\[
U(y) = \begin{cases} 
0 & |y| < 1/2, \\
1 & |y| > 1/2.
\end{cases}
\]

The shear $dU/dy$ is represented by (Fig. 4.1):

\[
\frac{dU}{dy} = \delta(y + 1/2) - \delta(y - 1/2).
\] (4.14)

An IGW incident upon the ‘tophat’-jet in region I (Fig. 4.1) gives rise to a reflected wave in region I and a transmitted wave in region III. In region I the solution takes the form:

\[
h(y) = \exp(-i\chi y) + R \exp(i\chi y).
\] (4.15)

The amplitude of the incident wave is equal to unity, $R$ is the amplitude of the reflected wave and

\[
l = ((\omega^2 - 1)(m^2 + \chi^2))^{1/2}.
\] (4.16)

In region III the solution takes the form

\[
h(y) = T \exp(-i\chi y),
\] (4.17)
Propagating inertia-gravity waves in barotropic shear layers

Figure 4.1: The piece-wise continuous velocity profile $U(y)$ (the ‘tophat’ jet) for the background flow (a) and its derivative $dU/dy$ consisting of two $\delta$-functions at $y = -1/2$ and $y = 1/2$ (b). Three regions can be distinguished; regions I and III being outside the ‘jet’ and region II being the ‘jet’.

with $T$ the amplitude of the transmitted wave. In region II we have the solution

$$h(y) = A \exp(-iy) + B \exp(iy),$$

(4.18)

Corresponding to waves with amplitude $A$ and $B$ travelling to the right and left, respectively. The first boundary condition requires that $h(y)$ is continuous in $y = -1/2$ and $y = 1/2$. Thus

$$\lim_{\epsilon \to 0} [h(-1/2 - \epsilon) - h(-1/2 + \epsilon)] = 0,$$

(4.19)

and

$$\lim_{\epsilon \to 0} [h(1/2 - \epsilon) - h(1/2 + \epsilon)] = 0.$$

(4.20)

The second boundary condition in $y = -1/2$ and $y = 1/2$ can be found by integration of Eq. (4.11) from $-1/2 - \epsilon$ to $-1/2 + \epsilon$ and from $1/2 - \epsilon$ to $1/2 + \epsilon$, respectively, and consider the limit $\epsilon \to 0$, i.e.

$$\lim_{\epsilon \to 0} \int_{-1/2-\epsilon}^{-1/2+\epsilon} \left[ \frac{d^2 h}{dy^2} + (m^2 + \chi^2) \left( (\omega^2 - 1) + \text{Ro}\delta(y + 1/2) \right) \right] dy = 0.$$
Integration by parts yields

$$
\lim_{\varepsilon \to 0} \int_{-1/2 - \varepsilon}^{-1/2 + \varepsilon} \frac{d^2 h}{dy^2} dy + \lim_{\varepsilon \to 0} \int_{-1/2 - \varepsilon}^{-1/2 + \varepsilon} (m^2 + \chi^2) (\omega^2 - 1) h dy +
$$

$$= 0 \text{ from first condition}
$$

$$\lim_{\varepsilon \to 0} \int_{-1/2 - \varepsilon}^{-1/2 + \varepsilon} (m^2 + \chi^2) \text{Ro} \delta(y + 1/2) h dy = 0,
$$

so that at $y = -1/2$

$$\lim_{\varepsilon \to 0} \left[ \frac{dh}{dy} \bigg|_{-1/2 + \varepsilon} - \frac{dh}{dy} \bigg|_{-1/2 - \varepsilon} \right] + (m^2 + \chi^2) \text{Ro} h(-1/2) = 0,
$$

(4.21)

holds and at $y = 1/2$ it is similarly found that

$$\lim_{\varepsilon \to 0} \left[ \frac{dh}{dy} \bigg|_{1/2 + \varepsilon} - \frac{dh}{dy} \bigg|_{1/2 - \varepsilon} \right] - (m^2 + \chi^2) \text{Ro} h(1/2) = 0,
$$

(4.22)

where the following general property of delta functions was used

$$\int_{-\infty}^{\infty} \delta(y - y') h(y) dy = h(y').
$$

(4.23)

Application of the boundary conditions (4.19) and (4.20) to the general solutions (4.15) and (4.18) and (4.17) and (4.18), respectively, yields

$$\exp(il/2) + R \exp(-il/2) = A \exp(il/2) + B \exp(-il/2),
$$

(4.24)

$$A \exp(-il/2) + B \exp(il/2) = T \exp(-il/2),
$$

(4.25)

Application of the boundary conditions (4.21) and (4.22) to identities (4.15) and (4.18) and (4.17) and (4.18), respectively, leads to

$$il(\exp(il/2) - R \exp(-il/2) - A \exp(il/2) + B \exp(-il/2)) =
$$

$$= (m^2 + \chi^2) \text{Ro} (\exp(il/2) + R \exp(-il/2)),
$$

(4.26)

$$-il A \exp(-il/2) + il B \exp(il/2) + il T \exp(-il/2) =
$$

$$= (m^2 + \chi^2) \text{Ro} T \exp(-il/2).
$$

(4.27)

By combining Eqs. (4.24)-(4.25) and (4.26)-(4.27), expressions for the amplitudes $R$ and $T$ can be derived. The reflection and transmission coefficients are defined by
Figure 4.2: $|R|^2$ and $|T|^2$ as a function of the wavenumber $l$ for typical atmospheric values of $m, \chi$ and the Rossby number (Ro).

After some algebra one finds

$$|R|^2 = RR^*$$ and $$|T|^2 = TT^*$$, respectively, where the superscript, $'*'$, indicates complex conjugation.

After some algebra one finds

$$|R|^2 = \frac{\eta^2}{l^2} \left( 1 + \frac{\eta^2}{4l^2} \right) \sin^2(l) \left[ 1 + \frac{\eta^2}{l^2} \left( 1 + \frac{\eta^2}{4l^2} \right) \sin^2(l) \right]^{-1},$$ \hspace{1cm} (4.28)

$$|T|^2 = \left[ 1 + \frac{\eta^2}{l^2} \left( 1 + \frac{\eta^2}{4l^2} \right) \sin^2(l) \right]^{-1},$$ \hspace{1cm} (4.29)

with $\eta = (m^2 + \chi^2)Ro$. Note from (4.28) and (4.29) that

$$|R|^2 + |T|^2 = 1,$$ \hspace{1cm} (4.30)

which expresses conservation of wave energy.

In figure 4.2 an example of $|R|^2$ and $|T|^2$ as a function of the wavenumber $l$ is plotted for values of $m, \chi$ and the Rossby number that are typical for the earth’s atmosphere.

### 4.2 Inertia-gravity waves incident upon a hyperbolic tangent wind profile

Here analytical expressions for the reflection and transmission coefficients will be derived for a hyperbolic tangent wind profile in case of perpendicular wave incidence.
4.2.1 Reduction to the hypergeometric equation

Now, Eq. (4.9) in section 4.1 will be studied in case the background flow is represented by a hyperbolic tangent velocity profile i.e.

\[ U(y) = \frac{1}{2} \left(1 + \tanh \left(\frac{y}{2}\right)\right). \]  

(4.31)

The shear \(dU/dy\) is then given by

\[ \frac{dU}{dy} = \frac{1}{4} \text{sech}^2 \left(\frac{y}{2}\right). \]  

(4.32)

Now introduce a transformation of the dependent variable

\[ \eta = -\exp(-y), \]  

(4.33)

and rewrite Eq. (4.9) in terms of the new variable \(\eta\) i.e.

\[ \eta^2 \frac{d^2 h}{d\eta^2} + \eta \frac{dh}{d\eta} + \left(A^2 - \frac{B\eta}{(\eta - 1)^2}\right) h = 0, \]  

(4.34)

where

\[ A = \left((m^2 + \chi^2)(\omega^2 - 1)\right)^{1/2} \]  

(4.35)

\[ B = (m^2 + \chi^2)\text{Ro}. \]  

(4.36)

Eq. (4.34) is a special form of the hypergeometric equation, which has three regular singular points \(\eta = 0, \eta = 1\) and \(\eta = \infty\). The Frobenius exponents of Eq. (4.34), corresponding to the singularities at \(\eta = 0, \eta = 1\) and \(\eta = \infty\), are

\[ \lambda_0 = iA, \quad \lambda_1 = \frac{1}{2} + \frac{1}{2}(1 + 4B)^{1/2}, \quad \lambda_\infty = iA. \]  

(4.37)

If we introduce a power transformation of the dependent variable (Sluijter, 1966; Sluijter and van Duin, 1980; Ronveaux, 1995)

\[ h(\eta) = \eta^{\lambda_0}(\eta - 1)^{\lambda_1} \tilde{h}(\eta), \]  

(4.38)

substitution of (4.38) in Eq. (4.34) and using that \(\lambda_0^2 + A^2 = 0\) and \((\lambda_1^2 - \lambda_1) - B = 0\) yields after some algebra

\[ \frac{d^2 \tilde{h}}{d\eta^2} + \left(\frac{2\lambda_0 + 1}{\eta} + \frac{2\lambda_1 U}{\eta - 1}\right) \frac{d\tilde{h}}{d\eta} + \left(\frac{(2\lambda_0 + 1)\lambda_1 + B}{\eta(\eta - 1)}\right) \tilde{h} = 0. \]  

(4.39)

Eq. (4.39) is the hypergeometric equation in its canonical form. For details on the properties of the hypergeometric equation the reader is referred to Snow (1952), Erdélyi (1953) and Sluijter (1966); Sluijter and van Duin (1980); van Duin and Sluijter (1983).
4.2.2 Reflection and transmission coefficients

From the asymptotic behaviour of the solutions \( h \) in Eq. (4.9), the asymptotic behaviour of the solutions \( \tilde{h} \) in Eq. (4.39) can be derived. If \( |y| \to \infty \), Eq. (4.9) reduces to

\[
\frac{d^2h}{dy^2} + l^2h = 0, \quad (4.40)
\]

with

\[
l^2 = (\omega^2 - 1)(m^2 + \chi^2). \quad (4.41)
\]

An IGW incident upon the background flow from \( y \to -\infty \) gives rise to a reflected wave there and a transmitted wave for \( y \to \infty \). As \( y \to -\infty \) the solution \( h \) takes the form

\[
h(y) = \exp(-\imath y) + R \exp(\imath y), \quad (4.42)
\]

\[
= (-\eta)^{\lambda_0} + R(-\eta)^{-\lambda_0}, \quad (4.43)
\]

where the identities in (4.33) and (4.37) were used. The incident wave has unit amplitude and the reflected wave has amplitude \( R \). For \( y \to \infty \) the solution takes the form

\[
h(y) = T \exp(-\imath y), \quad (4.44)
\]

\[
= (-\eta)^{\lambda_0}, \quad (4.45)
\]

where \( T \) is the amplitude of the transmitted wave.

The asymptotic behaviour of the solutions \( \tilde{h} \) now follows from (4.38). Further note that from (4.33) it follows that \( y \to -\infty \) corresponds with \( \eta = 0 \) and \( y \to \infty \) corresponds with \( \eta \to -\infty \). Thus for \( \eta = 0 \) we have

\[
\tilde{h}(\eta) = (1 - \eta)^{-\lambda_1} \exp(i\pi(\lambda_0 - \lambda_1)) + R\eta^{-2\lambda_0}(1 - \eta)^{-\lambda_1} \exp(i\pi(-\lambda_0 - \lambda_1)), \quad (4.46)
\]

and similarly for \( \eta \to -\infty \)

\[
\tilde{h}(\eta) = T(1 - \eta)^{-\lambda_1} \exp(i\pi(\lambda_0 - \lambda_1)). \quad (4.47)
\]

From the analysis above it is clear that we have to search for solutions of Eq. (4.39) relative to \( \eta = 0 \) and \( \eta \to -\infty \). Solutions of the hypergeometric equation and relations between them are listed in Erdélyi (1953). Here we will closely follow the analysis performed in Sluijter (1966), Chapter 3. The solutions of Eq. (4.39) relative
Inertia-gravity waves incident upon a hyperbolic tangent wind profile

Figure 4.3: $|R|^2$ and $|T|^2$ plotted as a function of the Rossby number for typical values of $m$, $\chi$ and the wave frequency $\omega$.

to the singular points $\eta = 0$ and $\eta \to -\infty$ follow directly from relations (3.2.2) through (3.2.5) in Sluijter (1966) i.e.

$$\tilde{h}_1^0(\eta) = (1-\eta)^{-\lambda_1} F \left( \lambda_1, \lambda_1 - 1, 2\lambda_0 - 1; \frac{\eta}{\eta-1} \right),$$  \hspace{1cm} (4.48)

$$\tilde{h}_2^0(\eta) = \eta^{-2\lambda_0} (1-\eta)^{-\lambda_1} F \left( \lambda_1 - 2\lambda_0 + 2, 1 - 2\lambda_0 - \lambda_1, 3 - 2\lambda_0; \frac{\eta}{\eta-1} \right),$$  \hspace{1cm} (4.49)

$$\tilde{h}_1^\infty(\eta) = (1-\eta)^{-2\lambda_0 - \lambda_1} F \left( \lambda_1, \lambda_1 - 1, 1 - 2\lambda_0; \frac{1}{\eta-1} \right),$$  \hspace{1cm} (4.50)

$$\tilde{h}_2^\infty(\eta) = (1-\eta)^{-\lambda_1} F \left( 2\lambda_0 + \lambda_1, 2\lambda_0 - \lambda_1 - 1, 2\lambda_0 + 1; \frac{1}{\eta-1} \right).$$  \hspace{1cm} (4.51)

The superscripts refer to the specific singular point and $F$ represents the hypergeometric function (see e.g. Erdélyi (1953) for details). The solutions (4.48) through (4.51) are analytic for Re$(\eta) < 0$ (Sluijter, 1966). Also, note that the asymptotical behaviour of $\tilde{h}_1^0$, $\tilde{h}_2^0$ and $\tilde{h}_2^\infty$ corresponds with that in (4.46) and (4.47), respectively. Thus we may also write

$$T\tilde{h}_2^\infty = \tilde{h}_1^0 + R \exp(-2i\pi\lambda_0) \tilde{h}_2^0$$  \hspace{1cm} (4.52)

Analytical continuation of the solutions $\tilde{h}_1^0$, $\tilde{h}_2^0$ and $\tilde{h}_2^\infty$ yields a linear relation with constant coefficients (Erdélyi, 1953; Sluijter, 1966) from which expressions for
Propagating inertia-gravity waves in barotropic shear layers

the reflection coefficient $R$ and transmission coefficient $T$ in terms of Gamma functions can be derived i.e.

$$R = -\frac{\Gamma(1 - 2\lambda_0 - \lambda_1)\Gamma(\lambda_1 - 2\lambda_0)}{\Gamma(1 - \lambda_1)\Gamma(\lambda_1)}, \quad (4.53)$$

$$T = \frac{\Gamma(1 - 2\lambda_0 - \lambda_1)\Gamma(\lambda_1 - 2\lambda_0)}{\Gamma(-2\lambda_0)\Gamma(1 + 2\lambda_0)}. \quad (4.54)$$

In figure 4.3 an example of $|R|^2 = RR^*$ and $|T|^2 = TT^*$ as a function of the Rossby number is plotted for typical values of $m$, $\chi$ and the wave frequency $\omega$.

4.3 Inertia-gravity waves incident upon a hyperbolic cosine wind profile

In this section some preliminary results are presented, concerning the propagation of IGWs in a barotropic jetlike background wind profile. The work presented in this section is a first approach towards an understanding of the dynamical aspects concerning the propagation properties of IGWs in a jetlike flow.

Reduction to Heun’s equation

Eq. (4.9) in section 4.1 will now be studied for a jetlike background wind profile of the type

$$U(y) = \frac{1}{\cosh(y)}. \quad (4.55)$$

The shear $dU/dy$ is represented by

$$\frac{dU}{dy} = -\sinh(y)\text{sech}^2(y). \quad (4.56)$$

Introduce now the following transformation of the independent variable:

$$\eta = e^y. \quad (4.57)$$

Substitution of the identities (4.57) into (4.9) yields

$$\eta^2 \frac{d^2h}{dy^2} + \eta \frac{dh}{d\eta} + \left[ P + 2Q \frac{(1 - \eta^2)\eta}{(1 + \eta^2)(1 + \eta^2)} \right] h = 0, \quad (4.58)$$
Inertia-gravity waves incident upon a hyperbolic cosine wind profile

where

\[ P = (m^2 + \chi^2)(\omega^2 - 1), \quad (4.59) \]
\[ Q = (m^2 + \chi^2)\text{Ro}. \quad (4.60) \]

Eq. (4.58) is a second-order differential equation of Fuchsian type (Kamke, 1959; Whittaker and Watson, 1978) with four regular singularities \( \eta = 0, \eta = +i, \eta = -i \) and \( \eta = \infty \). An efficient method to reduce equations of Fuchsian type, is to apply a power transformation of the dependent variable (Sluijter, 1966; Sluijter and van Duin, 1980; van Duin and Kelder, 1982). To apply this method successfully to Eq. (4.58), the singularities \( \eta = 0, \eta = +i, \eta = -i \) in Eq. (4.58) need to be shifted onto the real axis by means of a coordinate transformation of the independent variable i.e.

\[ \eta = i\xi. \quad (4.61) \]

Substitution of (4.61) into Eq. (4.58) yields

\[ \xi^2 \frac{d^2 h}{d\xi^2} + \xi \frac{dh}{d\xi} + \left[ P + 2iQ \frac{(1 + \xi_0^2)\xi}{(1 - \xi_0^2)^2} \right] h = 0. \quad (4.62) \]

The singularities at \( \eta = 0, \eta = +i, \eta = -i \) are now on the real axis i.e. at, respectively, \( \xi = 0, \xi = +1, \xi = -1 \). The singularity at \( \eta = \infty \) is now at \( \xi = \infty \).

The next step is to determine the Frobenius exponents of the corresponding singularities \( \xi = 0, \xi = +1, \xi = -1 \) and \( \xi = \infty \). The method of Frobenius utilizes the idea that a solution of a differential equation near a singular point can be written as

\[ h(\xi) = (\xi - \xi_0)^s \sum_{n=0}^{\infty} \gamma_n (\xi - \xi_0)^n, \quad (4.63) \]

where \( \xi = \xi_0 \) refers to the singular point in question.

Substitution of (4.63) into Eq. (4.62) at each of the four singular points and collecting terms of the lowest order in \( \xi \) (i.e. \( \xi^0 \)) yields the following equations for the index \( s \),

- \( \xi_0 = 0 \):
  \[ s(s - 1) + s + P = 0 \quad \Rightarrow \quad s = \rho_{1,2} = \pm i P^{1/2}. \quad (4.64) \]

- \( \xi_0 = -1 \):
  \[ s(s - 1) - iQ = 0 \quad \Rightarrow \quad s_{1,2} = \frac{1 \pm (1 + 4iQ)^{1/2}}{2}. \quad (4.65) \]
\[ s(s-1) + iQ = 0 \quad s_{0}^{*} = \frac{1 \pm (1 - 4iQ)^{1/2}}{2} \]  
\[ \sigma_{1,2}^{*} = \frac{1 \pm (1 - 4iQ)^{1/2}}{2} \]  
\[ (4.66) \]

where the asterisk, ‘\( * \)’, indicates complex conjugation.

\[ \xi_{0} = +1: \]

\[ s(s-1) + iQ = 0 \quad s_{0}^{*} = \frac{1 \pm (1 - 4iQ)^{1/2}}{2} \]

\[ (4.66) \]

\[ \xi_{0} = \infty: \]

At infinity Eq. (4.62) is transformed by substituting \( \xi = 1/\zeta \) yielding

\[ \zeta^{2} \frac{d^{2}h}{d\zeta^{2}} + \zeta \frac{dh}{d\zeta} + \left[ P + 2iQ \frac{(1 + \zeta^2)\zeta}{(1 - \zeta^2)^2} \right] h = 0. \]

\[ (4.67) \]

The local behaviour of the solutions of Eq. (4.67) can now be studied at \( \zeta = 0 \).

Substitution of the Frobenius series (4.63) in terms of \( \zeta \) yields to lowest order

\[ s(s-1) + s + P = 0 \quad s_{1,2} = \pm iP^{1/2}. \]

\[ (4.68) \]

Note that the singularities at \( \xi = 0 \) and \( \zeta = 0 \) (i.e. \( \xi = \infty \)) have the same Frobenius exponents.

Now that the Frobenius exponents have been found, Eq. (4.62) can be transformed to the canonical form of Heun’s equation by using a power transformation of the dependent variable i.e.

\[ h(\xi) = \xi^{\rho}(\xi-1)^{\sigma} (\xi+1)^{\sigma_{+}} \hat{h}(\xi), \]

\[ (4.69) \]

where \( \rho \), \( \sigma^{*} \), and \( \sigma \) are the Frobenius exponents of the singularities at \( \xi = 0, \xi = +1 \) and \( \xi = -1 \), respectively, and are given by

\[ \rho = iP^{1/2}, \quad \sigma^{*} = \frac{1 + (1 - 4iQ)^{1/2}}{2}, \quad \sigma = \frac{1 + (1 + 4iQ)^{1/2}}{2}. \]

\[ (4.70) \]

Substitution of (4.69) into Eq. (4.62) yields

\[ \frac{d^{2}\hat{h}}{d\xi^{2}} + A(\xi) \frac{d\hat{h}}{d\xi} + B(\xi) \hat{h} = 0, \]

\[ (4.71) \]

with

\[ A(\xi) = \frac{2\rho + 1}{\xi} + \frac{2\sigma^{*}}{(\xi-1)} + \frac{2\sigma}{(\xi+1)} \]

\[ (4.72) \]
and
\[
B(\xi) = \frac{(2\rho + 1)\sigma^*}{\xi(\xi - 1)} + \frac{(2\rho + 1)\sigma}{\xi(\xi + 1)} + \frac{2\sigma\sigma^*}{(\xi + 1)(\xi - 1)} + \frac{2iQ}{\xi} - \frac{iQ}{(\xi - 1)} - \frac{iQ}{(\xi + 1)} \tag{4.73}
\]

In the derivation above, the last term of the coefficient of \(h\) of Eq. (4.62) has been rewritten through division with \(\xi\) i.e.
\[
2iQ\frac{(1 + \xi^2)}{\xi(1 - \xi)^2(1 + \xi)^2} = \frac{2iQ}{\xi(1 - \xi)^2(1 + \xi)^2} + \frac{2iQ}{(1 - \xi)^2(1 + \xi)^2}. \tag{4.74}
\]

Decomposition of \(F_1\) and \(F_2\) into partial fractions yields
\[
F_1 + F_2 = 2iQ \left( \frac{1}{\xi} + \frac{1}{2(\xi - 1)^2} - \frac{1}{2(\xi - 1)} - \frac{1}{2(\xi + 1)^2} - \frac{1}{2(\xi + 1)} \right),
\]
where we made use of the fact that
\[
\rho(\rho - 1) + \rho + P = 0, \tag{4.75}
\]
\[
\sigma(\sigma - 1) - iQ = 0, \tag{4.76}
\]
\[
\sigma^*(\sigma^* - 1) + iQ = 0. \tag{4.77}
\]

We will now bring all terms of the coefficient \(B(\xi)\) on the same denominator viz. \(\xi(\xi + 1)(\xi - 1)\) for purposes that will become clear later. Eq. (4.71) finally takes the form
\[
\frac{d^2\tilde{h}}{d\xi^2} + C(\xi) \frac{d\tilde{h}}{d\xi} + D(\xi)\tilde{h} = 0, \tag{4.78}
\]

with
\[
C(\xi) = \frac{\lambda}{\xi} + \frac{2\sigma^*}{\xi - 1} + \frac{2\sigma}{\xi + 1}, \tag{4.79}
\]
\[
D(\xi) = \frac{(\lambda(\sigma^* + \sigma) + 2\sigma\sigma^*)\xi + \lambda(\sigma^* - \sigma) - 2iQ}{\xi(\xi + 1)(\xi - 1)}, \tag{4.80}
\]
and \(\lambda = 2\rho + 1\) in (4.80). Eq. (4.78) is a special form of Heun’s equation (for details see the Appendix). Its solutions can be characterized by the Riemann \(P\)-symbol
\[
\tilde{h} = P \left\{ \begin{array}{cccc}
0 & 1 & -1 & \infty \\
0 & 0 & 0 & \sigma + \sigma^* + 2i\sqrt{P} \\
-2i\sqrt{P} & -\sqrt{1 - 4iQ} & -\sqrt{1 + 4iQ} & \sigma + \sigma^* \\
\end{array} \right\}.
\]

The Riemann \(P\)-symbol lists the four regular singularities of Heun’s equation and its Frobenius exponents relative to them.
Reflection and transmission coefficients

Analytical expressions for the reflection and transmission coefficients can be found from the asymptotic behaviour of the solutions of Heun’s equation. Heun’s equation has two solutions around each of the regular singular points. To determine which of these solutions represents the correct asymptotical behaviour, let us return to Eq. (4.9).

From Eq. (4.9) follows that for \( y \to -\infty \) there exists an incident wave with amplitude \( a_1 \) propagating to the right and a reflected wave with amplitude \( a_2 \) propagating to the left i.e.

\[
h(y) = a_1 e^{-iy} + a_2 e^{iy}, \tag{4.81}\]

with \( \ell = ((\omega^2 - 1)(m^2 + \chi^2))^{1/2} \). Similarly for \( y \to \infty \) it is found that

\[
h(y) = a_3 e^{-iy}, \tag{4.82}\]

corresponding to a transmitted wave with amplitude \( a_3 \) propagating to the right.

Furthermore note that it follows from (4.57) and (4.61) that \( b_1 \) corresponds to \( \phi \) and \( b_2 \) corresponds to \( \phi \).

From the investigations above it follows that the correct solutions of Eq. (4.78) is determined by the asymptotical behaviour around \( \phi \) and \( \phi \). Using (4.107-4.108), the solution around \( \phi \) takes the form

\[
h(x) = f(\phi)\tilde{h}(x),
\]

\[
= f(\phi)(b_1\tilde{h}_1(\phi) + b_2\tilde{h}_2(\phi)),
\]

\[
\xi = 0 \quad \xi \to \infty \quad (\xi + 2i\rho^{1/2}),
\]

\[
= (-i)^{P^{1/2}}e^{\frac{i\rho^{1/2}}{B-y}}\left(b_1 + b_2(-i)\right)\left(e^{2i\rho^{1/2}}e^{-2i\rho^{1/2}}\right),
\]

\[
= e^{\frac{i\rho^{1/2}}{B-y}}\left(b_1 e^{\frac{i\rho^{1/2}}{B-y}} + b_2 e^{-\pi B^{1/2}} e^{\frac{i\rho^{1/2}}{B-y}}\right). \tag{4.83}\]

where we have used that the facts \( f(\phi) \equiv \xi^{\eta}(\xi - 1)^{\eta}(\xi + 1)^{\eta} \) and \( \xi = -ie^y \). Note the similarity of the result in the last line of (4.83) with (4.81). Both an incident wave with amplitude \( b_2 e^{-\pi P^{1/2}} \) and a reflected wave with amplitude \( b_1 \) are present in (4.83). The reflection coefficient \( R \) is defined as the ratio of the amplitudes of the reflected and incident wave i.e.

\[
R = \frac{b_1}{b_2 e^{-\pi P^{1/2}}}. \tag{4.84}\]
Suppose now that \( b_2 = e^{\pi P^{1/2}} \) then \( b_1 = R \). Substitution of this in the second identity of (4.83) yields the solution \( h(\xi) \) around \( \xi = 0 \)

\[
h^0(\xi) = f(\xi) \left( R \tilde{h}_1^0(\xi) + e^{\pi P^{1/2}} \tilde{h}_2^0(\xi) \right). \tag{4.85}
\]

Along the same lines the solution \( h \) for \( \xi \to \infty \) can be found. From (4.122-4.123) and (4.101) it is found that

\[
h(\xi) = f(\xi) \tilde{h}(\xi),
\]

\[
\xi \to \infty \Rightarrow \xi^{\sigma+\sigma^*} \left( b_3 \xi^{-\alpha} + b_4 \xi^{-\beta} \right),
\]

\[
= b_3 (\xi^{-iP^{1/2}} e^{-iP^{1/2} y}) + b_4 (\xi^{-iP^{1/2}} e^{iP^{1/2} y}),
\]

\[
= b_3 e^{\pi P^{1/2}/2} e^{-iP^{1/2}/2 y} + b_4 e^{\pi P^{1/2}/2} e^{iP^{1/2}/2 y}. \tag{4.86}
\]

The last identity in (4.86) corresponds to waves of amplitude \( b_3 e^{\pi P^{1/2}/2} \) and \( b_4 e^{\pi P^{1/2}} \) propagating to the right and left, respectively. However, far away from the jet (\( \xi \to \infty \)) only the transmitted wave should be observed. It follows that \( b_1 = 0 \). The transmission coefficient \( T \) is then defined as

\[
T = b_3 e^{-\pi P^{1/2}}. \tag{4.87}
\]

From (4.87) it follows that \( b_3 = T e^{\pi P^{1/2}/2} \); and after substitution in the second identity of (4.86), the solution \( h(\xi) \) around \( \xi = \infty \) takes the form

\[
h^\infty(\xi) = f(\xi) Te^{\pi P^{1/2}/2} \tilde{h}_1^\infty(\xi). \tag{4.88}
\]

Unfortunately, there is no straightforward relationship between \( h^0(\xi) \) and \( h^\infty(\xi) \), which means that

\[
R h_1^0 + e^{\pi P^{1/2}} h_2^0 \neq T e^{\pi P^{1/2}/2} \tilde{h}_1^\infty. \tag{4.89}
\]

This can be understood by the fact that solution \( h^0(\xi) \) and \( h^\infty(\xi) \) do not have a common region of analyticity. From the Appendix it is found that solution \( h^0(\xi) \) is analytic inside the unit circle (Fig. 4.4) whereas solution \( h^\infty(\xi) \) is analytic outside the unit circle.

However by means of the solutions \( h^{-1}(\xi) \) or \( h^+1(\xi) \) it is possible to construct a relationship between the solutions \( h^0(\xi) \) and \( h^\infty(\xi) \). For example, from the Appendix it follows that solution \( h^{-1}(\xi) \) is analytic within a circle with its centre at \( \xi = -1 \) and radius one. The solutions \( h^0(\xi) \) and \( h^{-1}(\xi) \) are now both analytical and
Eqs (4.92) through (4.95) can be further reduced by elimination of $\mu_1$ and $\mu_2$ and

$$
R \tilde{h}_1^0 + e^{\pi P^{1/2}} \tilde{h}_2^0 = \mu_1 \tilde{h}_1^{-1} + \mu_2 \tilde{h}_2^{-1}, \\
\mu_1 \tilde{h}_1^{-1} + \mu_2 \tilde{h}_2^{-1} = T e^{\pi P^{1/2}} \tilde{h}_1^\infty,
$$

(4.90) (4.91)

where $\mu_1$ and $\mu_2$ are constants.

Analytical expressions for the reflection $R$ and transmission coefficient $T$ can be found, by choosing two points in the common area of convergence of the solutions $\tilde{h}_1^0(\xi), \tilde{h}_2^0(\xi)$ and the solutions $\tilde{h}_1^{-1}(\xi)$ and $\tilde{h}_2^{-1}(\xi)$ for example $\xi = -1/2$ and $\xi = -3/4$. Two points in the common area of convergence for the solutions $\tilde{h}_1^{-1}(\xi)$, $\tilde{h}_2^{-1}(\xi)$ and $\tilde{h}_1^\infty(\xi)$ are for example $\xi = -5/4$ and $\xi = -7/4$.

The following equations for analytic continuation are now obtained:

$$
R \tilde{h}_1^0 + e^{\pi P^{1/2}} \tilde{h}_2^0 = \mu_1 \tilde{h}_1^{-1} + \mu_2 \tilde{h}_2^{-1} \quad \text{at } \xi = -1/2, \\
R \tilde{h}_1^0 + e^{\pi P^{1/2}} \tilde{h}_2^0 = \mu_1 \tilde{h}_1^{-1} + \mu_2 \tilde{h}_2^{-1} \quad \text{at } \xi = -3/4, \\
\mu_1 \tilde{h}_1^{-1} + \mu_2 \tilde{h}_2^{-1} = T e^{\pi P^{1/2}} \tilde{h}_1^\infty \quad \text{at } \xi = -5/4, \\
\mu_1 \tilde{h}_1^{-1} + \mu_2 \tilde{h}_2^{-1} = T e^{\pi P^{1/2}} \tilde{h}_1^\infty \quad \text{at } \xi = -7/4.
$$

(4.92) (4.93) (4.94) (4.95)

Eqs (4.92) through (4.95) can be further reduced by elimination of $\mu_1$ and $\mu_2$ and
solve the resulting equations for $R$ and $T$. After some algebra one yields the following expressions for $R$ and $T$:

$$R = -\frac{e^{\pi P^{1/2}}\hat{h}_2^{-1}(-7/4)\hat{h}_1^\infty(-5/4) - \hat{h}_2^{-1}(-5/4)\hat{h}_1^\infty(-7/4)}{\hat{h}_1^{-1}(-7/4)\hat{h}_1^\infty(-5/4) - \hat{h}_1^{-1}(-5/4)\hat{h}_1^\infty(-7/4)}, \quad (4.96)$$

$$T = \frac{e^{\pi P^{1/2}}\hat{h}_1^{-1}(-7/4)\hat{h}_2^{-1}(-5/4) - \hat{h}_1^{-1}(-5/4)\hat{h}_2^{-1}(-7/4)}{\hat{h}_1^{-1}(-7/4)\hat{h}_1^\infty(-5/4) - \hat{h}_1^{-1}(-5/4)\hat{h}_1^\infty(-7/4)}. \quad (4.97)$$

From (4.96) and (4.97) it is clear that $R$ and $T$ depend in a complicated way on the solutions of Heun’s equation around different singularities and at different points. The expressions for the reflection and transmission coefficients, $|R|^2 = RR^* \quad$ and $|T|^2 = TT^*$ will be even more complicated. A numerical approach to solve (4.9) and study the behaviour of $|R|^2$ and $|T|^2$ is therefore preferred. In chapter 5 equation (4.9) will be integrated numerically for the hyperbolic tangent (section 4.2) as well as the hyperbolic cosine wind profile, not only for the case of perpendicular incidence but also for the case of oblique wave incidence ($k \neq 0$). For a wide range of parameter values the behaviour of $|R|^2$ and $|T|^2$ will be studied.
Appendix: General features of Heun’s equation

Here some aspects of the theory of Fuchsian differential equations and in particular
the hypergeometric and Heun’s differential equation will be briefly reviewed. For a
detailed overview of Fuchsian differential equations and singular differential equa-
tions in general the reader is referred to Erdélyi (1942), Ince (1956) and Kamke
(1959). A comprehensive overview of Heun’s differential equation is given in Ron-
veaux (1995). The properties of Heun’s equation presented below will be used in the
derivation in the previous section.
A differential equation of Fuchsian type is an ordinary linear differential equation
in which every singular point, including the point at infinity, is a regular singularity
(Ince, 1956). The singular points do not necessarily have to be real. More specific,
a Fuchsian equation of second order with four regular singular points including the
point at infinity is called Heun’s equation. A Fuchsian equation of second order with
three regular singular points including the point at infinity is called the hypergeome-
tric equation. The latter is in fact a limiting case of Heun’s equation just like other
differential equations, for example the Lamé, Weber and Legendre equations.

Here we will consider Heun’s equation in its general canonical form (Arscott,
1995; Snow, 1952).

\[
\frac{d^2 \tilde{h}}{d\xi^2} + \left( \frac{\gamma}{\xi} + \frac{\delta}{\xi - 1} + \frac{\epsilon}{\xi - a} \right) \frac{d\tilde{h}}{d\xi} + \frac{\alpha \beta \xi - q}{\xi (\xi - 1)(\xi - a)} \tilde{h} = 0,
\]  

(4.98)

with \( \tilde{h} = h(\xi) \) is a complex function of the complex variable \( \xi \), and has complex
parameters \( \alpha, \beta, \gamma, \delta, \epsilon \) and \( a \), and \( q \) the accessory parameter. Parameter \( a \) is allowed
to take all values except \( a = 0 \) and \( a = 1 \) nor should \( a = \infty \). The parameters are
linked by the relation

\[
\alpha + \beta + 1 = \gamma + \delta + \epsilon.
\]  

(4.99)

Eq. (4.98) has regular singularities at \( \xi = 0, \xi = 1, \xi = a \) and \( \xi = \infty \).

In general, solutions of Eq. (4.98) can be found in terms of Frobenius series

\[
\tilde{h}(\xi) = (\xi - \xi_0)^s \sum_{n=0}^{\infty} c_n (\xi - \xi_0)^n,
\]  

(4.100)

around the singular point \( \xi = \xi_0 \). The Frobenius exponents \( s \) can found as the roots
of the so-called indicial equation, arising from the lowest power in \( \xi - \xi_0 \).

For \( \xi = 0, \xi = 1, \xi = a \) and \( \xi = \infty \) the Frobenius exponents \( (s_1, s_2) \) are
\( \{s_1, s_2\} = \{0, 1 - \gamma\}; \{0, 1 - \delta\}; \{0, 1 - \epsilon\} \) and \( \{\alpha, \beta\} \), respectively. The solution
of Heun’s equation can be characterized by the Riemann P-symbol

\[ \tilde{h} = P \left\{ \begin{array}{ccc}
0 & 1 & a \\
0 & 0 & 0 \\
1 - \gamma & 1 - \delta & 1 - \epsilon \\
\end{array} \right\} \xi, \]

which lists the singular points with its accompanying exponents, and by the accessory parameter.

Note that from the discussion in the main text that \( a = -1 \) and that

\[
\gamma = 1 + 2iP, \\
\delta = 1 + \sqrt{1 - 4iQ}, \\
\epsilon = 1 + \sqrt{1 + 4iQ}, \\
\alpha = 2iP + (\sigma + \sigma^*), \\
\beta = (\sigma + \sigma^*), \\
q = -(1 + 2iP)(\sigma^* - \sigma) + 2iQ.
\]

with \( P, Q, \sigma^* \) and \( \sigma \) defined in (4.58) and (4.70).

**Solutions of Heun’s equation**

The general solution of Heun’s equation can be given in terms of Frobenius series. From the asymptotic analysis in section 3 of the main text, solutions of Heun’s equation around \( \xi = 0, \xi = -1 \) and \( \xi = \infty \) are considered. Here the local solutions around \( \xi = 0, \xi = -1 \) and \( \xi = \infty \) will be analysed in more detail. Recurrence relations will be derived and the asymptotic behaviour near of the solutions is investigated.

- **Solutions around \( \xi = 0 \)**

  Consider Heun’s equation in case \( a = -1 \) and \( \alpha, \beta, \gamma, \delta, \epsilon \) and \( q \) as given in (4.101). Multiply equation (4.98) with \( \xi^{-s} \) into equation (4.98) and collect equal powers in \( \xi \). After some algebra and rearranging of terms one obtains:

\[
- c_0[(s - 1)s + \gamma s]\xi^{s-1} - c_1[(s + 1)s + \gamma(s + 1)]\xi^s - c_0[-\delta s + \epsilon s + q]\xi^s - \sum_{n=1}^{\infty}[-\delta + \epsilon](s + n) + q]c_n\xi^{s+n} - \sum_{n=1}^{\infty}[(s + n + 1)(s + n + \gamma)]c_{n+1}\xi^{s+n} + \sum_{n=1}^{\infty}(s + n - 2)(s + n - 1) + (\gamma + \delta + \epsilon)(s + n - 1) + \alpha(\beta)c_{n-1}\xi^{s+n} = 0.
\]

(4.102)
The indicial equation arises from the lowest power in $\xi$ (i.e. $\xi^{s-1}$) and is found to be

$$(s - 1)s + \gamma s = 0 \implies s = 0 \lor s = 1 - \gamma,$$  

(4.103)

where we have adopted the normalization (Heun, 1889)

$$c_0 = 1.$$  

(4.104)

The coefficient of the next power yields

$$c_1 = \frac{-\delta s + \epsilon s + q}{-(s + 1)(s + \gamma)},$$  

(4.105)

with $\gamma \neq 0$ if $s = 0$. The next power yields the three-term recurrence relation for $\gamma \geq 1$:

$$-[(\delta + e)(s + n) + q]c_n - [(s + n + 1)(s + n + \gamma)]c_{n+1} + [-(s + n - 2)(s + n - 1) + (\gamma + \delta + e)(s + n - 1) + \alpha\beta]c_{n-1} = 0.$$  

(4.106)

Note that the coefficient of $c_{n+1}$ becomes zero if $\gamma = 0, -1, -2, ...$ and if $\gamma \neq 3, 4, ...$ in case $s = 1 - \gamma$.

With (4.104)-(4.106) the solutions around $\xi = 0$ can be written as:

$$\tilde{h}_1^0(\xi) = 1 + c_1\xi + \sum_{n=2}^{\infty} c_n\xi^n \quad \text{for } s = 0,$$  

(4.107)

and

$$\tilde{h}_2^0(\xi) = \xi^{1-\gamma} \left[ 1 + c_1\xi + \sum_{n=2}^{\infty} c_n\xi^n \right] \quad \text{for } s = 1 - \gamma,$$  

(4.108)

where the subscript refers to the value of the exponent. These solutions may be represented more generally by the symbol $H$ as follows

$$\tilde{h}_1^0(\xi) = H(a, q; \alpha, \beta, \gamma, \delta; \xi),$$  

(4.109)

and

$$\tilde{h}_2^0(\xi) = \xi^{1-\gamma} H(a, q^4; 1 + \alpha - \gamma, 1 + \beta - \gamma, 2 - \gamma, \delta; \xi),$$  

(4.110)

1In some textbooks the notation $F(a, q; \alpha, \beta, \gamma, \delta; \xi)$ is used. However this symbol is generally used to represent the solutions of the hypergeometric equation. Here we prefer to use the notation with $H$ instead of $F$. 

90 Propagating inertia-gravity waves in barotropic shear layers
Inertia-gravity waves incident upon a hyperbolic cosine wind profile

with

\[ q^s = q + (1 - \gamma)(-\delta + \epsilon). \quad (4.111) \]

The series (4.109) and (4.110) converge inside the unit circle in the complex \( \xi \)-plane centered at the origin and with a radius up to the nearest other singular point.

- Solutions around \( \xi = -1 \)

The construction of the solutions around \( \xi = -1 \) can in principle be done along the same lines as shown above for the solution around \( \xi = 0 \). After substitution of (4.100) with \( \xi = -1 \) and collecting equal powers in \( \xi + 1 \), the result for the lowest power in \( \xi + 1 \) yields the indicial equation:

\[ s(s - (1 - \epsilon)) = 0 \iff s = 0 \vee s = 1 - \epsilon, \quad (4.112) \]

Again we adopt the normalization (4.104). The coefficient of the next power then becomes

\[ c_1 = \frac{3s(s - 1) + \gamma s + (\gamma + \delta + \epsilon)s + \alpha \beta + q}{2(s + 1)(s + \epsilon)}. \quad (4.113) \]

Note that \( \epsilon \neq 0 \) if \( s = 0 \) and \( \epsilon \neq 2 \) if \( s = 1 - \epsilon \). The next power yields the three-term recurrence relation for \( n \geq 1 \)

\[
\begin{align*}
[-3(s + n)(s + n - 1) - 2\gamma(s + n) - (\delta + \epsilon)(s + n) - \alpha \beta - q]c_n &+ \\
[(s + n - 1)(s + n - 2) + (\gamma + \delta + \epsilon)(s + n - 1) + \alpha \beta]c_{n-1} &+ \\
2[(s + n + \epsilon)(s + n + 1)]c_{n+1} &= 0. \quad (4.114)
\end{align*}
\]

with \( \epsilon \neq 0, -1, -2, \ldots \) in case \( s = 0 \) and \( \epsilon \neq 3, 4, \ldots \) in case \( s = 1 - \epsilon \) otherwise the coefficient of \( c_{n+1} \) becomes zero.

From (4.112)-(4.114) we may construct the solutions around \( \xi = -1 \)

\[ \tilde{h}_1^{-1}(\xi) = 1 + c_1(\xi + 1) + \sum_{n=2}^{\infty} c_n(\xi + 1)^n \quad \text{for } s = 0, \quad (4.115) \]

and

\[ \tilde{h}_2^{-1}(\xi) = (\xi + 1)^{1-\epsilon} \left[ 1 + c_1(\xi + 1) + \sum_{n=2}^{\infty} c_n(\xi + 1)^n \right] \quad \text{for } s = 1 - \epsilon. \quad (4.116) \]
The region of analyticity of the series (4.115) and (4.116) is inside a circle in the complex $\xi$-plane with its centre at $\xi = -1$ and with a radius up to the nearest other singular point.

- Solutions near the singular point $\xi = \infty$

In order to find solutions at $\xi = \infty$ a transformation of the independent variable $\xi$ is required, i.e. replace $\xi$ by $1/\chi$ in Heun’s equation and study the resulting equation in $\chi = 0$. The transformed Heun’s equation (4.98) for $\alpha = -1$ reads

$$\frac{d^2 w}{d\chi^2} + \left[\frac{(2 - \gamma)}{\chi} - \frac{\delta}{1 - \chi} - \frac{\epsilon}{1 + \chi}\right] \frac{dw}{d\chi} + \left[\frac{\alpha \beta \chi^{-1} - q}{\chi(1 - \chi)(1 + \chi)}\right] w = 0.$$  

(4.117)

We may construct solutions around $\chi = 0$ using the same procedures as used above. First of all we multiply equation (4.117) by $\chi^{-(3/2)}$ and substitute into (4.117) a Frobenius series of the type:

$$w(\xi) = \sum_{n=0}^{\infty} c_n \chi^{s+n}.$$  

(4.118)

Finally we collect equal powers of $\chi$. The lowest power in $\chi$ yields the indicial equation:

$$s^2 - (\alpha + \beta)s + \alpha \beta = 0 \implies s = \alpha \quad \lor \quad s = \beta.$$  

(4.119)

Again we adopt the normalization (4.104). The result of the coefficient of the next power in $\chi$ yields:

$$c_1 = \frac{(\delta + \epsilon)s + q}{(s + 1)(s + 2) - (s + 1)(\gamma + \delta + \epsilon) + \alpha \beta}.$$  

(4.120)

The three-term recurrence relation for $n \geq 1$ reads

$$[(-\delta + \epsilon)(s + n) - q]c_n - [(s + n - 1)(s + n - 2) + (2 - \gamma)(s + n - 1)]c_{n-1} + [(s + n)(s + n + 1) + (2 - \gamma)(s + n + 1) - (\delta + \epsilon)(s + n + 1) + \alpha \beta]c_{n+1} = 0.$$  

(4.121)

The solutions around $\xi = \infty$ can now be written down with the use of (4.119)-(4.121) and using the transformation $\chi = 1/\xi$

$$h_1^\infty(\xi) = \xi^{-\alpha} \left[1 + \frac{c_1}{\xi} + \sum_{n=2}^{\infty} c_n \xi^{s-n}\right] \quad \text{for } s = \alpha.$$  

(4.122)
Inertia-gravity waves incident upon a hyperbolic cosine wind profile

and

\[ \tilde{h}_2^\infty(\xi) = \xi^{-\beta} \left[ 1 + \frac{c_1}{\xi} + \sum_{n=2}^{\infty} c_n \xi^{-n} \right] \quad \text{for } s = \beta. \quad (4.123) \]

The series (4.122) and (4.123) converge outside the unit circle in the complex \( \xi \)-plane.
Chapter 5

Propagation properties of inertia-gravity waves through a barotropic shear layer and application to the Antarctic polar vortex

The propagation of inertia-gravity waves (IGWs) through a dynamical transport barrier, such as the Antarctic polar vortex edge has been investigated using a linear wave model. The model is based on the linearized, inviscid hydrostatic equations on an $f$-plane. Typical values for the parameters that are appropriate to the Antarctic polar vortex are assumed. The background flow $U$ is assumed to be barotropic and its horizontal shear is represented by a hyperbolic tangent background wind profile. The wave equation that describes the latitudinal structure of a monochromatic disturbance contains two singularities. The first corresponds to the occurrence of a critical level where the intrinsic wave frequency $\Omega = \omega - kU$ becomes zero. $\omega$ is the absolute wave frequency and $k$ its longitudinal wavenumber in the direction of $U$. The second is an apparent singularity and does not give rise to singular wave behaviour. It becomes zero whenever the square of the intrinsic wave frequency $\Omega^2 = (f - U_y)$, $f$ being the Coriolis frequency and $U_y$ the horizontal shear of the flow. The wave equation is solved numerically for different values of the angles of incidence of the wave upon the background flow, and for different values of the wave frequency, the horizontal wavenumber and the Rossby number. Reflection ($|R|$) and transmission ($|T|$) coefficients are determined as a function of these parameters. The results de-
pend on whether the flow is inertially stable or not. They also depend on the presence and location of the turning levels, where the wave becomes evanescent, with respect to the location of the $\Omega$-critical levels. For inertially stable flows, the wave totally reflects at the turning level and never reaches the critical level. If the background flow is inertially unstable, turning levels can disappear and the wave can now reach the critical level. Then overreflection, overtransmission and absorption can occur.

5.1 Introduction

The dynamics of the Antarctic polar vortex in the lower and middle stratosphere has been much studied during the last 20 years (e.g. Schoeberl and Hartmann, 1991; Randel, 1993; McIntyre, 1995). They follow early studies by e.g. Juckes and McIntyre (1987) and McIntyre (1989) which have shown that such a vortex behaves like a dynamical barrier for atmospheric minor constituents.

The Antarctic polar vortex is characterized by a strong gradient in potential vorticity (the vortex edge) which is nearly impermeable to transports induced by large scale motions (e.g. Bowman, 1993a; Chen, 1994; Öllers et al., 2002b), but can be more porous through transports related to smaller scale IGWs (McIntyre, 1995). Nevertheless, very few studies have actually addressed how IGWs dynamically interact with such a vortex. Dunkerton (1984) investigated the propagation and refraction properties of stationary IGWs in zonal mean flows in the stratosphere using a ray-tracing technique. Pierce et al. (1994) performed Lagrangian material line calculations and investigated the impact of an idealized IGW-field on irreversible mixing and stretching of material lines near the Arctic and Antarctic vortex edges.

More fundamentally but still in the context of wave-barrier interaction, Staquet and Huerre (2002) have analyzed the breaking of IGWs in a rotating barotropic flow $U$ with horizontal shear near an $N$-critical level, where the intrinsic wave frequency ($\Omega = \omega - kU$, $\omega$ being absolute wave frequency and $k$ is the wavenumber in the direction of $U$) equals the Brunt-Väisälä frequency ($N$), by using a three-dimensional Boussinesq fully non-linear model. The interaction between oceanic internal gravity waves and barotropic background flows with horizontal shear has been studied by Ivanov and Morozov (1974); Olbers (1981) and Basovich and Tsimring (1984). Basovich and Tsimring (1984) showed using the linearized Boussinesq equations that, depending on specific wave and background flow characteristics, a wave may be partially absorbed (wave energy is released into the background flow), totally reflected (no energy exchange) or overreflected (the wave extracts energy from the background flow) at $N$-critical levels ($\Omega = N$) or $\Omega$-critical levels ($\Omega = 0$). These type of critical levels have not much been studied in the past for IGWs propagating towards a barotropic flow. A plausible reason for this is that in most circumstances
the critical levels are preceded by turning levels where WKB theory predicts total reflection. Depending on whether the Coriolis frequency $f$ is assumed to be constant or changing with latitude, waves can, respectively, be absorbed (Jones, 1967) or reflected (e.g. Kitchen and McIntyre, 1980) at Jones critical levels ($\Omega = f$).

Although the IGWs interaction with a barotropic jet is little studied, the propagation of internal gravity waves in background flows with vertical shear has been extensively studied since the seminal paper of Booker and Bretherton (1967). In the presence of an $\Omega$-critical level the value of the Richardson number plays a key role in the stability of the flow and determines whether waves will be absorbed (Booker and Bretherton, 1967) or overreflected at the critical level (Jones, 1968; Acheson, 1976; van Duin and Kelder, 1982).

In the present study, the interaction between monochromatic IGWs and a barotropic background flow is studied in a linear hydrostatic model. The background flow is represented by a hyperbolic tangent profile. The same qualitative behaviour, with respect to the reflection and transmission properties, is found for jet-type profiles mimicking for example the Antarctic polar vortex edge.

From the basic model equations a wave equation is derived that describes this interaction. The wave equation, that describes the disturbance field, is the $f$-plane version of Laplace’s tidal equation with shear in the zonal background flow in the latitudinal direction (Flattery, 1967; Longuet-Higgins, 1968; Dunkerton, 1990). It contains two singularities that are of different nature. One is similar to that found by Booker and Bretherton (1967). The second one is described by Boyd (1978) and Dunkerton (1990) for the case the background flow has merely latitudinal shear, by Yamanaka and Tanaka (1984) for the case of a vertically sheared background flow and the combined case of latitudinally and vertically sheared background flows is described by Kitchen and McIntyre (1980). This singularity, denoted as $A$ here or $-\Delta$ in Boyd (1978) and Dunkerton (1990), occurs whenever the square of the intrinsic wave frequency $\Omega$ equals $f(f - dU/dy)$, $dU/dy$ being the shear of the background flow $U$ in latitudinal direction. It has been shown by Boyd (1976) that the zeroes of $A$ are apparent singularities. No jump in the momentum flux is associated with these apparent singularities.

The parameters are chosen to compare to the Antarctic polar vortex. Reflection and transmission coefficients are determined for waves incident upon the background flow as a function of several characteristic parameters, and are discussed with regard to the degree of inertial instability, that is measured here by a Rossby number.

Although our model is used to study the basic interactions between IGWs and the Antarctic polar vortex, its range of application is much wider and may also be used to study the interaction between IGWs and horizontally-sheared background flows on an $f$-plane in more general conditions such as at midlatitudes and the tropics (see
Inertia-gravity waves in barotropic shear layers

e.g. Dunkerton (1990) and references therein).

This paper should be considered as a first approach in understanding the basic mechanisms of the interaction between IGWs and strong barriers like the Antarctic polar vortex edge and their effect upon its permeability.

The paper is organized as follows. In section 2 the mathematical background of our model is presented. In section 3 some details are given on the numerical integration procedure. The choice of typical values for the parameters is outlined and the main numerical results are at the end of section 3. In the final section we summarize our major results.

5.2 Model formulation

5.2.1 Basic model equations

The linearized equations in log-pressure coordinates for hydrostatic perturbations in a horizontal shear flow $U(y)$ are,

\[ \frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} + v \frac{dU}{dy} - f v = -\frac{\partial \phi}{\partial x}, \quad (5.1) \]

\[ \frac{\partial v}{\partial t} + U \frac{\partial v}{\partial x} + f u = -\frac{\partial \phi}{\partial y}, \quad (5.2) \]

\[ \frac{\partial^2 \phi}{\partial t \partial z} + U \frac{\partial^2 \phi}{\partial x \partial z} + N^2 w = 0, \quad (5.3) \]

\[ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} - \frac{\partial w}{\partial z} - \frac{w}{H} = 0, \quad (5.4) \]

where $u, v$ and $w$ are latitudinal, longitudinal and vertical velocity perturbations, respectively, $\phi$ corresponds to the geopotential, $f$ is the Coriolis frequency and $H$ is the vertical scale height. In Eq. (5.3) $N^2 = (g/\theta_0)d\theta_0/dz$ is the Brunt–Väisälä frequency, where $g$ is the gravitational acceleration and $\theta_0$ is the background potential temperature. Hereafter it is assumed that $N^2$ is constant, which is a reasonable approximation since in the lower and middle stratosphere $\theta_0$ is a nearly linear function of height (Andrews et al., 1987). The assumption of hydrostatic equilibrium is justified here since we are primarily interested in low frequency IGWs with horizontal scales much larger than the vertical scales. Thus the possibility that $N$-critical levels occur is excluded.

Finally $U(y)$ is the background flow, which is assumed to be independent of $z$ and to have a barotropic shear in the horizontal $y$-direction only. In the following, $U(y)$ will be modelled by a hyperbolic tangent or transitional profile,

\[ U(y) = \frac{U_0}{2}(1 + \tanh(y/2L)). \quad (5.5) \]
In Eq. (5.5), $U_0$ is a characteristic maximum velocity of the background flow and $L$ is a measure of the shear layer width.

Assume now wave-like disturbances $\zeta(x,y,z,t) = \hat{\zeta}(y)e^{i(\omega t-kx-mz)+z/2H}$, where $\zeta$ is a general notation for $u,v,w$ and $\phi$. After substitution in Eqs. (5.1) through (5.4), the following wave equation for $\phi$ can be derived

$$\frac{d^2 \phi}{dy^2} - \frac{1}{A} \frac{dA}{dy} \frac{d\phi}{dy} + \left[ \frac{A}{\lambda^2} - \frac{k f}{\lambda \Omega} \frac{dA}{dy} - k^2 \right] \phi = 0,$$

(5.6)

with

$$\Omega = (\omega - k U), \quad A = \Omega^2 - \omega_i^2, \quad \lambda^2 = N^2 / (1/4H^2 + m^2).$$

(5.7)

In (5.7) $\omega_i^2$ is defined as

$$\omega_i^2 = f(f - dU/dy).$$

(5.8)

Eq. (5.6) becomes singular whenever the intrinsic wave frequency $\Omega$ is equal to zero. In that case the horizontal phase speed of the wave equals that of the background flow. Note that this type of singularity is similar to the classical Booker and Bretherton-type singularity for internal gravity waves propagating in a horizontally uniform wind field with vertical shear (Booker and Bretherton 1967). However, the $\Omega$-singularity in our case is a logarithmic singularity (see appendix), whereas the Booker and Bretherton-type singularity is non-logarithmic. Note that $A = 0$ will not give rise to singular behaviour in $\phi$, since $A$ is an apparent singularity. We note that the background flow becomes inertially unstable if $\omega_i^2 < 0$ for certain values of $y$.

Eq. (5.6) can be written into its normal form using $\hat{\phi}(y) = A^{1/2} \psi(y)$ i.e.

$$\frac{d^2 \psi}{dy^2} + \left[ -\frac{3}{4A^2} \left( \frac{dA}{dy} \right)^2 + \frac{1}{2A} \frac{d^2 A}{dy^2} + \frac{A}{\lambda^2} - \frac{k f}{\lambda \Omega} \frac{dA}{dy} - k^2 \right] \psi = 0.$$

(5.9)

Turning levels occur at locations where $Q(y) = 0$. Note that $Q(y)$ will become singular if $\Omega = 0$ and $A = 0$.

5.2.2 Conservation of momentum fluxes

Eq. (5.6) can be rewritten in the form

$$\frac{d}{dy} \left( \frac{1}{A} \frac{d\phi}{dy} \right) + \left[ \frac{1}{\lambda^2} - \frac{k f}{\lambda \Omega} \frac{dA}{dy} - \frac{k^2}{A} \right] \phi = 0.$$

(5.10)
Now, multiply Eq. (5.10) by $i\hat{\phi}^*$ yielding
\[
\frac{d}{dy} \left( \frac{i\hat{\phi}^*}{A} \frac{d\hat{\phi}}{dy} \right) - \frac{i}{A} \frac{d\hat{\phi}^*}{dy} + iP\hat{\phi}^* = 0,
\]
where the asterisks '*' indicate complex conjugation. It now follows that
\[
C = \text{Re} \left( \frac{i\hat{\phi}^*}{A} \frac{d\hat{\phi}}{dy} \right), \tag{5.11}
\]
which does not depend on $y$ as long as $\Omega \neq 0$. Also, the quantity in (5.11) is closely related to a momentum flux. From Eqs. (5.1) and (5.2) it can be derived that the momentum flux
\[
\tau \equiv \langle \rho_0 u v \rangle = \rho_0 \text{Re} \left( \frac{i k \hat{\phi}^*}{A} \frac{d\hat{\phi}}{dy} \right), \tag{5.12}
\]
where $\rho_0$ corresponds to the background density of air and $\langle . \rangle$ denotes an average over a wave cycle.

An IGW incident upon the transitional wind profile (5.5) from $y \to -\infty$ gives rise to a reflected wave there and transmitted wave as $y \to \infty$. The solution for $y \to -\infty$ reads
\[
\hat{\phi}(y) = \exp(-iz_{-\infty}y) + R \exp(iz_{-\infty}y), \tag{5.13}
\]
the amplitudes of the incident and reflected waves being 1 and $R$, respectively. $z_{-\infty}$ corresponds to the horizontal wavenumber in $y$-direction evaluated at $y = -\infty$ and is defined by
\[
l_{-\infty} = \left( \frac{(\omega^2 - f^2)}{A^2} - k^2 \right)^{1/2}, \tag{5.14}
\]
As $y \to \infty$ the solution takes the form
\[
\hat{\phi}(y) = T \exp(-iz_{+\infty}y), \tag{5.15}
\]
where $T$ is the amplitude of the transmitted wave and $l_{+\infty}$ is defined by
\[
l_{+\infty} = \left( \frac{(\omega - kU_0)^2 - f^2}{A^2} - k^2 \right)^{1/2}, \tag{5.16}
\]
Numerical results

and means the value \( l \) takes as \( y \to \infty \). If Eq. (5.6) has no real singularities (\( \Omega \neq 0 \)), we have

\[
C_{-\infty} = C_{+\infty},
\]

(5.17)

with

\[
C_{-\infty} = \text{Re} \left( \frac{i}{A} \frac{d\hat{\phi}}{dy} \right) \bigg|_{y=-\infty}, \quad C_{+\infty} = \text{Re} \left( \frac{i}{A} \frac{d\hat{\phi}}{dy} \right) \bigg|_{y=+\infty}.
\]

(5.18)

Using the results (5.13) through (5.16) one derives for the hyperbolic tangent profile (5.5) that

\[
|R|^2 + \frac{l_{+\infty}}{A_{+\infty}} \frac{A_{-\infty}}{l_{-\infty}} |T|^2 = 1.
\]

(5.19)

For IGWs incident upon a symmetrical jet-type wind profile we would have that \( A_{-\infty} = A_{+\infty} \) and \( l_{-\infty} = l_{+\infty} \), resulting in

\[
|R|^2 + |T|^2 = 1.
\]

(5.20)

If \( \Omega \) does become zero (i.e. \( \Omega \)-critical levels are present) for particular values of \( y \) between \( y = -\infty \) and \( y = +\infty \), it is expected that the quantity \( C \) in (5.11) is discontinuous at these locations (critical levels). In that case (5.19) should fulfil

\[
|R|^2 + \frac{l_{+\infty}}{A_{+\infty}} \frac{A_{-\infty}}{l_{-\infty}} |T|^2 = 1 + \frac{A_{-\infty}}{l_{-\infty}} (C_{+\infty} - C_{-\infty}).
\]

(5.21)

Eq. (5.21) shows that the sum of \(|R|^2\) and \((l/A)_{-\infty} (A/l)_{-\infty} |T|^2\) may differ from 1 depending on the magnitude of the jump \((A_{-\infty}/l_{-\infty})(C_{+\infty} - C_{-\infty})\) in passing critical levels. Whenever \( \Omega \) becomes zero somewhere between \( y = -\infty \) and \( y = +\infty \), the momentum flux \( \tau \) in (5.12) will jump in passing these \( \Omega \)-critical levels. Analytical expressions for such jumps are given in the appendix. There it is shown that the sign of \((C_{+\infty} - C_{-\infty})\) depends upon the sign of \((d^2U/dy^2)/(dU/dy)\)|\(y=y_c\), where \( y_c \) is the location of the \( \Omega \)-critical level. If \((d^2U/dy^2)/(dU/dy)\)|\(y=y_c\) is negative, that is in our case \( y_c > 0 \), \((C_{+\infty} - C_{-\infty}) > 0\) resulting in overreflection and overtransmission. For \( y_c < 0 \), \((d^2U/dy^2)/(dU/dy)\)|\(y=y_c\) is positive and \((C_{+\infty} - C_{-\infty}) < 0\), implying resonant absorption of wave energy at the \( \Omega \)-critical level.

5.3 Numerical results

For \( k \neq 0 \), the singular wave equation in (5.6) seems too complicated to allow for any analytical solutions. Therefore, Eq. (5.6) is integrated numerically. Reflection and transmission coefficients are then calculated for several values of the wave parameters.
5.3.1 Numerical solution method

Eq. (5.6) is solved numerically using a 7–8th-order Runge–Kutta algorithm. The stepsize is automatically reduced when the desired accuracy is not found. An imaginary component of about $10^{-8}$ for the wave frequency $\omega$ is introduced, which acts as a small linear Rayleigh damping. As a consequence, the singularities are then shifted off the real $y$-axis which enabled integration along this axis. The integration is started for large positive values of $y$ with a transmitted wave. Eq. (5.6) is then integrated backwards along the $y$-axis to large negative values of $y$ where the solution is split into an incident and a reflected wave.

The ability of the numerical code to integrate across regions where singularities occur is tested in two ways. First, Eq. (5.6) with $k = 0$ and for a hyperbolic tangent profile (5.5) is integrated numerically and reflection and transmission coefficients are computed. Analytical expressions for reflection and transmission coefficients are also derived by solving Eq. (5.6) in terms of hypergeometric functions (see chapter 4, section 4.2). The numerical and analytical results are compared and the difference is less than 0.005%. Note that for $k = 0$, Eq. (5.6) contains no $\Omega$-type singularity but only an apparent $A$-type singularity. The code integrates through regions where $A = 0$, but as expected, no singular behaviour is observed in the solution $\phi$.

Another test for the code is provided by the analytical results of van Duin and Kelder (1982). Their Eq. (2.1) is integrated using the Runge–Kutta code, after which reflection and transmission coefficients are determined. The numerical results are compared with their analytical expressions for the reflection and transmission coefficients (their Eqs. (4.11) and (4.12)). The differences between the numerical and analytical results are of the same order as above.

5.3.2 Choice of parameters

Realistic values appropriate to the austral winter and early spring polar vortex, are chosen for the parameters $f$, $N$, $L$ and $U_0$. The value of the Coriolis parameter is calculated at 65°S yielding $-1.3 \times 10^{-4}$ s$^{-1}$. 65°S corresponds roughly with the average location of the polar vortex derived from ECMWF analyses for August 1998. From these same analyses, maximum zonal wind speeds in the Antarctic polar vortex are found to be around 60 ms$^{-1}$, and the width of the polar vortex is estimated to be of the order 5-10$^3$ m. The Brunt–Väisälä frequency $N$ in the lower stratosphere is of the order 2-10$^{-2}$ s$^{-1}$ (Pfenninger et al., 1999; Yoshiki and Sato, 2000). The scale height $H$ is taken to be 7 km. The Rossby number ($Ro$) is defined as $Ro = U_0 / f |L|$. Hereafter, $Ro$ is set to 1.0 for inertially stable flows and for inertially unstable flows $Ro$ is kept constant at 4.5. The horizontal wavelength, corresponding with $k$, will vary between approximately 100 and 1000 km and the horizontal wavelength, corre-
Numerical results

sponding with $l_{-\infty}$, is taken to be approximately 314 km. The vertical wavelength is
taken to be 5 km. An IGW incident upon a background flow will be determined by
its wave frequency $\omega$ and its parallel, transverse and vertical wavenumbers $k, l$ and
$m$, respectively. We assume an IGW incident from $y = -\infty$ having a real transverse
wavenumber $l_{-\infty} > 0$, to ensure propagation toward the shear layers, on the back-
ground flow. The subscript ‘$-\infty$’ in $l_{-\infty}$ indicates the value of $l$ at $y = -\infty$. Then
two angles of incidence can be defined for an IGW incident upon a background flow
viz.

$$\alpha = \arctan \left( \frac{m}{l_{-\infty}} \right), \quad \beta = \arctan \left( \frac{k}{l_{-\infty}} \right). \quad (5.22)$$

Using the parameter values above, $\alpha$ is then equal and taken constant to approxi-
mately $89^\circ$. The angle of incidence $\beta$ will vary. Typical values for the wave fre-
quency $\omega$ are between 1 and 5 times the Coriolis frequency $f$. Overall these values
are in good agreement with recent observations made by Guest et al. (2000) for IGWs
in the Southern Hemisphere lower stratosphere.

5.4 Results

5.4.1 Inertially stable background flow

In Figure 5.1 the locations of the turning levels and $\Omega$-critical levels ($\Omega$-CLs hereafter)
are plotted as a function of the angle of incidence $\beta$. From this figure, the following
four cases can be distinguished for increasing $\beta$:

1a. No turning levels and no $\Omega$-CLs
2a. 1 turning level and no $\Omega$-CLs
3a. 1 turning level followed by 1 $\Omega$-CL
4a. 1 $\Omega$-CL surrounded by 2 turning levels

For $\beta$ larger than $22^\circ$, the $\Omega$-CL gradually shifts towards negative values of $y$, but is
always surrounded by 2 turning levels. In fact, close inspection of the behaviour of
$Q$ in Eq. (5.9) shows that a third turning level is present in the flow which nearly
coincides with the $\Omega$-CL. However, the distance between this turning level and the
$\Omega$-CL is much smaller than the wavelength. Therefore, this turning level will not
have any significant effect on the propagational behaviour of the wave and will be
excluded from any further discussion below. In Figure 5.2 reflection and transmission
coefficients are plotted for the same values of $\beta$ as in Figure 5.1. For all values of $\beta$
Figure 5.1: Locations of the turning levels and $\Omega$-critical levels as a function of the angle of incidence $\beta$. The background flow is inertially stable.

Figure 5.2: Reflection and transmission coefficients as a function of the angle of incidence $\beta$. The background flow is inertially stable.
Numerical results

1 turning level  no $\Omega$-critical level, stable profile in tanh

$|R|=1.0, |T|=0.0, \beta = 11 \text{ degrees}$

Figure 5.3: Situation with 1 turning level and no $\Omega$-CLs ($\Omega^2/f^2 > 0$) for a stable

background flow ($\omega^2/f^2 > 0$ for all $y/L$). The wave field is represented by $\text{Re}(\phi)$. The

turning level (‘TL’) occurs at the location where $Q = 0$. $QN^2/f^2L^2$ becomes singular

where $A = 0$, which follows from relation (5.9). The angle of incidence $\beta$ is equal to $11^\circ$.

in Figure 5.2, $|R|^2 + (i/A)_\infty (A/l)_\infty |T|^2 = 1$. This may appear rather surprising,

since one might expect overreflection, overtransmission or absorption to occur in the

presence of an $\Omega$-CL (see appendix). However, for the inertially stable flow it is

always found that the distance between the first turning level and the $\Omega$-CL is so

large, that the wave has become evanescent ($\phi^1 = 0$) at the time it reaches the $\Omega$-CL.

Furthermore, note that $|R|^2 + (i/A)_\infty (A/l)_\infty |T|^2 = 1$ is also found for other values

of $\alpha$ and $\beta$.

If there are no turning levels and no $\Omega$-CLs in the flow (case 1a), then we have

for the parameters considered always $|R| = 0$.

For a small range of values of $\beta$, we have one turning level and no $\Omega$-CLs (case

2a). In these cases we have $|R| = 1$. An example of this is shown in Figure 5.3 for

\text{hereafter we will omit the superscript}^\dagger \text{ for convenience}
the inertially stable flow ($\omega^2/f^2 > 0$ for all $y/L$). The angle of incidence $\beta = 11^\circ$. The turning level (indicated as ‘TL’) occurs where $Q = 0$. Note that $QN^2/f^2L^2$ becomes singular where $A = 0$ in agreement with (5.9). The wave (Re($\phi$)) is totally reflected ($|R| = 1$) by the turning level. Since there is only one turning level, the wave becomes evanescent at $y = +\infty$, which explains $|T| = 0$.

For $\beta$ larger than approximately $13^\circ$, we have a situation in which there is 1 turning level followed by 1 $\Omega$-CL in the flow (case 3a). In these cases we have $|R| = 1$. An example of such a situation is shown in Figure 5.4. The parameter values are the same as in Figure 5.3, except that $\beta = 17^\circ$. The flow is inertially stable ($\omega^2/f^2 > 0$ for all $y/L$). As the wave propagates towards the background flow, it first encounters the turning level ($Q = 0$) and then the $\Omega$-CL ($\Omega = 0$). Similar as in Figure 5.3, $QN^2/f^2L^2$ becomes singular where $A = 0$, but also where $\Omega = 0$. Note that the wave is again completely reflected. At the turning level the wave becomes evanescent and by the time it reaches the $\Omega$-CL, its amplitude is insignificantly small
Inertially unstable profile in tanh

Angle of incidence $\beta$

Figure 5.5: Locations of the turning levels and $\Omega$-critical levels as a function of the angle of incidence $\beta$. The background flow is inertially unstable.

Inertially unstable profile in tanh

Reflection coefficient $|R|$  
Transmission coefficient $|T|$

Figure 5.6: Reflection and transmission coefficients as a function of the angle of incidence $\beta$. The background flow is inertially unstable. The range of values of $\beta$ where overreflection and overtransmission ($|R|^2 + (l/A)_\infty(A/l)_\infty|T|^2 > 1$) and absorption ($|R|^2 + (l/A)_\infty(A/l)_\infty|T|^2 < 1$) occurs is also indicated.
(Re(ϕ) = 0). As mentioned above, the distance between the turning level and the Ω-CL is so large, that the Ω-CL has no significant effect on the propagation properties of the wave.

A situation in which the Ω-CL is surrounded by two turning levels (case 4a) is reached for β larger than approximately 20°. As can be seen from Figure 5.2, we have then always |R| = 1 and |T| = 0. The wave is, similar as in the case with 1 turning level and 1 Ω-CL, totally reflected. The distance between the (first) turning level and the Ω-CL is in this case also so large that the wave has become evanescent when it reaches the Ω-CL. As a consequence, the wave behaviour is not significantly influenced by the presence of the Ω-CL.

5.4.2 Inertially unstable background flow

The background flow becomes inertially unstable, around the inflection point γ = 0, for Rossby numbers larger than 4.0. Now reflection and transmission coefficients can become larger or smaller than 1, indicating exchange of momentum and energy between the wave and the background flow. In Figure 5.5 the locations of the turning levels and the Ω-CLs are plotted as a function of the angle of incidence β. From this figure, six different cases can be identified:

1b. No turning levels and no Ω-CL
2b. 1 turning level and no Ω-CL
3b. 1 turning level followed by 1 Ω-CL
4b. 1 Ω-CL followed by 1 turning level
5b. No turning levels and 1 Ω-CL
6b. 1 Ω-CL surrounded by 2 turning levels

Similar as in the inertially stable case, turning levels are also present very close to the location of the Ω-CLs. For the same reasons as in the inertially stable case, we will not further discuss these turning levels. Reflection and transmission coefficients are plotted in Figure 5.6 for the same values of β as in Figure 5.5. Note from Figure 5.5 and Figure 5.6 that overreflection and overtransmission occur in case the location of the Ω-CL \( y_c/L > 0 \), and that absorption occurs in case the location of the Ω-CL \( y_c/L < 0 \). In the appendix it is shown that the location of \( y_c/L \), is related to the sign of \(- (d^2U/dy^2)/(dU/dy)\) at the location of the critical level. The sign of \(- (d^2U/dy^2)/(dU/dy)\) at the critical level determines on its turn the sign in the jump of the momentum flux. For example if \( y_c/L > 0 \), then \(- (d^2U/dy^2)/(dU/dy) > 0 \).
Figure 5.7: An example of overreflection with $|R| = 3.304$, $|T| = 0.906$ and $|R|^2 + (l/A)_\infty (A/l)_{-\infty}|T|^2 = 12.689$. The plotted quantities are equal to those in Figure 5.3 and Figure 5.4. The flow is inertially unstable ($\omega f^2 < 0$) between approximately $-1 < y/L < 1$. Singular behaviour in $QN^2/f^2L^2$ occurs at locations where $\Omega = 0$. The angle of incidence $\beta \approx 4.5^\circ$.

Figure 5.8: An example of an inertially unstable background flow, with 1 $\Omega$-critical level and no turning levels. The plotted quantities are equal to those in Figure 5.3 and Figure 5.4. The reflection coefficient $|R|$ is equal to 0.448 and $|T|$ is equal to 0.891 and $|R|^2 + (l/A)_\infty (A/l)_{-\infty}|T|^2 = 0.991$. Singular behaviour in $QN^2/f^2L^2$ occurs at the location where $\Omega = 0$. The angle of incidence $\beta \approx 5.7^\circ$. 
yielding a positive jump in the momentum flux. From Eq. (5.30) and Eq. (5.21) it can then be seen that a positive jump in the momentum flux corresponds to overreflection and overtreatment. A similar reasoning can be followed to explain the occurrence of absorption in case \( \frac{y_c}{L} < 0 \).

Based on the different cases above it can be seen that in case 1b, \(|R|\) is nearly zero and \(|T|\) is approximately between 0.6 and 1.0 and \( |R|^2 + (l/A)_{\infty}(A/l)\infty|T|^2 = 1 \).

In case 2b and 3b we have always \( |R| = 1 \) and \(|T| = 0\) as in the inertially stable case.

Overreflection (\(|R| > 1\)) and overtransmission (\(|T| > 1\)) occurs for \( \beta \) between approximately 4° and 5°. Note, from Figure 5.6 that overreflection and overtreatment is most pronounced in case there is only 1 \( \Omega\)-CL followed by a turning level (case 4b) or in case there is 1 \( \Omega\)-CL only (case 5b). An example in which spectacular overreflection occurs is shown in Figure 5.7, with \( |R| = 3.304 \), \(|T| = 0.906\) and \( |R|^2 + (l/A)_{\infty}(A/l)\infty|T|^2 = 12.689 \). Note that the \( \Omega\)-CL is located in the region where the flow is inertially unstable (\( \omega^2/f^2 < 1 \) for certain values of \( y/L \)).

For \( \beta \) between approximately 5° and 14° absorption with \( |R| < 1 \) and \(|T| < 1\) is dominant (case 5b and 6b). An example of this is shown in Figure 5.8 for \( \beta \approx 5.7\)° with \( |R| = 0.448 \), \(|T| = 0.891\) and \( |R|^2 + (l/A)_{\infty}(A/l)\infty|T|^2 = 0.991 \). There is only 1 \( \Omega\)-CL located in the flow and no turning levels. Similar as in Figure 5.7, the \( \Omega\)-CL is located in the region where the flow is inertially unstable (\( \omega^2/f^2 < 0 \) for certain values of \( y/L \)) and \( QN^2/f^2L^2 \) becomes singular at the \( \Omega\)-CL.

Another example of absorption is shown in Figure 5.9 for \( \beta \approx 10\)°. Here \( |R| = 0.912 \), \(|T| = 0.379\) so \( |R|^2 + (l/A)_{\infty}(A/l)\infty|T|^2 = 0.886 \). Now we have 1 \( \Omega\)-CL that is surrounded by 2 turning levels. The quantity \( QN^2/f^2L^2 \) shows singular behaviour at the \( \Omega\)-CL and where \( A = 0 \) (see relation (5.9)). It is interesting to note that the \( \Omega\)-CL is now located just outside the region where the flow is inertially unstable (\( \omega^2/f^2 < 0 \)).

For values of \( \beta \) larger than approximately 14°, the distance between the first turning level and the \( \Omega\)-CL increases, so we finally recover that \( |R| \approx 1 \) and \(|T| \approx 0\). This is in agreement with the results for the stable background profile. As soon as the distance between the first turning level and the \( \Omega\)-CL becomes too large, the wave has become completely evanescent when it reaches the \( \Omega\)-CL. The \( \Omega\)-CL will then have an insignificant effect on the wave behaviour.

In Figure 5.7, where an example of overreflection is shown, the \( \Omega\)-CL is located in the region where the background flow is inertially unstable. This is also the case for other values of \( \beta \) where overreflection occurs. In case absorption occurs (Figure 5.8 and Figure 5.9), the \( \Omega\)-CL can be located either inside or outside the region where the background flow is inertially unstable. These results are in analogy with the results obtained for horizontally homogeneous background flows with vertical shear.
For example Jones (1968); Acheson (1976); van Duin and Kelder (1982) have shown that unstable shear flows with Richardson numbers smaller than 1/4 are in many cases a necessary condition for overreflection. Absorption occurs in stable flows for Richardson numbers larger than 1/4 (Booker and Bretherton, 1967). Note however, that in our case the actual occurrence of overreflection, overtransmission and absorption is determined by the location of the $\Omega$-CL in the flow, which determines the sign of the jump in the momentum flux (appendix) and by the distance between the first occurring turning level and the $\Omega$-CL.

Other experiments with different values for $\alpha$ and $Ro (> 4.0)$ do not significantly alter the results presented above.
5.5 Conclusions

In this study a linear model is presented describing the propagation of IGWs on an $f$-plane in a vertically homogeneous background flow with shear in the latitudinal direction. The resulting wave equation contains two types of singularities. One corresponds to the occurrence of a critical level in the flow when the intrinsic wave frequency $\Omega$ is equal to zero, here referred to as the $\Omega$-CL. This singularity is similar to the classical Booker and Bretherton-type of singularity for horizontally homogeneous background flows with vertical shear. The other type of singularity is an apparent singularity and has been discussed in the literature in different contexts.

Numerical integration of the singular wave equation enables the calculation of reflection and transmission coefficients for a wide range of parameter values and for two angles of incidence $\alpha = m/l_\infty$ and $\beta = k/l_\infty$.

The main results can be summarized as follows:

- The following distinction can be made for an inertially stable background flow:
  - $|R| = 0$ and $|T| < 1$ in cases when there are no turning levels and no $\Omega$-CLs.
  - $|R| = 1$ and $|T| = 0$ when the wave encounters 1 turning level only.
  - $|R| = 1$ and $|T| = 0$ when the wave first encounters a turning level followed by an $\Omega$-CL.
  - In all cases above $|R|^2 + (l/A)_\infty (A/l)_{-\infty}|T|^2 = 1$.

Also, in all cases considered for the stable background flow where $|R| = 1$ and $|T| = 0$, the distance between the (first) turning level and the $\Omega$-CL played an important role. When the wave passes the turning level it becomes evanescent and its amplitude has become insignificantly small when reaching the $\Omega$-CL. The wave reflects completely at the turning level and the $\Omega$-CL has no significant effect on the wave propagation. This explains the robust result $|R| = 1$ and $|T| = 0$.

- For inertially unstable background flows the following distinction can be made:
  - $|R| = 0$ and $|T| < 1$ when there are no turning levels and no $\Omega$-CLs.
  - $|R| = 1$ and $|T| = 0$ when the wave encounters a turning level only.
  - In the two cases above we have $|R|^2 + (l/A)_\infty (A/l)_{-\infty}|T|^2 = 1$.
  - Overreflection ($|R| > 1$) and overtransmission ($|T| > 1$) occur mostly in case there is only 1 $\Omega$-CL in the flow. In a few cases this $\Omega$-CL is followed by a turning level.
Absorption (|R|, |T| < 1) occurs mostly in case 1 Ω-CL is surrounded by two turning levels. As the distance between the first turning level and the Ω-CL increases for increasing Ω, finally |R| ≈ 1 and |T| ≈ 0.

The sign of $(-d^2U/dy^2)/(dU/dy)$ at the location of the Ω-CL determines the sign of the jump in the momentum flux, that on its turn determines whether overreflection/overtransmission or absorption occurs.

An important finding in this study, is the fact that in inertially stable flows the wave will always totally reflect at the turning level and never reach the critical level. In inertially unstable flows the turning level can disappear and the wave can reach the critical level. At the critical level overreflection, overtransmission and absorption can occur.

In this study we have not accounted for the effects of wave saturation and eventually wave breaking, processes that may significantly contribute to the exchange across the vortex edge. Nevertheless, our results show a wide range of behaviour that may occur when IGWs propagate in shear flows like the polar vortex edge.

5.6 Appendix: Jump in momentum flux across the Ω-critical level

In section 5.2.2 it is shown that whenever Ω becomes zero somewhere between $y = -\infty$ and $y = +\infty$, the momentum flux in Eq. (5.12) can jump in passing critical levels. These jumps are related to logarithmic singularities that are analysed below.

At the critical level $y = y_c$ where $\Omega(y) = 0$, a Taylor expansion of $\Omega(y)$ around $\hat{y} = y - y_c$ can be made i.e.,

$$\Omega(y) = \Omega_1 \hat{y} + \Omega_2 \hat{y}^2 + \ldots$$

where $\Omega_1 = \Omega_{\hat{y}}(0)$ and $\Omega_2 = 1/2\Omega_{\hat{y}\hat{y}}(0)$.

In a similar way the Taylor expansion for $A(y)$ around $\hat{y} = 0$ is given by

$$A(y) = A_1(\hat{y}) + A_2(\hat{y})^2 + \ldots$$

where $A_1 = A_{\hat{y}}(0)$ and $A_2 = 1/2A_{\hat{y}\hat{y}}(0)$. We note that the expansion coefficients $\Omega_1$, $\Omega_2$, ..., $A_1$, $A_2$, ... are all related to the Taylor expansion coefficients of the background flow $U(y)$.

Application of the method of Frobenius to Eq. (5.6) and using of (5.23) and (5.24) yields Frobenius exponents 0 and 1. The first solution reads

$$\hat{\phi}_{\Omega_1}(\hat{y}) = \hat{y} \sum_{n=0}^{\infty} c_n \hat{y}^n, \quad c_0 \neq 0, \quad \text{where} \quad \hat{y} = y - y_c,$$

$$\text{(5.25)}$$
and a second linearly independent solution is given by

$$
\hat{\phi}_{\Omega_2}(\bar{y}) = \hat{\phi}_{\Omega_1}(\bar{y})(B_{\Omega_1} \ln \bar{y} + B_{\Omega_2} \bar{y}^{-1}) + \hat{\phi}_{\Omega_1}(\bar{y}) \sum_{n=2}^{\infty} \frac{d_n}{n-1} \bar{y}^{n-1},
$$

(5.26)

where $B_{\Omega_1}$ and $B_{\Omega_2}$ are constants defined as

$$
B_{\Omega_1} = \frac{A_1}{a_0^2} - \frac{2a_1}{a_0^3} A_0, \quad B_{\Omega_2} = -\frac{A_0}{a_0^2}.
$$

(5.27)

The solution can again be written as a linear combination of (5.25) and (5.26) i.e.

$$
\hat{\phi}_\Omega(\bar{y}) = F_\Omega \hat{\phi}_{\Omega_1}(\bar{y}) + G_\Omega \hat{\phi}_{\Omega_2}(\bar{y}).
$$

(5.28)

Recalling (5.11), the jump across $\Omega$-critical levels can now be determined from

$$
\Delta \text{Re} \left( \frac{i \hat{\phi}^* d \hat{\phi}}{A} \frac{d}{dy} \right) = \Delta \text{Re} \left( \frac{i}{A} \left[ F_{\Omega_1}^2 \hat{\phi}_{\Omega_1} \frac{d \hat{\phi}_{\Omega_1}}{dy} + G_{\Omega_1}^2 \hat{\phi}_{\Omega_2} \frac{d \hat{\phi}_{\Omega_2}}{dy} + F_{\Omega_2}^2 \hat{\phi}_{\Omega_1} \frac{d \hat{\phi}_{\Omega_2}}{dy} + F_{\Omega_2}^2 \hat{\phi}_{\Omega_2} \frac{d \hat{\phi}_{\Omega_1}}{dy} \right] \right).
$$

(5.29)

All terms in (5.29) give zero contributions at $\bar{y} = 0$, except the term proportional to $|G_{\Omega_1}^2$. After some algebra (5.29) takes the form

$$
C_{+\infty} - C_{-\infty} \equiv \Delta \text{Re} \left( \frac{i \hat{\phi}^* d \hat{\phi}}{A} \frac{d}{dy} \right) = |G_{\Omega_1}^2| \Delta \text{Re}(i J_\Omega \ln \bar{y}),
$$

(5.30)

where $J_\Omega$ is given by

$$
J_\Omega = \frac{2a_1 |A_0|^2 - a_0 A_1 A_0^*}{a_0 |A_0|^2 |a_0|^2}.
$$

(5.31)

Evaluating $A_0$, $A_1$, $a_0$ and $a_1$ in terms of the Taylor expansion coefficients of the background flow $U(y)$, reveals that the sign of $J_\Omega$ is determined by the sign of $-(d^2U/dy^2)/(dU/dy)$ at the location of the critical level $y = y_c$. From (5.12) the momentum flux can be written as

$$
\Delta \tau = \rho_0 k \left| G_{\Omega_1}^2 \right| < \Delta \text{Re}(i J_\Omega \ln \bar{y}) >,
$$

(5.32)

for real $k$. Obviously the sign of the jump is determined by the sign of $J_\Omega$. For the hyperbolic tangent profile this means that we have a positive jump in the momentum flux for $y_c > 0$ and a negative jump for $y_c < 0$. 


Chapter 6

Summary, conclusions, and outlook

6.1 Summary and conclusions

In this section the main results of the investigations described in this thesis will be briefly summarized. An extensive summary and discussion of the results obtained, is given at the end of each of the chapters 2, 3 and 5. The main objective of the work presented in this thesis is to investigate the role of inertia-gravity waves upon the permeability of the Antarctic polar vortex. A number of questions have been posed that serve as a guideline for these investigations:

1. How well is the polar vortex interior isolated from its surroundings?
2. How much air leaks from the Antarctic polar vortex horizontally across the edge and vertically through the vortex bottom?
3. Is there a seasonal and interannual variability in the amount of exchange from the polar vortex?
4. What is the effect of inertia-gravity waves upon the leakage of air from the polar vortex and how does the leakage depend on the wave parameters?
5. How do inertia-gravity waves propagate in idealized background flows?

The first three questions are addressed in chapter 2, question four in chapter 3 and question five in chapter 5.
In chapter 2 it is found that the Antarctic polar vortex edge is a very effective barrier to transport. Earlier studies using different analysis methods, have already found in a more qualitative sense that the vortex interior is well isolated from its surrounding midlatitude air. The number of studies that have estimated the amount of exchange from the polar vortex in a quantitative sense is rather sparse. The study in chapter 2 is one of the few accounts using trajectories, in which quantitative exchange rates have been calculated from high-resolution \( (1^\circ \times 1^\circ) \) meteorological data. Previous studies have used rather coarse-gridded \( (5^\circ \times 3.75^\circ) \) meteorological data. Better horizontal and vertical resolution of these data generally contributes to a more accurate estimation of the amount of exchange across the vortex edge.

Furthermore, it is shown that there is little transport of air from the vortex interior to midlatitudes. However, the amount of air that enters the vortex interior from midlatitudes is even less. This rather peculiar fact implies that the outer vortex edge is more robust to air mass intrusions from midlatitudes into the polar vortex, than the inner vortex edge is to air from the vortex interior to midlatitudes. This so-called ‘one-sidedness’ of air mass intrusions is important for the following reasons. A relatively larger permeability of the inner vortex edge, implies that ozone-poor air from the vortex interior (ozone hole) can more easily escape towards midlatitudes than warm ozone-rich air is able to enter the vortex interior. It is clear that the greater permeability of the inner edge has a negative impact on the midlatitude ozone content.

Large-scale breaking planetary waves are virtually absent in the months August and September in the neighborhood of the vortex edge. In these months the winds in the polar vortex reach maximum speeds, thereby shifting the critical ‘wave breaking’ lines toward midlatitudes. The absence of breaking planetary waves is reflected in the very low exchange rates in August and September. At the end of the polar night in October and November, the wind velocities in the polar vortex decrease somewhat. In these months planetary waves may break in the neighborhood of the polar vortex edge, creating filaments that are wrapped around the vortex. These filaments contain vortex air and may eventually break off the main vortex. It is shown that the specific filamentation events in October 1998 are responsible for the relatively large exchange rates in that month. Generally, the polar vortex edge is very resistant to transport even in the presence of large-scale (breaking) planetary waves. Only in periods of intense filamentation is its near impermeability affected.

This question has also been investigated in chapter 2. It is shown that the amount of exchange from the polar vortex depends on whether air parcel trajectories are calculated in the ‘isentropic’ mode or in the ‘3D-mode’. In case
the air parcel trajectories are calculated in the isentropic mode, only cross-edge leakage rates can be determined. In the late winter months August and September 1998, cross-edge maximum leakage rates of 0.37%/week are found. In October and November 1998, filamentation of the outer vortex edge leads to somewhat higher rates of 1.95%/week or less. The calculation of isentropic trajectories is inherently connected to the neglect of diabatic heating or cooling effects. It is shown however, that diabatic motions become already important at short time scales. Trajectory calculations for the polar vortex are usually performed for periods of one month or longer. The neglect of cross-isentropic air motions might therefore lead to an under-estimation of the leakage of air from the polar vortex. A comparison between the isentropic and 3D cross-edge exchange rates show, that the 3D rates are significantly larger than the isentropic ones.

The question of how much air leaked through the bottom of the polar vortex at 340 K, can only be answered from the results obtained from the 3D-trajectory calculations. Typical vertical exchange rates at 450 K are of the order 0.03%/week, which is much lower than typical chemical loss rates for ozone (1-2%/day). The calculated average descent rates are in good agreement with other values reported in the literature. It is also found that the descent rates increase with increasing potential temperature (K).

The calculated exchange rates in chapter 2 depend of course to an important degree on the analysis method used. The question on how to define correctly the polar vortex edge is still difficult and controversial. Nevertheless, the method described in section 2.3.1 seems a good choice. Although the calculated exchange rates are low, they are still significant. Mass or area fluxes have been determined at regular time intervals, such that an increasing amount of air leaves or enters the polar vortex as time progresses. Furthermore, the threshold residence time defined in section 2.3.2 of chapter 2 has been introduced to prevent frequent reversible parcel displacements. The exchange rates in chapter 2 are therefore quite robust estimates.

The results in chapter 2 show that exchange rates are small in the late winter months August and September and increase in the spring months October and November. Maximum rates are found in October. A clear trend in the amount of exchange on a seasonal time scale has not been observed. However, the calculated exchange rates for the year 1993 are smaller than those for 1998. This is in agreement with the findings of other authors. Exchange rates and diabatic cooling rates are generally larger in even years (1992, 1994, ...) than in odd years (1993, 1995, ...). This variability in the exchange rates is thought
Summary, conclusions, and outlook

to be related to variations in planetary wave activity and to periodic forcings like the Quasi-Biennial Oscillation (QBO).

The effect of inertia-gravity waves upon the leakage of air from the polar vortex is investigated using the KNMI trajectory model. 3D-trajectories are calculated using high-resolution (1° x 1°) meteorological data. The effect of an idealized inertia-gravity wave field has been investigated in a late winter situation (August), when there is no significant exchange across the vortex edge due to large-scale processes. All exchange that will then occur, will be caused by the waves. The wave field has been determined analytically and is superimposed upon the background wind of the ECMWF. For different values of the horizontal and vertical wavelength and amplitude of the wave, exchange rates are determined. The analysis method is identical to the method discussed in chapter 2. Since the vortex edge has proven to be an effective barrier to transport, the wave field is superimposed in the vortex edge region in order to investigate whether inertia-gravity waves can induce cross-edge transport. The results show that inertia-gravity waves induce only little transport from the polar vortex, cross-edge as well as cross-bottom. Both cross-edge and cross-bottom exchange rates increase for decreasing horizontal and vertical wavelength. This is in agreement with the fact that the shortest waves have the largest wave frequencies and wave amplitudes. This clearly increases the chance for exchange from the polar vortex. Increasing the wave amplitude and keeping the horizontal and vertical wavelength constant, also leads to an increase in the amount of exchange. Most of the exchange occurs in the lower polar vortex. It is shown that (air) parcels moving in a finite amplitude transverse wave will experience a drift depending on the ratio between the wave amplitude and the phase velocity. This explains to some degree the poleward shift of air parcels in the presence of inertia-gravity waves. This poleward shift increases as the local horizontal angle of incidence with respect to the zonal background flow increases. As air parcels move poleward under the influence of the wave field, they will experience a larger vertical descent near the vortex core than near the edge of the vortex. The larger descent of air parcels in the presence of inertia-gravity waves than in the absence of these waves, is reflected in the average diabatic cooling rates. Average diabatic cooling rates increase for decreasing horizontal wavelength. This can be explained from the fact that the shortest waves have the largest horizontal and vertical wave amplitudes.

This question involved a detailed investigation of the interaction between inertia-gravity waves and a horizontally sheared barotropic background flow. From the inviscid linearized Navier-Stokes equations in a hydrostatic atmo-
Outlook

The problem concerning the permeability of the Antarctic polar vortex has many aspects. One aspect, which is the main topic of this thesis is to investigate the role of inertia-gravity waves upon the permeability of the Antarctic polar vortex. This topic in itself may be approached in several ways. In this thesis the topic is approached in two different ways, i.e. by calculating air parcel trajectories in a trajectory model and by using an analytical/numerical model with which the interaction between inertia-
gravity waves and the polar vortex has been investigated. Although we have come
to a better understanding of the degree of permeability of the Antarctic polar vortex
and the role of inertia-gravity waves, new issues have popped up that need further
investigation.

Potential vorticity (PV) has become a primary diagnostic tool for determining the
location of the polar vortex edge. However, exact determination of the vortex edge is
difficult. PV is a derived quantity, that is calculated from the wind and temperature
data of sophisticated weather forecast or climate models. These data have limited
space and time resolution, which limits the accuracy of the PV field. A better hor-
izontal and vertical resolution of the meteorological data will directly improve the
determination of the vortex edge. Uncertainties and errors in the determination of
PV and consequently the vortex edge, also have important implications for the cal-
culation of exchange from the polar vortex. In general, the use of coarse-gridded
data often leads to an overestimation in the amount of exchange. Due to the wide
range of techniques, methods and data used for calculating exchange from the polar
vortex, it is difficult to compare or validate different outcomes from numerical mod-
els. An additional problem arises from the fact that it is difficult to estimate directly
from observations how large the actual amount of exchange from the polar vortex is.
Large measurements campaigns like STRATEOLE should contribute to fill the
gap between model outcome and observations. Within the STRATEOLE/VORCORE
project, 25 balloons will be launched from McMurdo station during September and
October 2004. This experiment will provide important chemical and dynamical in-
formation about the Antarctic polar vortex core at the end of winter, when it is well
isolated, up to its final breakdown in November.

It would be interesting to perform trajectory calculations with more vertical lev-
els, including the upper stratosphere and lower mesosphere. Such calculations would
contribute to a better understanding of the total amount of air that is flowing through
and processed by the Antarctic polar vortex.

Weather forecast and climate models do not explicitly account for the effects of
small-scale inertia-gravity waves. Lagrangian type models are very well suited for
incorporating the sub-grid scale effects of such waves. Simple plane-parallel wave
data can only be implemented locally, otherwise the polarization relations would have
to be derived for inertia-gravity waves on a sphere (Francis, 1972). This inevitably
complicates the problem a lot.

A useful extension of the study performed in chapter 3 would be the development
of a Lagrangian type model in which the vertical and horizontal structure of the polar
vortex and the wave field is prescribed in an idealized way. Such a model would
enable a detailed study of the movement of air parcels due to the combination of
the polar vortex wind field and the inertia-gravity wave field. Additional laboratory
Outlook

experiments in fluid containers using dye could be helpful to gain further insight in the basic processes that play a role between an idealized vortex and inertia-gravity waves. Such experiments might also give a qualitative picture of mixing that results from the interaction between the vortex and the wave.

Very few studies have actually addressed the interaction between inertia-gravity waves and horizontally sheared barotropic background flows. Analytical models, even in the most simplified configurations, already lead to complex singular wave equations that in most cases can only be solved numerically. A clear drawback of this often inevitable numerical approach, is the absence of analytical benchmarks. These benchmarks are in the first place needed to demonstrate the ability of the numerical code to integrate across delicate regions where singularities occur. Furthermore, models for which analytical solutions can be found, provide a detailed picture of the behaviour of waves around critical levels. The model that is considered in chapter 5 is linear. Despite this linearity, the interaction between the wave and the mean flow (vortex edge) at the critical levels might give rise to irreversible parcel displacements and thus transport across the vortex edge. Nevertheless, it is recommended to consider in future models also non-linear effects. Especially, the interaction between breaking inertia-gravity waves and the vortex edge is of importance. Due to the decrease in density with height, wave amplitudes increase. At a certain level the amplitude becomes larger than a certain threshold for instability. Eventually the wave will break and induce transport and mixing of air across the polar vortex edge. This approach involves the use of more sophisticated numerical models that are able to integrate the full Navier-Stokes equations.

More in general, the causes of midlatitude lower stratospheric ozone trends have not yet been completely resolved, although the observed downward trend appears to be due to a combination of local chemical loss, enhanced by volcanic sulfuric aerosols. The temporal and spatial variations in the decline of midlatitude ozone is due to halogens, like chlorine, which can be released by UV radiation from for example, CFCs. Transport of ozone-poor air to midlatitudes following the break-up of the Antarctic polar vortex may contribute to the downward trend in ozone not only locally and for short periods of time. The transport and mixing of ozone-poor air with midlatitude air is much faster than the slow chemical replacement time of ozone (months in the lower stratosphere). This may lead to a long term contribution of polar ozone loss to the downward midlatitude ozone trend (WMO-UNEP, scientific assessment of ozone depletion 1998/2002).

Only recently, it has been shown that an increased amount of greenhouse gases brought into the troposphere, can lead to cooling of the lower stratosphere (Shindell et al., 1998). Also, changes in temperature and wind induced by greenhouse gases can alter the propagation of planetary waves in such a way that they no longer disturb
the Arctic vortex as often. The cooling of the stratosphere and the decreased planetary wave activity lead to the formation of a strong and isolated Arctic vortex. Due to the strong cooling in the vortex core, polar stratospheric clouds on which ozone is rapidly depleted, can more easily form. However, Arctic ozone depletion is highly variable from year to year and difficult to predict. Current chemistry-climate models have shown that it is unlikely that an Arctic ozone hole will develop similar to the Antarctic one. These same models predict that ozone levels in the Antarctic will gradually increase after 2010 to values like those before 1980, due to the expected decreases of halogens in the atmosphere. This will lead over the next 50 years or so to a gradual recovery of the ozone layer in general (WMO-UNEP, scientific assessment of ozone depletion 2002).
Bibliography


— 1996, Rossby wave speeds and mixing barriers in the stratosphere. Part I: Obser-

Bowman, K. P. and P. Chen: 1994, Mixing by barotropic instability in a nonlinear

Boyd, J. P.: 1976, *Planetary waves and the semiannual wind oscillation in the tropi-

— 1978, The effects of latitudinal shear on equatorial waves. Part 1: Theory and


Brewer, A. W.: 1949, Evidence for a world circulation provided by the measurements


20,563–20,571.


Chubachi, S.: 1984, Preliminary results of ozone observations at Syowa Station from

Dalhberg, S. P. and K. P. Bowman: 1994, Climatology of large-scale isentropic mixing
in the Arctic winter stratosphere from analyzed winds. *J. Geophys. Res.*, **99**,
20,585–20,599.

Danielsen et al., E. F.: 1991, Irreversible transport in the stratosphere by internal

Deshler, T., D. J. Hofmann, and J. V. Hereford: 1990, Ozone profile measurements
within, at the edge of, and outside the Antarctic polar vortex in the spring of 1988.

Dobson, G. M. B.: 1956, Origin and distribution of the polyatomic molecules in the
Bibliography


Samenvatting

Tijdens de wintermaanden juni tot en met september op het Zuidelijk Halfrond bestaat er een groot temperatuurverschil tussen de equatoriale en de poolatmosfeer. Als gevolg hiervan vormt zich ieder jaar in de lagere stratosfeer (13-30 km) op ongeveer 60° Zuid een straalstroom die ook wel de Antarctische circumpolaire straalstroom wordt genoemd. De Antarctische circumpolaire straalstroom wordt gekenmerkt door zeer hoge windsnelheden tot 300 km/uur. In de literatuur wordt hiervoor ook wel de benaming Antarctische polaire wervelrand gebruikt. Deze is gebaseerd op de ruimtelijke verdeling van de potentiële vorticiteit (PV)\(^1\). De wervelrand wordt gekarakteriseerd door een nauwe circumpolaire zone waarin een scherpe verandering in PV optreedt. Het geheel bestaande uit de Antarctische polaire wervelrand en de lucht die aan de poolwaardse zijde opgesloten zit, wordt over het algemeen aangeduid als de Antarctische polaire wervel. De lucht die in de Antarctische polaire wervel gevangen zit, kan in de poolwinter sterk afkoelen tot temperaturen rond -88 °Celsius. Bij deze temperaturen gaan verschillende chemische processen een rol spelen, die in het voorjaar (september-november) zorgen voor een snelle afbraak van ozon en de vorming van het bekende ozongat.

Sinds de ontdekking van het ozongat omtrent 1985 is de doorlaatbaarheid van de fysieke begrenzing van de wervel voor transport van ozon-arme lucht in de belangstelling komen te staan. Afhankelijk van de mate van doorlaatbaarheid zou die ozonarme lucht uit de polaire wervel kunnen ontsnappen en de waargenomen neerwaartse trend in ozonwaarden op gematigde breedten kunnen verklaren. De doorlaatbaarheid van de wervelrand kan worden beïnvloed door in de atmosfeer aanwezige golven. Uit de interactie tussen de golven en de wervel kan transport en menging van lucht uit het ozongat naar gematigde breedten optreden en omgekeerd.

De afgelopen 20 jaar is het effect van planetaire golven met golflengten van duizenden kilometers, op de doorlaatbaarheid van de wervelrand veelvuldig en met

\(^1\)PV zegt iets over hoeveelheid draaiing die een stabiel gelaagde kolom lucht heeft onafhankelijkheid of deze wordt samengedrukt of uitgerekt. PV is een grootheid die over een periode van weken behouden is in de stratosfeer en daarom een nuttig hulpmiddel is om de beweging van luchtdoeltjes in de tijd en ruimte te volgen.
verschillende methoden onderzocht. Studies hebben aangetoond dat planetaire golven die van gematigde breedten omhoog propageren uit de troposfeer, kunnen breken in de nabijheid van de wervelrand. Tijdens dit proces kan erosie van de buitenste wervelrand optreden, wat zich manifesteert in langgerekte slierten lucht ofwel filamenten, die de wervel omcirkelen. Uiteindelijk kunnen deze filamenten van de polaire wervel losscheuren en voor anomalie ozonwaarden op gematigde breedten zorgen. Deze processen spelen met name een rol tijdens het opbreken van de polaire wervel in de maanden november en december op het Zuidelijk Halfrond. Het is gebleken dat eerder in het seizoen, van juli tot en met oktober, filamentatie van de wervelrand veel minder vaak optreedt en er praktisch geen uitwisseling van lucht door de wervelrand optreedt.

Een paar studies hebben gesuggereerd dat in deze periode wel filamentatie op kleinere schaal kan optreden als gevolg van zogenaamde inertiaal-zwaartegolven. Deze golven kunnen bestaan dankzij de gelaagdheid van de lucht en hebben golflengten die groot genoeg zijn om de rotatie van de aarde te ‘voelen’. Het potentiële belang van inertiaal-zwaartegolven op de doorlaatbaarheid van de polaire wervelrand in de maanden augustus tot en met oktober wordt onderstreept door het feit dat juist in deze periode de laagste ozonwaarden in het ozongat bereikt worden. Echter, het effect van inertiaal-zwaartegolven op de uitwisseling van lucht door de Antarctische polaire wervelrand was nog niet in detail onderzocht. Het is de doelstelling van het onderzoek beschreven in dit proefschrift om dit nader te onderzoeken. Het probleem is op twee manieren benaderd.

In de eerste plaats is de doorlaatbaarheid van de wervelrand onderzocht met behulp van een trajectoriemodel. Zo’n model kan gebruikmakend van de wind- en temperatuurvelden van het Europees weercentrum in Reading (ECMWF), de verplaatsing van luchtdoeltjes in de tijd berekenen. De wind- en temperatuurvelden waren beschikbaar op een rooster van 1° × 1° (ruwweg 100 x 100 km in lengte x breedte). Zo’n hoge resolutie was niet eerder gebruikt in studies met trajectoriën, die de doorlaatbaarheid van de Antarctische wervelrand tot onderwerp hadden. De experimenten die wij met zulke modellen hebben uitgevoerd en de verkregen resultaten zijn beschreven in de hoofdstukken 2 en 3 van dit proefschrift. De doelstelling van de methode beschreven in hoofdstuk 2, is om met behulp van wind- en temperatuurvelden van hoge resolutie tot een betere schatting van de lekkagesnelheden (als fractie van de initiële massa ingenomen door de trajectoriën in de wervel, en uitgedrukt in %/week) te komen. De lekkage wordt zowel horizontaal door de wervelrand als aan de onderkant van de polaire wervel bepaald. De resultaten geven een nieuw kwantitatief beeld van de uitwisseling van lucht door de polaire wervelrand. Ze versterken de conclusies uit eerdere studies dat de wervelrand nagenoeg ondoorlaatbaar is voor grootschalige mengprocessen. Het blijkt dat vertikale bewegingen in de po-
Samenvatting

slaire wervel van groot belang zijn om tot juiste schattingen te komen van de mate van lekkage uit de polaire wervel. De variabiliteit van jaar tot jaar van de mate van uitwisseling door de wervelrand is verassend groot.

De nieuwe schattingen van de mate van lekkage ten gevolge van grootschalige golven worden beschreven in hoofdstuk 2. Deze waren mede nodig om tot nauwkeurige schattingen van het effect van inertiaal-zwaartegolven op de doorlaatbaarheid van de polaire wervelrand te komen. In hoofdstuk 3 wordt dit effect onderzocht. Het effect van een expliciet inertiaal-zwaartegolfveld wordt onderzocht in een maand waarin geen aantoonbare lekkage door grootschalige mengprocessen optreedt. De snelheidscomponenten van het inertiaal-zwaartegolfveld zijn analytisch bepaald en sterk geïdealiseerd en worden gesuperponeerd op het achtergrondwindveld uit het ECMWF model. Alle lekkage die vervolgens optreedt is dus het gevolg van het gesuperponeerde golfveld. Voor verschillende waarden van de horizontale en vertikale golflengte en amplitude van de golf zijn lekkages gevonden van 1.06%/maand of minder. De meeste lekkage treedt op in de onderste lagen van de polaire wervel. Inertiaal-zwaartegolven hebben slechts een geringe invloed op de uitwisseling van lucht uit de polaire wervel, waarbij de vereenvoudigde weergave van het golfveld niet uit het oog verloren mag worden.

Vervolgens wordt in hoofdstuk 4 een compilatie van een aantal eenvoudige analytische lineaire golfmodellen gepresenteerd. De modellen zijn eenvoudig in die zin dat de fasevlakken van de golf loodrecht staan op de stromingsrichting van de achtergrondstroming. Voor de achtergrondstroming is een aantal zeer geïdealiseerde profielen bekeken. Deze modellen waren feitelijk voorstudies voor het werk dat in hoofdstuk 5 is beschreven.

Als alternatief voor de methode met behulp van trajectoriën, is in hoofdstuk 5 de interactie tussen inertiaal-zwaartegolven en de polaire wervelrand in een analytisch-numeriek model onderzocht. Uit de gelineariseerde Navier-Stokes vergelijkingen voor een niet-visceuze hydrostatische atmosfeer is een homogene singuliere golfvergelijking afgeleid. Deze golfvergelijking is vervolgens numeriek opgelost. De oplossingen van de golfvergelijking beschrijven de propagatie van een inertiaal-zwaartegolf in een achtergrondstroming. Voor de achtergrondstroming werd een windprofiel met horizontale schering gekozen, dat representief is voor de polaire wervelrand. De singulariteit van de golfvergelijking beschrijft locaties in de achtergrondstroming, zogeheten kritieke lagen, waar de golf bijzonder gedrag vertoont in de vorm van energieoverdracht van de golf naar de achtergrondstroming of omgekeerd. De locatie van een kritieke laag in de achtergrondstroming speelt een grote rol in het voortplantingsgedrag van de golf. Ook de aanwezigheid en locatie van zogenaamde omkeerlagen ten opzichte van de kritieke lagen speelt hierin een belangrijke
Rol.

In hoofdstuk 6 worden tenslotte de conclusies van het onderzoek samengevat.
Dankwoord

Van de mensen die een rol hebben gespeeld in de totstandkoming van dit proefschrift, wil ik graag een aantal met naam noemen. In de eerste plaats wil ik mijn promotoren Hennie Kelder en Frans Sluijter bedanken voor hun begeleiding in het algemeen. Het geheel en kritisch doorlezen van het manuscript door Frans Sluijter heeft het proefschrift op met name taalkundig gebied sterk verbeterd.

Daarnaast wil ik natuurlijk mijn copromotor Leon Kamp en Peter van Velthoven bedanken. Peter wil ik in het bijzonder bedanken voor zijn intensieve begeleiding, het nauwkeurig doorlezen van de manuscripten en het feit dat hij ondanks zijn eigen drukke agenda toch altijd ruim de tijd voor je nam om over wetenschappelijke maar ook over meer persoonlijke zaken te praten. Je steun en opbeurende woorden, met name in perioden dat het wat minder vlotte met het onderzoek, zal ik niet snel vergeten. Peter, hartelijk dank hiervoor!

Leon wil ik met name bedanken voor zijn raad en daad op mathematisch gebied. Je wist moeilijke concepten altijd op een eenvoudige heldere manier uit te leggen. Ik heb hier veel van geleerd.

Ook de leden van de leescommissie mag ik in dit dankwoord niet vergeten, voor het nauwkeurig doorlezen en becommentariëren van het concept-proefschrift.

Zonder de grote bijdrage van Francois Lott had hoofdstuk 5 niet tot stand kunnen komen. Francois, thanks a lot for your invaluable contribution to chapter 5.

De Nederlandse Organisatie voor wetenschappelijk Onderzoek (NWO) en met name mevrouw Kooi-De Bruijne hebben het mogelijk gemaakt dat ik na afloop van mijn contract nog 3 maanden betaald kon doorwerken. Hartelijk dank hiervoor.

Tijdens de reguliere werkbezoeken aan het KNMI, heb ik de plezierige informele sfeer binnen de sectie atmosferische samenstelling altijd als zeer prettig ervaren. Rinus Scheele was de vraagbaak als het ging om problemen met het trajectriemodel. Zijn grote hulpvaardigheid kwam de snelheid van het onderzoek vaak ten goede.

Michael Sigmond van het KNMI hield mij op de hoogte van alle nieuwtjes op het KNMI en is altijd een plezierige gesprekspartner geweest. Wie weet zien we ons in de nabije toekomst wel weer als collega’s ...
Ook de collega’s van het (voormalige) samenwerkingsverband theoretische natuurkunde aan de Technische Universiteit Eindhoven, mag ik niet vergeten te bedanken voor de gezellige jaren.

De komst van David Molenaar als AIO heeft de laatste twee jaar van mijn onderzoek bijzonder plezierig gemaakt. Tijdens onze dagelijkse ‘bakkies’ bespraken we, hetzij in het Noors of in het Nederlands, het wel en wee in de wereld en onze klim-en trekkingavonturen. David, ik zal deze ‘bakkies’ enorm gaan missen, maar ik hoop dat we in de toekomst gewoon verder blijven ouwehoeren hetzij hier in Nederland, in Noorwegen of ergens in de bergen.

De gezellige etentjes in Utrecht met Wouter Peters, Julius Spit en de paranimfen Garret Alberts en Koen Hos waren een welkome afwisseling na een dag hard werken bij het KNMI. Koen, ik hoop dat we in toekomst nog meer avontuurlijke trektochten zullen gaan maken, zoals we dat afgelopen oktober in Noorwegen gedaan hebben.

Mijn ouders, broer en alle andere goede vrienden of kennissen wil ik bedanken voor hun goede zorgen, steun en interesse.

Tot slot is er natuurlijk Angela en onze Maremma berghond Miska. Door hen heb ik steeds weer de motivatie op kunnen brengen om door te gaan. Door Angela’s steun heb ik de bergen en dalen van het ‘AIO-schap’ weten te overwinnen en Miska zorgde tijdens de overwerktijdjes al kauwend en knagend onder het bureau voor de ‘wijze tips’.

Dankwoord

I was born on October 2, 1973 in Schaesberg. From 1986 to 1993 I followed pre-university education at the Bernardinus college in Heerlen. In 1993 I started an undergraduate study in physics and astronomy and graduated in 1998 in meteorology and physical oceanography at the Institute of Marine and Atmospheric Research of Utrecht University. During my graduation project I stayed three months at the Max Planck Institute in Hamburg, Germany. The project concerned the longterm variability of the North-Atlantic thermohaline ocean circulation under the supervision of Dr.Leo Maas and Dr.Gerrit Lohmann. From October 1998 to January 2003, I carried out my PhD research at Eindhoven University of Technology and the Royal Dutch Meteorological Institute. Besides science, I was involved in activities concerning mountaineering, the norwegian language and my Maremma mountain dog.