ministerie van verkeer en waterstaat

koninklijk nederlands meteorologisch instituut

on waves in the upper atmosphere

H. Kelder

scientific reports WR 86-8

wetenschappelijke rapporten WR 86-8

KNMI

S T E L L I N G E N

behorende bij het
proefschrift van
H. Kelder

Eindhoven, 27 januari 1987

S T E L L I N G E N

I Bij de berekening van de voortplanting van zwaartegolven door hoogte-afhankelijke windvelden is het van belang te weten of het een monotoon toenemende wind is, dan wel een wind die na een zeker maximum weer afneemt.
(Dit proefschrift.)

II De invloed van de rotatie van de aarde op de passage van een zwaartegolf door een hoogte-afhankelijk windveld is niet altijd te verwaarlozen.
(Dit proefschrift.)

III De Boussinesq benadering leidt voor sommige golfvoortplantingsproblemen in de atmosfeer tot resultaten die sterk afwijken van die welke met andere benaderingen worden verkregen.
(Dit proefschrift.)

IV Slechts combinaties van ionosferische meetgegevens kunnen karakteristieken van zich voortplantende verstoringen in de ionosfeer onthullen.
(Dit proefschrift.)

V Resonante overreflectie treedt alleen maar op als de kritieke laag in een symmetrisch windveld zich in het vlak van symmetrie bevindt.
(Teitelbaum, H., Vial, F. & Kelder, H., 1986.
Submitted to Advances in Space Physics.)

VI De reflectie en transmissie van een zwaartegolf aan een zogenaamd kritieke laag verandert niet wezenlijk als aan de vloeistof in beperkte, doch voor de atmosfeer in realistische mate, moleculaire viscositeit en het vermogen tot warmtegeleiding wordt toegekend.
(Van Duin, C.A. & Kelder, H., 1986.
J. Fluid Mech. 169, 293-306.)

VII Dat Bougeret met zijn oost-west opstelling van radiotelescopen tot de conclusie komt, dat ionosfeerverstoringen zich oost-west bewegen, is wellicht niet verwonderlijk maar wel onjuist.
(Bougeret, J.L., 1981, Astron. Astrophys. 96, 259-266.)

VIII Evans & Wand beweren in de ionosfeer een nieuw soort zich voortplantende verstoring gevonden te hebben, die zich alleen manifesteert bij satellietwaarnemingen onder hoeken van kleine elevatie. Dit verschijnsel is echter met grote waarschijnlijkheid toe te schrijven aan interferentie tussen de directe straal en de stralen die een of meer reflecties aan het aardoppervlak hebben ondergaan.
(Evans, J.V. & Wand, R.H., 1983, J. atm. terr. Phys. 45, 255-265.
Sukumar, S., Spoelstra, T.A.Th. & Kelder, H., 1986. Accepted for publication by J. atm. terr. Phys.)

IX Aan het pionierswerk van Elias over de verklaring van het ontstaan van de ionosfeer wordt ten onrechte in de literatuur nauwelijks aandacht besteed. De Chapman-theorie voor de vorming van ionosfeerlagen dient eigenlijk Elias-theorie of tenminste Chapman-Elias-theorie te worden genoemd.
(Elias, G.J., 1925, Elektrische Nachrichten, 2, 351-358.)

X De conclusie van Lupini & Pellacani dat niet-lineariteit en diffusie het gedrag van zwaartegolven in een jet-type windveld domineren is onvoldoende gefundeerd.
(Lupini, R. & Pellacani, C., 1975, Il Nuovo Cimento 29B, no. 1, p. 47.)

XI Precieze kennis van de ionosfeer is voor satellietgeodetische waarnemingen van belang voor een nauwkeurige interpretatie van hun meetgegevens.
(Souriau, A., Piuzzi, A., Etchegorry, M., Machitel, P., 1984. Bull. Géod. 58, 53-72.)

XII De regel bij de C.N.R.S.* in Frankrijk om directeuren van research laboratoria ten hoogste drie keer 4 jaar aan het bewind te laten, verdient ook in Nederland aanbeveling.

XIII Een van de speciaal op de overheid gerichte aanbevelingen in het rapport 'Aardwetenschappen, een knooppunt in onze samenleving' luidt: "bevorder dat specifiek Nederlandse aardkundige kennis via ontwikkelingssamenwerking ten nutte komt aan arme landen". Voor wat betreft het aardmagnetisme en de aëronomie moeten wij echter hopen dat er arme landen zullen zijn die Nederland in de toekomst van kennis ter zake zullen willen voorzien.
(Aardwetenschappen, een knooppunt in onze samenleving, 1985.)

* Centre National de la Recherche Scientifique

C O N T E N T S

VOORWOORD i

1. INTRODUCTION 1
   1.1. Atmospheric Structure 1
   1.2. Tides 3
   1.3. Internal Gravity Waves 6
   1.4. Ionospheric Observations 9
   1.5. References 11

2. ELECTROMAGNETIC WAVES OF HIGH FREQUENCY IN THE IONOSPHERE 13
   2.1. Introduction 13
   2.2. The refractive index for the propagation of electromagnetic waves through the ionosphere 13
   2.3. The temperate plasma correction to high frequency waves 19
      2.3.1. The cold plasma case 21
      2.3.2. The temperate plasma 23
   2.4. Ray paths in the ionosphere 27
   2.5. The Doppler shift of the signals of the beacon satellites 32
   2.6. References 40

3. EXPERIMENTAL STUDY OF TRAVELLING IONOSPHERIC DISTURBANCES 43
   3.1. Introduction 43
   3.2. Observations 44
   3.3. Climatology 47
   3.4. Mean diurnal behaviour of TID amplitudes 50
   3.5. Discussion 52
   3.6. References 56

4. CRITICAL LEVELS FOR INTERNAL GRAVITY WAVES IN A JET TYPE FLOW 60
4.1. Introduction 60
4.2. On the hydrostatic and the Boussinesq approximations 62
4.3. On log-pressure coordinates 70
4.3.1. The momentum equation 71
4.3.2. The continuity equation 73
4.3.3. The thermodynamic energy equation 74
4.4. Critical levels in a jet-type flow 78
4.5. Two critical levels 79
4.6. The reflection and transmission coefficients 83
4.7. One critical level 91
4.8. Conclusion 94
4.9. References 95

5. PROPAGATION OF INTERNAL GRAVITY WAVES IN A ROTATING FLUID WITH SHEAR FLOW 99
5.1. Introduction 99
5.2. The wave equations in the Boussinesq and hydrostatic approximations 100
5.2.1. The Boussinesq approximation 100
5.2.2. The hydrostatic approximation 103
5.3. The singularities of the equation 106
5.4. The method of solution 109
5.5. Results 110
5.6. Conclusions 112
5.7. References 114

6. ON THE REFLECTION OF TIDES IN THE UPPER ATMOSPHERE 116
6.1. Introduction 116
6.2. The differential equation for the vertical propagation 117
6.3. Reflection of waves in an inhomogeneous medium 122
6.4. Results 126
6.5. Conclusions 129
6.6. References 130

SAMENVATTING 133

CURRICULUM VITAE 135

VOORWOORD

De velen die op enigerlei wijze hebben bijgedragen aan de totstandkoming van dit proefschrift wil ik langs deze weg hartelijk danken.
Een aantal wil ik hier ook noemen:

- de leiding van het Koninklijk Nederlands Meteorologisch Instituut voor de mogelijkheid mijn onderzoeksresultaten in deze vorm te presenteren;
- Henk Tennekes voor het scheppen van een klimaat voor goed onderzoek;
- Reinier Ritsema voor het geven van een grote vrijheid van werken;
- de promotoren prof.dr. F.W. Sluijter en prof.dr. C.J.E. Schuurmans voor de wijze van begeleiding;
- de lezers prof.dr. G. Vossers en prof.dr. M.P.H. Weenink voor hun opbouwende kritiek;
- Le "co-promotor" dr. H. Teitelbaum, qui par les nombreuses discussions a largement contribué à l'élaboration de cette thèse.
- Titus Spoelstra voor de samenwerking en stimulering;
- Cees van Duin voor de vele discussies over golfvoortplantingsproblemen;
- Frans Haaken, Rinus Rauw en Hans Theihzen voor de wijze waarop ze de satellietdata in de hier gebruikte ionosfeergegevens hebben getransformeerd;
- Jan Jehee voor het ontwerpen van de omslag.
- Jos van Bodegraven-Vermeulen voor haar adviezen en de nauwkeurigheid waarmee ze dit manuscript heeft getypt.

Druk: Drukkerij K.N.M.I.
Illustraties: Tekenkamer K.N.M.I.

CHAPTER 1

INTRODUCTION

1.1. Atmospheric Structure

The earth's atmosphere is commonly described as a series of layers defined by their thermal characteristics (fig. 1.1). Specifically, each layer is a region where the change in temperature with respect to altitude has a constant sign. The layers are called "spheres" and the boundary between the connecting layers is the "pause".

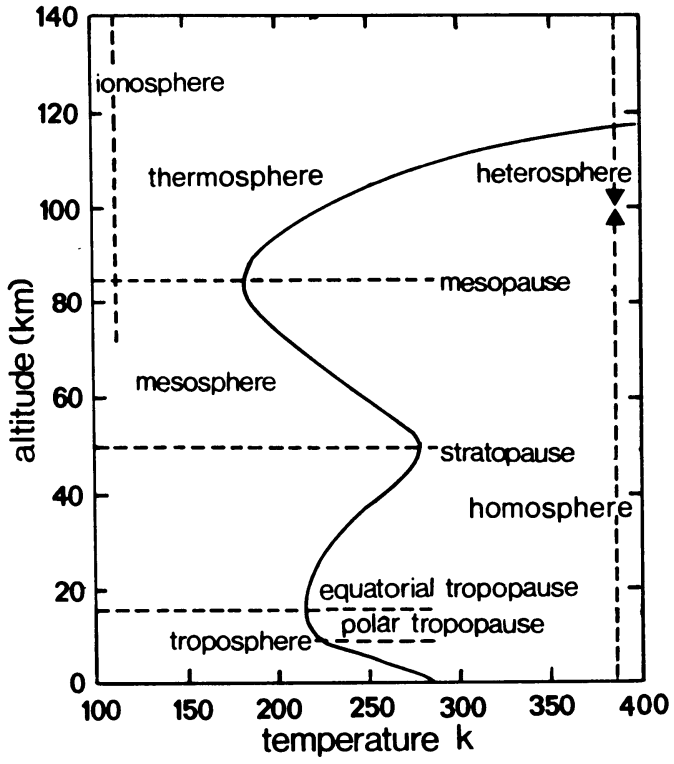

Figure 1.1 Thermal structure of atmospheric layers.

The lowest layer, called the troposphere, exhibits generally decreasing temperatures with increasing altitudes up to a minimum called the tropopause. The temperature and location vary with latitude and season. At the equator, its mean altitude is located near 18 km, and the temperature is roughly 190 K. In the polar regions its elevation is only about 8 km and the temperature roughly 220 K. Above the tropopause the stratosphere begins, exhibiting increasing temperature with altitude up to a maximum of about 270 K at the level of the stratopause located near 50 km. At still higher altitude, the temperature again decreases up to 85 km, where another temperature minimum is found. This layer is called the mesosphere and its upper boundary is the mesopause. In these layers the major constituents, N_2 and O_2, make up about 80 and 20% respectively of the total number density, so that the mean molecular weight of air varies little with altitude. Because of this common feature, the three layers are collectively referred to as the homosphere.

The region located above the mesopause is called the thermosphere. The temperatures there increase very rapidly with altitude and can reach daytime values of 500 to 2000 K, depending on the level of solar activity. The composition at these altitudes is very different from that of the lower regions due to an increasing proportion of atomic oxygen, whose density becomes comparable to and even greater than those of O_2 and N_2 above about 130 km. The abundances of O_2 and N_2 decrease, primarily as a result of rapid photo-dissociation. In contrast to the homosphere, the mean molecular weight of air in this region, therefore, varies with altitude; for this reason the region above 100 km is also called the heterosphere.

The atmosphere above the tropopause is called the upper atmosphere.

The upper atmosphere is the site of substantial motion. Limited evidence has been available to mankind throughout the centuries in the form of auroral displays and meteor trail distortions, but this evidence went largely unnoticed or unappreciated. Scientific consideration can be said to date from 1882, when Stewart advanced the important postulate that motions of the upper atmosphere are responsible through a dynamo action for the geomagnetic variations that are observed at ground level. A century later a wealth of data is available on motions in the upper atmosphere. The last three decades the available data base has explosively been augmented and extended by the development of space technology.

The interpretation of the motion of the upper atmosphere has also reached a rather mature level.

A division of the motions according to the time scales involved is as follows.

a. Prevailing winds. Change with seasons. Dependent on latitude, longitude and altitude. Strong winds, for example at 250 km height windspeeds of 300 m/s are measured.

b. Planetary waves. Time scales of a day or longer. Occur on a global scale. Amplitudes of tens of meters per second. Strongly dependent on season.

c. Tidal oscillations. Periods are integral fractions of either a lunar or a solar day. Amplitudes of tens of meters per second.

d. Acoustic-gravity waves. Periods from fractions of a second to hours. This class of waves contains the well-known sound and infrasound waves but also the internal gravity waves.

e. Turbulence. Time scales of seconds. Turbulence is revealed by variations in the diffusive growth of meteor trails. The cross-section of a trail increases first under the effects of molecular diffusion, but in a matter of seconds eddy diffusion becomes important and ultimately dominates.

With regard to the dynamics of the upper atmosphere, we will restrict ourselves to tides and internal gravity waves and these will be considered in some more detail.

1.2. Tides

The sea tides, with rise and fall of the water twice daily on most coasts, have been known from time immemorial. The explanation of tides was first indicated by Newton in his Principia Mathematica. They are a consequence of the lunar and solar gravitational forces. Newton realized that the tidal forces must affect the atmosphere as well as the oceans, but thought "with so small a motion that no sensible wind will be thence produced" (Newton, 1687).

Atmospheric tides were first measured by using a barometer. The vertical accelerations of the air are so small that the barometer effectively measures the weight of the overlying air, thus an above-normal barometric height implies a heaping up of air above the station. In the tropics, the barometer does show a marked semidiurnal variation, but its period is half a solar, not lunar day. This is illustrated in the historical figure 1.2 for five days of November 1919, at Batavia (presently Jakarta) in Indonesia, at 6.5°S latitude, and also at the temperate zone station Potsdam (52.4° N) where the barometer undergoes larger irregular variations associated with weather changes with small tidal variations superimposed (Bartels, 1928).

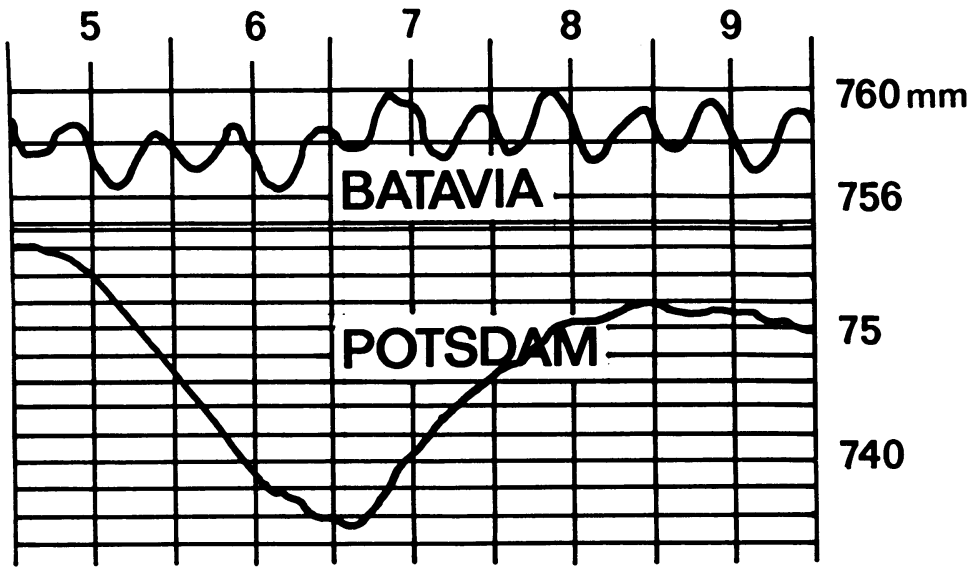

Figure 1.2 Historical registration of the barometric variations (on twofold different scales) at Batavia (6° S) and Potsdam (52° N) during November 1919. After Bartels (1928).

In the upper atmosphere important sources of tidal wind information are the ionized trails left by the numerous meteors disintegrating there. These trails are carried by the neutral wind and may be tracked from the ground by the observations of reflected radio signals.

Greenhow and Neufeld (1961) were the pioneers of a systematic analysis of the horizontal wind above Joddrel Bank (53.2° N, 2.3° E) averaged over the vertical range 80-100 km. Later on French groups have contributed largely to this kind of tidal wind measurements (Fellous et al., 1975). Diurnal and semidiurnal winds were found at these heights with amplitudes of tens of meters per second.

Comprehensive reviews of atmospheric tides can be found in Chapman & Lindzen (1970) and Kato (1980).

The introduction of tidal theory is greatly simplified if it is assumed that the background wind can be ignored and that the unperturbed atmospheric parameters (ρ,p,T) vary with height z only. The tidal variations are assumed to be small. The tidal oscillations may be analyzed into a number of eigenmodes. The longitudinal variation may be taken simply sinusoidal, while the latitudinal variation is the solution of the Laplace Tidal Equation. The solutions are called "Hough functions". With each of these eigenfunctions a specific parameter h is associated which is known as the equivalent depth. Greatest interest centers on modes which progress around the earth in step with the generating agency, solar or lunar as the case may be. The equivalent depth of such modes will here be denoted h_n^1 and h_n^2 for the diurnal and the semidiurnal components respectively. The "1" and the "2" identify the number of wavelengths in 360° of longitude, while n is an index identifying the form of the latitudinal variation of the mode; it increases in magnitude with increasing complexity of latitudinal variations.

The various modes are excited with various amplitudes by the forcing agencies. In the case of atmospheric tides the main forcing agency is solar heating of the troposphere and the stratosphere by absorption of bands in the solar spectrum by water vapour and ozon respectively (Chapman & Lindzen, 1970). The amplitude of the excited modes depends on the degree of fitting in their latitudinal and longitudinal variations with the forcing agency. It depends also on the degree of fitting in the vertical variation and on the degree to which energy input at one level manifests itself, through propagation, at other levels where further input may be found and where constructive or destructive interference may occur.

It is important to realize that the solar heating of the atmosphere is the main generator of the atmospheric tides. Gravitational attraction is much less important. The consequence is that atmospheric tides are mainly coupled to the sun where as sea tides by the gravitational influence are mainly coupled to the moon.

Here attention will be paid to the vertical propagation of the tidal oscillations.

The value of the equivalent depth of the most important semidiurnal mode, the 2.2 mode, is 7.85 km. This makes the local vertical wavenumber, which is a function of the temperature and the equivalent depth, imaginary over an interval of about 20 km around the temperature minimum in the upper mesosphere. The energy of the 2.2 mode thus tends to be trapped in between the mesopause and the earth surface and a standing oscillation with relatively slight leakage of energy is established.
Higher order semidiurnal modes have smaller equivalent depths and real values of the vertical wavenumber. Accordingly they are less efficiently trapped.

The calculation of the reflection of semidiurnal modes at the temperature profile of the upper atmosphere is treated in some detail in this thesis.

1.3. Internal Gravity waves

Any wave must be associated with some restoring mechanism in a medium in equilibrium. For acoustic waves, the restoring force arises from the compressibility of the medium. For internal gravity waves, it is the buoyancy exerted on a displaced fluid element in a stably stratified fluid.
Consider an element of fluid at some level z_o, in a fluid with density ρ decreasing with height at a rate $-\frac{d\rho}{dz}$.

The situation is depicted in fig. 1.3 a.

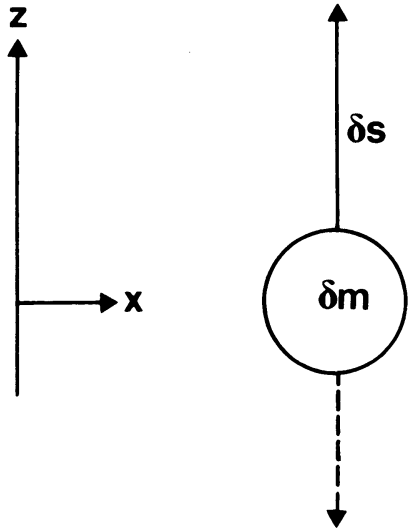

Figure 1.3 a Schematic description of fluid elements, their displacements and buoyancy forces per unit mass.

The mass of the fluid element at z_o is

$$\delta m = \rho(z_o)\, \delta\, v\ ,$$

where δv is the volume of a fluid element. If we displace δm over a small vertical distance δs , it will be subject to a buoyancy force:

$$g \frac{d\rho(z_o)}{dz}\ \delta s\ \delta v\ ,$$

acting to return δm to z_o; g is the acceleration of gravity. Variations of δv due to compressibility have, for simplicity, been neglected.
The equation of motion leads to

$$\frac{d^2\delta s}{dt^2} = \frac{g}{\rho(z_o)} \frac{d\rho}{dz}\, \delta s. \qquad (1.1)$$

In a stably stratified, incompressible fluid $\frac{d\rho}{dz} < 0$. Hence equation (1.1) describes a harmonic oscillation with a frequency ω_B , given by

$$\omega_B^2 = - \frac{g}{\rho} \frac{d\rho}{dz} , \tag{1.2}$$

ω_B is known as the Brunt-Väisälä frequency. The effect of adiabatic expansion is to change the expression for ω_B into the following

$$\omega_B^2 = \frac{g}{T} \left(\frac{g}{c_p} + \frac{dT}{dz}\right) , \tag{1.3}$$

where T is the temperature of the ambient fluid and c_p is the specific heat at constant pressure.

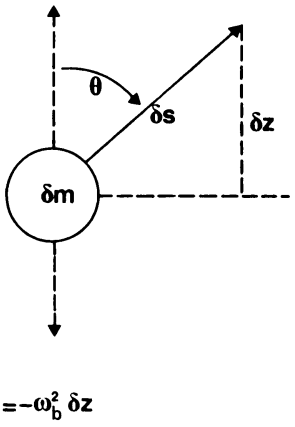

Figure 1.3 b Schematic description of fluid elements, their displacements and buoyancy forces per unit mass.

Let us designate the buoyancy force per unit volume on a displaced fluid element as

$$F_B = - \omega_B^2 \delta s \tag{1.4}$$

F_B is directed vertically. Now consider a fluid element that is somehow constrained to move at some angle θ with respect to the vertical (viz. fig. 1.3 b).

The force exerted on this fluid element will be the projection of the buoyancy force

$$F = -\omega_B^2 \cos^2\theta \, \delta s, \tag{1.5}$$

and the element will oscillate with a frequency ω, given by

$$\omega^2 = \omega_B^2 \cos^2 \theta \, . \tag{1.6}$$

Hence oscillations with all frequencies lower than the Brunt-Väisälä frequency are possible. If the frequency is low the neglecting of the Coriolis force is no longer justified and the simple picture sketched above is no longer valid.

The gravity waves, once excited, propagate through the atmosphere which is an inhomogeneous medium as temperature and wind are functions of the coordinates. Partial reflection and transmission occurs on gradients in the temperature and the wind. In layered wind fields gravity wave energy is not conserved. The waves can loose energy at a so-called critical level, that is the height where the wind velocity equals the horizontal component of the phase velocity. This was first treated by Booker & Bretherton (1967). But the contrary occurs also. Jones (1968) found that if the critical level is situated in a region with a sufficiently low value of the Richardson number, then a gravity wave can tap energy from the background wind.
Two chapters in this thesis are devoted to extensions of the theory of gravity waves propagating through height dependent windfields containing critical levels.

1.4. Ionospheric observations

Ionospheric observations raised also the interest in gravity waves. Frequently, wavelike Travelling Ionospheric Disturbances (TIDs) were observed. In 1950 Martyn suggested already that the TIDs might be the result of buoyancy or gravity waves in the ionosphere. Since then a lot of experimental and theoretical work has been done and is still going on on gravity waves at ionospheric heights. A pioneer in this field was Colin Hines (1960).

Radio-astronomical measurements contain information on TIDs (Kelder & Spoelstra, 1986 b). Radio-astronomical measurements, like those done in Westerbork (the Netherlands), yield very precise determinations of the angle of arrival of the radiation of radio sources. For point sources with well-known positions this angle of arrival can be calculated and differences with the measured values must be caused by refraction either in the troposphere or in the ionosphere. For wavelengths longer than 21 cm the ionospheric refraction in general dominates. Hence it is possible to subtract from angle of arrival measurements certain ionospheric parameters.
The signals of beacon satellites can also be used to measure ionospheric structure. Beacon satellites are in principle designed for position determinations. They are emitting continuously signals at high frequencies which propagate through the ionosphere. As the ionosphere is a dispersive medium the time of travel is dependent on the frequency. By using two frequencies a first order correction can be made for the ionospheric error in the position determination. However, this difference in time of travel can also be used for the determination of some characteristics of ionospheric irregularities.
A chapter is devoted to the interpretation of measurements of TIDs by different techniques.

In this thesis certain aspects of the propagation of high frequency radio waves through the ionosphere of some relevance for radio-astronomical and beacon satellite measurements are also discussed.

Some of the work that is presented in this thesis has been published previously. This is the case with chapter 3 (Spoelstra & Kelder, 1984; Kelder & Spoelstra, 1984 a,b; Kelder & Spoelstra, 1986 a), chapter 4 (Teitelbaum & Kelder, 1985) and chapter 5 (Teitelbaum, Kelder & Van Duin, 1986).

1.5. References

Bartels, J., 1928 - Gezeitenschwingungen der Atmosphäre. Handbuch der Experimental Physik 25 (Geophysik 1), 163-210.

Booker, J.R. & Bretherton, F.P., 1967 - The critical layer for internal gravity waves in a shear flow. J. Fluid Mech. 27, 513-529.

Chapman, S. & Lindzen, R.S., 1970 - Atmospheric Tides. Reidel, Dordrecht.

Fellous, J.L., Bernard, R., Glass, M., Massebeuf, M. & Spizzichino, A., 1975 - A study of the variations of atmospheric tides in the meteor zones. J. atm. terr. Phys. 37, 1511-1524.

Greenhow, J.S. & Neufeld, E.L., 1961 - Winds in the upper atmosphere. Quart. J. Roy. Meteor. Soc. A288, 564-574.

Hines, C.O., 1960 - Internal Atmospheric Gravity waves at Ionospheric Heights. Can. J. Phys., 38, 1441-1481.

Jones, W.L., 1968 - Reflexion and stability of waves in stably stratified fluid with shear flow: a numerical study. J. Fluid Mech. 34, 609-624.

Kato, S., 1980 - Dynamics of the upper atmosphere. Developments in Earth and Planetary Sciences 01, Reidel, Dordrecht.

Kelder, H. & Spoelstra, T.A.Th., 1984 a - Multi-technique study of medium scale TIDs. Kleinheubacher Berichte No. 27, 575-584.

Kelder, H. & Spoelstra, T.A.Th., 1984 b - Multi-technique study of ionospheric irregularities. Proc. Beacon Satellite Studies of the Earth's Environment, New Delhi, India, 457-463.

Kelder, H. & Spoelstra, T.A.Th., 1986 a - Medium-scale TIDs. Accepted by J. atm. terr. Phys.

Kelder, H. & Spoelstra, T.A.Th., 1986 b - Golven in de ionosfeer en hun invloed op sterrenkundige waarnemingen. Ned.Tijdschr.Natuurkunde A 52(2), 64-67.

Martyn, D.F., 1950 - Cellular atmospheric waves in the ionosphere and troposphere. Proc. Roy. Soc. London, Ser. A, 201, 216-234.

Newton, I. - Sir Isaac Newton's Mathematical Principles of Natural Philosophy and his System of the World. 3rd ed. (1929), A. Motte Trans., revised by F. Cajori, U. of California, P. Berkeley (1960).

Spoelstra, T.A.Th. & Kelder, H., 1984 - Effects produced by the ionosphere on radio interferometry. Radio Sci. 19, 779-788.

Stewart, B., 1882 - Terrestrial Magnetism. Encyclopaedia Brittanica, 9th ed. Chicago III.

Teitelbaum, H.& Kelder, H.,1985 - Critical levels in a jet-type flow. J. Fluid Mech. 159, 227-240.

Teitelbaum, H., Kelder, H. & Van Duin, C.A., 1986 - Propagation of internal gravity waves in a rotating fluid with shear flow. Accepted by J. atm. terr. Phys.

CHAPTER 2

ELECTROMAGNETIC WAVES OF HIGH FREQUENCY IN THE IONOSPHERE

2.1. Introduction

A large part of the literature on the ionosphere is devoted to the description of the propagation of electromagnetic waves of frequencies of kHz to tens of MHz. Waves with these frequencies are reflected in the ionosphere and hence suited for ground based remote sensing of the ionosphere. For high frequencies the ionosphere becomes transparent. This transparency is not perfect and in the last years some effort has been put in calculating the influence of the ionosphere on high frequency waves. The accuracy of geodesy with the help of navigation satellites is namely limited among other things by the influence of the ionosphere.
In this chapter the higher order terms in the refractive index are calculated and discussed. The influence of the finite temperature of the ionospheric plasma on the refractive index is also analysed.
Finally an analytic expression for the Doppler shift of signals propagating through a wavelike perturbation of the ionospheric electron density is derived and discussed.

2.2. The refractive index for the propagation of electromagnetic waves through the ionosphere

The theory of the propagation of electromagnetic waves in an ionized medium in the presence of an imposed magnetic field is sometimes called the magneto-ionic theory. This theory was developed during the first part of this century, following Marconi's experiments in long-distance radio propagation and Kennelly's and Heaviside's suggestions in 1902 that these waves are reflected from a conducting layer in the upper atmosphere. In the form used today the

theory is mainly the result of the work done by Appleton and Hartree between 1927 and 1932. A thorough discussion can be found in e.g. Ratcliffe's monograph (1951), Stix (1962), Allis et al. (1963). The equation for the refractive index n of an ionized tepid medium with an imposed magnetic field, taking into account the effect of collisions by introducing an effective collision frequency ν (see e.g. Ginzburg, 1961), is generally known as the Appleton-Lassen dispersion formula (Lassen, 1927; Appleton, 1928; Rawer & Suchy, 1967)

$$n^2 = 1-X/\left(1-jZ-\left[\tfrac{1}{2}Y_T^2 / (1-X-jZ)\right]\pm\sqrt{(\tfrac{1}{4}Y_T^4 / (1-X-jZ)^2+Y_L^2)}\right) . \qquad (2.1)$$

X, Y_L, Y_T and Z are dimensionless quantities defined as follows

$$X=\left(\frac{\omega_N}{\omega}\right)^2, \quad Y_L=\omega_L/\omega, \quad Y_T=\omega_T/\omega, \quad Z=\nu/\omega:$$

ω = frequency electromagnetic wave,
ω_N = angular plasma frequency,
ν = electron collision frequency,
$\omega_L = \omega_H \cos\theta$; $\omega_T = \omega_H \sin\theta$.
ω_H = electron gyro-frequency.

where θ is the angle between the direction of propagation of the wave and the direction of the geomagnetic field line. The plus sign corresponds in the quasi-transverse approximation to the so-called ordinary wave, the minus sign to the extraordinary wave (Stix, 1962). Although the Appleton-Lassen equation is strictly valid only for a homogeneous medium, we assume that the relation is also useful for application to slowly varying media, that is, where the changes in the refractive index are small over a distance of a wavelength, i.e. in the approximation of geometrical optics.
We shall first review the values of the different parameters.
The electron gyro-frequency ω_H equals $2\Pi f_H$ and

$$f_H = eB/(2\Pi m) , \qquad (2.2)$$

where e is the electron charge, m the electron mass and B the earth's magnetic field. The earth's magnetic field B can be approximated by the field of a dipole, that is

$$B(r,\lambda) = .32(r_E/r)^3 \sqrt{(1+3\sin^2\lambda)}\ 10^{-4}\ \mathrm{Wbm}^{-2}, \qquad (2.3)$$

B = Field strength of magnetic induction in Wbm^{-2},

r_E = Earth's radius,

r = Geocentric distance,

λ = Geomagnetic latitude.

The value of the field strength B varies roughly by a factor of 2 from the equator to the poles. The variation with height up to 600 km is in the order of 10%. The value of B at 45° latitude and 350 km height is $B = .43\ 10^{-4}$ Wbm^{-2}. The corresponding value of the electron gyro-frequency $f_H = 1.20$ MHz.

The angular plasma frequency ω_N is defined as

$$\omega_N^2 = Ne^2 / (\varepsilon_o\ 4\pi^2 m), \qquad (2.4)$$

where N is the electron density in m^{-3}, and ε_o is the dielectric constant in vacuum.

Substituting values in (2.4) yields a plasma frequency f_N equal to

$$f_N\ (\mathrm{Hz}) = \sqrt{(80.6N)}\ . \qquad (2.5)$$

A high value of N is $3\ 10^{12}\ \mathrm{m}^{-3}$ and the corresponding value of f_N is 15.5 MHz. Typical values of N are about $10^{12}\ \mathrm{m}^{-3}$ which corresponds to $f_N = 9.0$ MHz. The value of N varies considerably. For example, at night it drops to 10% of the daytime value. The average collision frequency of electrons with neutral particles and with ions ν_{en} and ν_{ei} respectively, is given by (see e.g. Banks & Kockarts, 1973)

$$\nu_{en} = 1.8\ 10^8 p_n\ \mathrm{Hz}\ , \qquad (2.6)$$

where p_n is the neutral pressure in Torr,

$$\nu_{ei} = 5.4\ 10^{-5}\ NT^{-3/2}\ \mathrm{Hz}\ , \qquad (2.7)$$

where T is in K.

Table 2.1 gives some specific values of the neutral and electron densities, the temperature, and the collision and plasma frequencies for the different ionospheric layers.

Table 2.1

Layer	D(80 km)	E(110 km)	F_1(180 km)	F_2(300 km)
T(K)	200	320	1130	1450
$N(m^{-3})$	$6\ 10^8$	$6\ 10^{10}$	$1\ 10^{11}$	$2\ 10^{12}$
$N_n(m^{-3})$	$4\ 10^{20}$	$3\ 10^{18}$	$1.5\ 10^{16}$	$1.5\ 10^{15}$
ν_{ei}(Hz)	$1\ 10^1$	$6\ 10^2$	$1\ 10^2$	$2\ 10^3$
ν_{en}(Hz)	$3\ 10^5$	$3\ 10^3$	7	$2\ 10^{-2}$
f_N(Hz)	$2\ 10^5$	$2\ 10^6$	$3.2\ 10^6$	$1.3\ 10^7$

The neutral density is indicated by N_n.
These values are valid at moderate latitudes with a 1500 K thermopause temperature (see Banks & Kockarts, 1973).

Below, we shall discuss some approximations in magnetoionic theory, a subject that has recently drawn some attention (De Munck, 1982; Budden, 1983; Hartmann & Leitinger, 1984; Heading, 1984).

Let us first assume that X, Y_L, Y_T and Z are all in the order of ε, with $\varepsilon << 1$. The electron collision frequency ν in the dispersion equation is assumed to be equal to the sum of the electron-neutral collision frequency ν_{en} and the electron-ion collision frequency ν_{ei}.
The expression (2.1) for the refractive index can then be approximated by

$$n^2 = 1-X/\left(1-jZ-\left\{\frac{Y_T^2}{2}(1+X+jZ) \mp Y_L \left(1+\frac{Y_T^4}{8Y_L^2}\right)+O(\varepsilon^4)\right\}\right) . \qquad (2.8)$$

Equation (2.8) can be reduced to

$$n^2 = 1 - X \pm XY_L - jXZ - X\left(Y_L^2 + \frac{Y_T^2}{2} - Z^2\right) \pm 2jXZY_L + O(\varepsilon^4) \ . \qquad (2.9)$$

For the refractive index n we have

$$n = 1 - \tfrac{1}{2}X \pm \tfrac{1}{2}XY_L - \frac{jXZ}{2} - \frac{X^2}{8} - \frac{X}{2}\left(Y_L^2 + \frac{Y_T^2}{2} - Z^2\right) \pm jXZY_L$$

$$\mp \frac{1}{4} X^2 Y_L + \frac{1}{4} jX^2Z - \frac{1}{16} X^3 + O(\varepsilon^4) \ . \qquad (2.10)$$

This can also be written as

$$n = 1 - \frac{f_N^2}{2f^2} \pm \frac{f_N^2 f_B \cos\theta}{2f^3} - \frac{f_N^2}{2f^4}\left(f_B^2(\cos^2\theta + \tfrac{1}{2}\sin^2\theta) - \nu^2\right) - \frac{1}{8}\frac{f_N^4}{f^4}$$

$$-j\frac{f_N^2}{2f^3}\nu\left(1 \mp 2\frac{f_B\cos\theta}{f}\right) \mp \frac{1}{4}\frac{f_N^4 f_B\cos\theta}{f^5} + \frac{1}{4} j\frac{f_N^4 \nu}{f^5} - \frac{1}{16}\frac{f_N^6}{f^6} + O(\varepsilon^4) \ . \qquad (2.11)$$

From Table 2.1 it can be inferred that for a height of 300 km and for a frequency f of 100 MHz: X=0.017, Y = ω_H/ω =0.012 and Z=10^{-5}. That isZ ~ $O(\varepsilon^2)$.

Hence (2.11) can be written as

$$n = 1 - \frac{f_N^2}{2f^2} \pm \frac{f_N^2 f_B\cos\theta}{2f^3} - \frac{1}{8}\frac{f_N^4}{f^4} - \frac{f_N^2 f_B^2}{2f^4}\left(\cos^2\theta + \frac{\sin^2\theta}{2}\right)$$

$$- j\frac{f_N^2}{2f^3}\nu - \frac{1}{4}\frac{f_N^4 f_B\cos\theta}{f^5} - \frac{1}{16}\frac{f_N^6}{f^6} + O(\varepsilon^4) \ . \qquad (2.12)$$

In measuring distances with the help of satellites and astronomical data some effort has been put into calculating higher order ionospheric corrections (Bertel, 1969; St. Etienne, 1981; De Munck, 1982; Lohmar, 1985). Expressions similar to (2.12) are then used. These authors, however, calculate higher order corrections by making higher order developments in an approximated refractive index equation. In some papers first the quasi-longitudinal approximation has been made and then a higher order development is made. This procedure offers no guarantee that all relevant terms are obtained.

In general the effects of a temperate plasma, deviations of geometrical optics and the influence of ions on the propagation of electromagnetic waves has to be carefully evaluated in the context of higher order developments in an expression for the refractive index. This domain seems to be not fully explored.

A less rigorous result than (2.12) is also obtained by Leitinger (1974) and Hartmann & Leitinger (1984). In the latter paper a worst case estimate of residual ionospheric errors for vertical incidence is given. In their calculation Hartmann and Leitinger use a model of the ionosphere of 200 km thickness and an uniform electron density of $10^{13}\ m^{-3}$. A value of 1.74 MHz is taken for the electron gyro-frequency. The expression used for the refractive index consists of the first five terms of the right-hand side of equation (2.12). From this expression, Hartmann and Leitinger calculated by integrating optical path lengths.

The contributions of:

$\pm \frac{f_N^2 f_B}{2f^3} \cos\theta$, $\frac{f_N^4}{8f^4}$ and $\frac{f_N^2 f_B^2}{2f^4} (\cos^2\theta + \frac{1}{2} \sin^2\theta)$ are called ΔS_A, ΔS_B and ΔS_C

respectively.

Let the contributions of:

$\frac{f_N^2}{2f^3} \nu$, $\mp \frac{1}{4} \frac{f_N^4 f_B \cos\theta}{f^5}$ and $-\frac{1}{16} \frac{f_N^6}{f^6}$ be named ΔS_D, ΔS_E and ΔS_F

respectively.

Inserting the values of the ionospheric parameters (ν is taken to be equal to 2.10^3 Hz, $\theta = 0$):

$$\Delta S_A = \pm \frac{1.4\ 10^{26}}{f^3} \text{ m},$$

$$\Delta S_B = \frac{-1.6\ 10^{34}}{f^4} \text{ m},$$

$$\Delta S_C = \frac{-2.4\ 10^{32}}{f^4} \text{ m},$$

$$\Delta S_D = \frac{-1.6\ 10^{23}}{f^3} \text{ m},$$

$$\Delta S_E = \mp \frac{5.7\ 10^{40}}{f^4} \quad m,$$

and

$$\Delta S_F = \frac{-6.5\ 10^{48}}{f^6} \quad m,$$

where f is in Hz.

Taking a frequency of 100 MHz, then

$\Delta S_A = \pm$ 140m, ΔS_B = -160 m, ΔS_C = -2.4 m ,ΔS_D = -0.16 m, $\Delta S_E = \mp$ 5.7 m and ΔS_F = -6.5 m. Hence ΔS_E and ΔS_F are comparable with ΔS_C and may not be ignored as was done by Hartmann & Leitinger (1984). This also has consequences for the 150/400 MHz corrections, of importance in geodesy; but this aspect will not be pursued here.

In the next section we will deal exclusively with first order corrections and we can simplify (2.12) further to

$$n = 1 - \frac{f_N^2 (x, y, z, t)}{2f^2} . \qquad (2.13)$$

2.3. The temperate plasma correction to high frequency waves

The dispersion relation for electromagnetic waves in a temperate plasma is given by (Yeh & Liu, 1972, ch. 4):

$$\det \vec{\vec{D}} = 0, \qquad (2.14)$$

with

$$\vec{\vec{D}} = k^2 \vec{\vec{I}} - \vec{k}\,\vec{k} - k_o^2 \vec{\vec{K}}, \qquad (2.15)$$

$$k_o^2 = \frac{\omega^2}{c^2} ,$$

$$\vec{k} = k_o \vec{n} = k_o (n \sin\theta, o, n \cos\theta),$$

where θ is the angle with the magnetic field vector (//z axis), and $\vec{k}$ lies in the x,z plane.

$\vec{\vec{K}}$ is given by Yeh & Liu (1972) in equations (4.18.16) and (4.18.17).

Substituting this in (2.15) results in the following expression.

$$\vec{\vec{D}} = \frac{\omega^2}{c^2}\left[\begin{pmatrix} n^2 & 0 & 0 \\ 0 & n^2 & 0 \\ 0 & 0 & n^2 \end{pmatrix} - \begin{pmatrix} n^2\sin^2\theta & 0 & n^2\sin\theta\cos\theta \\ 0 & 0 & 0 \\ n^2\sin\theta\cos\theta & 0 & n^2\cos^2\theta \end{pmatrix} + \right.$$

$$\left. - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \frac{X}{1-Y^2-n^2\delta\,(1-Y^2\cos^2\theta)} * \right.$$

$$\left. \begin{pmatrix} 1-n^2\delta\cos^2\theta & jY\,(1-n^2\delta\cos^2\theta) & n^2\delta\sin\theta\cos\theta \\ -jY\,(1-n^2\delta\cos^2\theta) & 1-n^2\delta & -jY\,n^2\delta\sin\theta\cos\theta \\ n^2\delta\sin\theta\cos\theta & jY\,n^2\delta\sin\theta\cos\theta & 1-Y^2-n^2\delta\sin^2\theta \end{pmatrix} \right] \quad (2.16)$$

The parameter δ takes into account the thermal velocity v_t of the plasma particles, in formula: $\delta = \frac{v_t^2}{c^2} = \frac{\gamma kT}{mc^2}$. In the high frequency approximation only electrons contribute in the propagation of the wave and therefore only the δ corresponding to the electrons appears in (2.16).

Define:

$$X^1 = X/(1-Y^2-n^2\delta(1-Y^2\cos^2\theta)),$$

$$K_{XX} = 1 - X^1\,(\,1 - n^2\,\delta\,\cos^2\theta\,)\,,$$

$$K_{XY} = -\,jX^1Y\,(\,1 - n^2\,\delta\,\cos^2\theta)\,, \quad (2.17)$$

$$K_{XZ} = -\,n^2\,\delta\,X^1\sin\theta\cos\theta\,,$$

$$K_{YY} = 1 - X^1\,(\,1 - n^2\,\delta)\,,$$

$$K_{ZZ} = 1 - X^1\,(1 - Y^2 - n^2\,\delta\sin^2\theta)\,,$$

$$K_{YZ} = +\,jX^1Y\,n^2\,\delta\sin\theta\cos\theta\,.$$

Then $\vec{\vec{D}}$ can be written as

$$\vec{\vec{D}} = \frac{\omega^2}{c^2} \begin{pmatrix} n^2\cos^2\theta - K_{XX} & + K_{XY} & - K_{XZ} - n^2 \sin\theta\cos\theta \\ - K_{XY} & n^2 - K_{YY} & - K_{YZ} \\ - K_{XZ} - n^2 \sin\theta\cos\theta & + K_{YZ} & n^2\sin^2\theta - K_{ZZ} \end{pmatrix} \quad (2.18)$$

A reasonable estimate for the electron temperature in the ionosphere above 200 km height is 2000°K (e.g. Banks & Kockarts, 1973, Aeronomy II, p. 288).
If the adiabatic condition is assumed, then for high frequency plasma waves $\gamma = 3$ is predicted (Yeh & Liu, 1972). With these values the parameter δ is roughly $\approx 10^{-6}$.
This value of δ should, in principle, lead to corrections in the refractive index of the order discussed in the preceding section. However, below we shall show that this is not the case.

2.3.1. The cold plasma

If we take $\delta = 0$ we get the cold plasma equations

$$K_{XX}^{c} = 1 - \frac{X}{1-Y^2}, \quad K_{YY}^{c} = 1 - \frac{X}{1-Y^2},$$

$$K_{ZZ}^{c} = 1-X, \qquad K_{XZ}^{c} = K_{YZ}^{c} = 0,$$

$$K_{XY}^{c} = \frac{jXY}{1-Y^2},$$

and

$$\vec{\vec{D}}^{c} = \frac{\omega^2}{c^2} \begin{pmatrix} n^2\cos^2\theta - K_{XX}^{c} & - K_{XY}^{c} & - n^2\sin\theta\cos\theta \\ + K_{XY}^{c} & n^2 - K_{YY}^{c} & 0 \\ - n^2\sin\theta\cos\theta & 0 & n^2\sin^2\theta - K_{ZZ}^{c} \end{pmatrix}.$$

The dispersion relation follows again from

$$\det \vec{\vec{D}}^{c} = 0 .$$

Introducing $K_o^c = K_{ZZ}^c$, $K_I^c = K_{XX}^c + jK_{XY}^c$, $K_{II}^c = K_{XX}^c - jK_{XY}^c$,

ultimately yields

$$(n^2 - K_o^c) \left[\frac{n^2 (K_I^c + K_{II}^c)}{2} - K_I^c K_{II}^c \right] \sin^2 \theta +$$

$$K_o^c \left[(n^2 - K_I^c) (n^2 - K_{II}^c) \right] \cos^2 \theta = 0 . \qquad (2.19)$$

The longitudinal case: $\theta = 0$,

then (2.19) reduces to

$$n^2 = 1 - \frac{X}{1+Y} ,$$

$$n^2 = 1 - \frac{X}{1-Y} .$$

These expressions describe the refractive index for left- and right-handed circularly polarized waves respectively.

The transversal case: $\theta = \frac{\pi}{2}$,

$$n^2 = K_o^c = 1-X ,$$

$$n^2 = \frac{2 K_I^c K_{II}^c}{K_I - K_{II}} ,$$

or

$$n^2 = 1 - \frac{X (1-X)}{1-Y^2-X} .$$

These expressions describe the ordinary and extraordinary waves respectively.

2.3.2. The temperate plasma

We have to solve

$$\det \vec{\vec{D}} = 0 \quad , \qquad \text{with } \vec{\vec{D}} \text{ given by (2.18).}$$

We consider only two special cases:
$\theta = 0$ and $\theta = \frac{\pi}{2}$.

The longitudinal case: $\theta = 0$.

The expressions in (2.17) reduce to

$$K_{XX} = 1 - \frac{X}{1-Y^2} \quad , \qquad K_{XZ} = K_{YZ} = 0 \quad ,$$

$$K_{XY} = - \frac{jXY}{1-Y^2} \quad ,$$

$$K_{YY} = 1 - \frac{X}{1-Y^2}$$

and

$$K_{ZZ} = 1 - \frac{X}{1-n^2\delta} \ .$$

The determinant becomes:

$$\begin{vmatrix} n^2 - K_{XX} & - K_{XY} & 0 \\ K_{XY} & n^2 - K_{XX} & 0 \\ 0 & 0 & - K_{ZZ} \end{vmatrix} .$$

The only place where δ acts is in K_{ZZ}; that is we recover the two cold plasma modes and one extra mode:

$$n^2 = \frac{1-X}{\delta} \ .$$

This last mode is just the electron plasma mode.

The transversal case: $\theta = \frac{\pi}{2}$.

With $\theta = \frac{\pi}{2}$ the expressions in (2.17) reduce to

$$K_{XX} = 1 - \frac{X}{1 - Y^2 - n^2\delta} \quad , \quad K_{XZ} = K_{YZ} = 0 \quad ,$$

$$K_{XY} = \frac{-jXY}{1 - Y^2 - n^2\delta} \, ,$$

$$K_{YY} = 1 - \frac{X(1 - n^2\delta)}{1 - Y^2 - n^2\delta}$$

and

$$K_{ZZ} = 1 - X \, .$$

Hence we have to solve:

$$\begin{vmatrix} -K_{XX} & -K_{XY} & 0 \\ K_{XY} & n^2 - K_{YY} & 0 \\ 0 & 0 & n^2 - K_{ZZ} \end{vmatrix} = 0 \, ,$$

or

$$(n^2 - K_{ZZ}) \, (K_{XX} \, (n^2 - K_{YY}) - K_{XY}^2) = 0 \, .$$

A One solution is

$$n^2 = K_{ZZ} = 1 - X \quad ,$$

that is just the ordinary cold plasma electromagnetic wave.

B The other solutions are

$$(1 - \frac{X}{1 - Y^2 - n^2\delta}) \, (n^2 - 1 + \frac{X(1 - n^2\delta)}{1 - Y^2 - n^2\delta}) + \frac{X^2Y^2}{(1 - Y^2 - n^2\delta)^2} = 0 \, ,$$

that is

$$(1 - X - Y^2 - n^2\delta) \, ((n^2 - 1) \, (1 - Y^2 - n^2\delta) + X(1 - n^2\delta)) + X^2Y^2 = 0 \, .$$

This can be reduced to

$$\delta^2 n^6 - \delta n^4 ((2 - 2Y^2 - X) + \delta(1 - X)) +$$

$$n^2 (1 - Y^2 + \delta(2-X)) (1 - X - Y^2) + X^2Y^2 - (1 - X - Y^2)^2 = 0 . \qquad (2.20)$$

The equation (2.20) should correspond to expression (4.19.5b), p 208 of Yeh & Liu, but it does not.

Let us first look at the plasma wave.
Define

$$N = n^2,$$

then equation (2.20) can be written as

$$\delta^2 N^3 - \delta N^2 ((2-2Y^2-X) + \delta(1-X)) + N(1-Y^2+\delta(2-X)) (1-X-Y^2) +$$

$$X^2Y^2 - (1-X-Y^2)^2 = 0 . \qquad (2.20a)$$

Suppose $N = \frac{N_o}{\delta}$,

$O(\frac{1}{\delta})$:

$$N_o^3 - N_o^2 (2-2Y^2-X) + N_o(1-Y^2) (1-X-Y^2) = 0.$$

The first solution is

$$N_o = 0 .$$

The other solutions are

$$N_o^2 - (2-2Y^2-X) N_o + (1-Y^2) (1-X-Y^2) = 0 ,$$

hence

$$N_o = \frac{(2-2Y^2-X)}{2} \pm \tfrac{1}{2}\sqrt{(2-2Y^2-X)^2 - 4(1-Y^2)(1-X-Y^2)} ,$$

that is

$$N_o = 1 - Y^2 - X ,$$

$$N_o = 1 - Y^2 .$$

For N we have:

$$N^{(1)} = \frac{1-Y^2-X}{\delta} , \quad N^{(2)} = \frac{1-Y^2}{\delta} .$$

These are again plasma waves.

O(1), O(δ) :

$$N = N_1 + \delta N_2 .$$

Substituted in (2.20a) we obtain

$$N_1 = 1 - \frac{X(1-X)}{1-X-Y^2} ,$$

$$N_2 = \frac{(1 - 2X - Y^2 + X^2)}{(1 - X - Y^2)^3(1 - Y^2)} (-XY^2 + 2Y^4 + XY^4) .$$

Hence the lowest order of the warm plasma correction in the transverse case is

$$\delta XY^2.$$

For high frequency waves (f ~ 100 MHz)

$$X \simeq 10^{-2} ,$$

$$Y^2 \simeq 10^{-4} ,$$

$$\Rightarrow \delta XY^2 \simeq 10^{-12} .$$

This can be ignored. The expression (4.19.8a) of Yeh & Liu (1972) yields a correction of $n^2 = 1 - \delta + O(\delta^2)$. This correction cannot be ignored. As shown above, this is not correct and the temperate plasma correction can indeed be ignored.

2.4. Ray paths in the ionosphere

In discussing ionospheric parameters a spherical coordinate system is often useful. Spherical coordinates r, ϕ and θ are used, with the origin at the centre of the earth, ϕ being the longitude and θ the colatitude.
Let us consider electromagnetic waves of high frequency in the ionosphere. Their wavelength is much shorter than the scale at which the electron density changes. Approximations based on such a condition show that the energy is propagated mainly along special trajectories (group rays), which are not necessarily straight lines. The approximations in question are termed geometric-optical. If we ignore the earth's magnetic field or investigate only propagation perpendicular to it, the geometric-optical treatment may be based on a single scalar equation with a refractive index n varying from point to point.
The optical length of a phase ray path between the points a and b is

$$P = \int_a^b n(r, \phi, \theta)\, ds = \int_{r_a}^{r_b} [1 + r^2\dot{\theta}^2 + r^2\dot{\phi}^2 \sin^2\theta]^{\frac{1}{2}}\, n(r, \phi, \theta)\, dr \,, \qquad (2.21)$$

where

$$\dot{\phi} = \frac{d\phi}{dr} \,, \quad \dot{\theta} = \frac{d\theta}{dr}$$

and

$$(ds)^2 = (dr)^2 \left[1 + r^2 \left(\frac{d\theta}{dr}\right)^2 + r^2 \sin^2\theta \left(\frac{d\phi}{dr}\right)^2\right] .$$

Fermat's principle, known also as the principle of the shortest optical path or the principle of least time, asserts that the optical length

$$P = \int_a^b n \, ds \, ,$$

of an actual phase ray between any two points a and b is shorter than the optical length of any other curve that joins these points and that lies in a certain regular neighbourhood of it. By a regular neighbourhood is meant one that may be covered by rays in such a way that one (and only one) ray passes through each point of it.

From Fermat's principle the following ray equations can be derived

$$\frac{d}{ds}\left(nr^2 \sin^2\theta \frac{d\phi}{ds}\right) = \frac{\partial n}{\partial \phi} \, , \tag{2.22}$$

$$\frac{d}{ds}\left(nr^2 \frac{d\theta}{ds}\right) = \left[\frac{\partial n}{\partial \theta} + nr^2 \sin\theta \cos\theta \left(\frac{d\phi}{ds}\right)^2\right] \, , \tag{2.23}$$

$$\frac{d}{ds}\left(n\frac{dr}{ds}\right) = \frac{\partial n}{\partial r} + nr\left(\left(\frac{d\theta}{ds}\right)^2 + \sin^2\theta \left(\frac{d\phi}{ds}\right)^2\right) \, , \tag{2.24}$$

Let us first treat the case of a refractive index dependent on r only. The equations (2.22), (2.23) and (2.24) simplify to

$$\frac{d}{ds}\left(nr^2 \sin^2\theta \frac{d\phi}{ds}\right) = 0 \, , \tag{2.25}$$

$$\frac{d}{ds}\left(nr^2 \frac{d\theta}{ds}\right) = nr^2 \sin\theta \cos\theta \left(\frac{d\phi}{ds}\right)^2 \, , \tag{2.26}$$

and

$$\frac{d}{ds}\left(n \frac{dr}{ds}\right) = \frac{\partial n}{\partial r} + nr\left(\left(\frac{d\theta}{ds}\right)^2 + \sin^2\theta \left(\frac{d\phi}{ds}\right)^2\right) \, . \tag{2.27}$$

The spherical symmetry allows a choice of axes such that the ray passes through $\theta = 0$. Then it follows from (2.25) that $\frac{d\phi}{ds} = 0$, that is the ray stays in a longitudinal plane.

This allows us to write (2.26) and (2.27) as

$$\frac{d}{ds}\left(nr^2 \frac{d\theta}{ds}\right) = 0 \ , \tag{2.28}$$

$$\frac{d}{ds}\left(n \frac{dr}{ds}\right) = \frac{\partial n}{\partial r} + nr\left(\frac{d\theta}{ds}\right)^2 \ . \tag{2.29}$$

Integration of equation (2.28) yields

$$n\, r \sin i = p \ , \tag{2.30}$$

where p is a constant and i is the angle between the ray and the vector $\vec{r}$. That is n(r) r sin i is constant along the ray path. This can also be proved from (2.29). Note that (2.30) is Snel's law for spherical surfaces.

Since r sin i represents the perpendicular distance d from the origin to the tangent, (2.30) may also be written as

$$nd = \text{constant.} \tag{2.31}$$

This relation is sometimes called the formula of Bouguer and is the analogue of a well-known formula in dynamics, which expresses the conservation of angular momentum of a particle moving under the action of a central force. From (2.30) the ray path can be calculated, namely

$$\frac{d\theta}{dr} = \frac{\tan i}{r} = \frac{p}{r\sqrt{n^2r^2 - p^2}} \ . \tag{2.32}$$

We will now examine the refraction of high frequency radio waves more closely. Following the general lines of a derivation given by Born & Wolf (1970), it can be shown that the radius of curvature ρ of a ray in general (within the limits of ray theory), in an isotropic ionosphere, can be written as

$$\frac{1}{\rho} = \frac{1}{n}\left|\nabla n - \vec{t}\,\frac{dn}{ds}\right| \ , \tag{2.33}$$

where n is the refractive index, $\vec{t}$ is the unit tangent vector to the ray at the point of interest, and s is the arc length of the ray.

This expression can be simplified to

$$\frac{1}{\rho} = \frac{1}{n} |\text{grad } n \sin z| \ , \tag{2.34}$$

where z is the angle between the ray and the direction of the gradient of the refractive index.

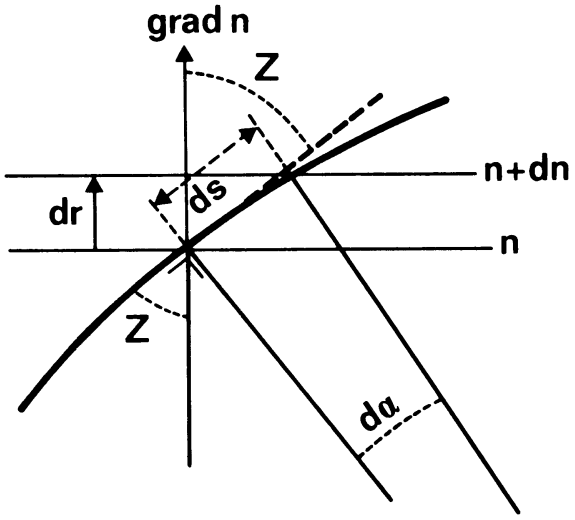

Figure 2.1 The relation between the curvature and the angular refraction of a ray in a medium with a gradient in the refractive index (Snel's law).

With the geometry as shown in figure 2.1 and n dependent on r only, we have

$$\frac{1}{\rho} = \left|\frac{d\alpha}{ds}\right| \ , \tag{2.35}$$

and

$$\frac{d\alpha}{dr} = - \frac{1}{n}\frac{dn}{dr} \frac{p}{\sqrt{n^2r^2-p^2}} \ . \tag{2.36}$$

This formula was also derived by De Munck (1982) (Except for a factor $\frac{1}{n}$).
In the high frequency limit this equation can be simplified to

$$\frac{d\alpha}{dr} = - \frac{dn}{dr} \frac{p}{\sqrt{r^2-p^2}} \ . \tag{2.37}$$

Hence

$$\alpha = - \int_{r_e}^{r} \frac{dn}{dr} \frac{p}{\sqrt{r^2-p^2}} \, dr \, . \qquad (2.38)$$

This formula was used by De Munck (1982) to calculate the refraction of high frequency radio waves in a spherically stratified ionosphere. He obtained analytical expressions for electron density distributions quadratic in r. However, the integral in (2.38) is also solvable in closed form for arbitrary polynomials in r (Gradshteyn & Ryzhik, 1965 p 68). Hence the class of easily integrable electron density profiles is larger.

The integral (2.38) was used by Spoelstra (1983) in calculating the refraction in an ionosphere that has horizontal gradients also. However, expression (2.38) is only strictly valid in a spherically stratified ionosphere. It has to be proved that this is a good approximation.

Another expression for refraction in a spherically stratified ionosphere was derived by Chvojková (1958) and used by Komesaroff (1960) and Spoelstra (1983). Equation (2.29) can be written as

$$\frac{dn}{dr} \cos i - n \sin i \frac{di}{dr} = \frac{1}{\cos i} \frac{dn}{dr} + \frac{n}{r} \frac{\sin^2 i}{\cos i} \, , \qquad (2.39)$$

or

$$\frac{di}{dr} = - \frac{\tan i}{n} \frac{dn}{dr} - \frac{\tan i}{r} \, . \qquad (2.40)$$

The second term on the right-hand side is nothing else than the change in i for the refractive index n = 1. Define $i(r) = i_{n=1}(r) + \alpha(r)$ then

$$\frac{d\alpha}{dr} = - \frac{\tan i}{n} \frac{dn}{dr} \, . \qquad (2.41)$$

This is nothing else than (2.36).

2.5 The Doppler shift of the signals of the beacon satellites

The phase path ϕ can be expressed as

$$\phi = \frac{l}{\lambda} + ft = \frac{fl}{c} + ft \ , \tag{2.42}$$

where the optical path $l = {}_o\int^s n \, ds$, f is the frequency, t is time, λ is the wavelength in vacuum, c the speed of light in vacuum, n is the real part of the refractive index, o is the position of the observer and s the position of the source. For high frequencies (i.e. GHz range), n can be approximated by (see section 2.2)

$$n = 1 - \frac{f_N^2}{2f^2} \ . \tag{2.43}$$

From equation (2.42) and (2.43) we get

$$\phi = ft + \frac{f}{c} {}_o\int^s ds - \frac{e^2}{8\pi^2 m\varepsilon_o cf} {}_o\int^s N \, ds \ . \tag{2.44}$$

The decrease of the phase path equals

$$\phi' = \frac{e^2}{8\pi^2 m\varepsilon_o cf} {}_o\int^s N \, ds \ . \tag{2.45}$$

If the values of the constants are substituted, equation (2.45) gives

$$\phi' = \frac{1.34 \quad 10^{-7}}{f} \, TEC \ . \tag{2.46}$$

The total electron content, TEC, between the satellite and the receiver is defined as

$$TEC = {}_o\int^s N(\underline{r}) ds \ . \tag{2.47}$$

The quantity TEC can be expressed as

$$TEC = \frac{1}{\cos \chi} {}_o\int^s N \, ds = D(\chi) N_{\perp} \ , \tag{2.48}$$

where χ corresponds roughly to the zenith angle at the altitude of maximum electron density. A formal definition of the geometrical factor D is

$$D(\chi) = \frac{1}{\cos \chi} = \left\{ \int_0^{h_s} N \frac{1}{\cos i(h)} \, dh \right\} / \left\{ \int_0^{h_s} N \, dh \right\} , \tag{2.49}$$

where i(h) is the zenith angle at altitude h and h_s is the height of the satellite. The geometrical factor D is a slowly varying function with values between 1 and 3 (Leitinger and Putz, 1978).

The frequency shift, Δf, being the consequence of the time-dependent phase shift can be written as

$$\Delta f = \frac{d\phi'}{dt} = \frac{1.34 \ 10^{-7}}{f} \frac{d}{dt}(TEC) . \tag{2.50}$$

The relative difference in frequency shift with respect to the frequency of 400 MHz equals

$$Df = \left(\frac{8}{3} \frac{1}{150 \ 10^6} - \frac{1}{400 \ 10^6} \right) 1.34 \ 10^{-7} \frac{d}{dt} (TEC) , \tag{2.51}$$

or

$$Df = 2.05 \ 10^{-15} \frac{d}{dt} (TEC) . \tag{2.52}$$

From (2.48)

$$\frac{d}{dt} (TEC) = \frac{d}{dt} D(\chi(t)) \, N_\perp(x(t),y(t)) + D(\chi(t)) \frac{d}{dt} N_\perp(x(t),y(t)) \tag{2.53}$$

where the horizontal coordinate x is the coordinate parallel to the meridian from north to south, while the horizontal coordinate y is orthogonal to x and from east to west. For satellites moving in a polar orbit with a sufficiently high velocity, only the variation of $N_\perp$ in the coordinate x has to be considered. We assume a small periodic variation of $N_\perp$ of the form:

$$N_\perp(x) = N_\perp^0 + N_\perp \sin kx . \tag{2.54}$$

With this, and assuming that $\frac{dD}{dt} << D$ and $N_\perp^0 >> N_\perp'$, equation (2.53) yields

$$\frac{d}{dt}(TEC) = N_\perp^0 \frac{dD}{dt} + D(\chi) \frac{dx}{dt} N_\perp' k \cos kx . \qquad (2.55)$$

Thus Df reduces to

$$Df = 2.05 \ 10^{-15} \left(N_\perp^0 \cos^{-2}\chi \ \frac{d\chi}{dt} + \frac{1}{\cos\chi} \frac{dx}{dt} \ N_\perp' k \cos kx\right). \qquad (2.56)$$

The first term of the right-hand side is a slowly varying quantity. The second term reflects wavelike perturbations. In equation (2.48) $N_\perp'$ also contains observational biases such as, for example, the projection of the line of sight along the wave fronts.

A more thorough discussion on the Doppler shift can be found in Bennet & Dyson (1982). We will retrace their derivation and show that in good approximation analytic expressions can be obtained which are useful for the interpretation of Doppler shift measurements. These authors use a general variational theory to obtain an expression for the contribution of irregularities to the Doppler shift. The basis of the method is that the phase path may be expanded in a multiple Taylor-McLaurin series and the coefficients evaluated assuming that for all combinations of values of the variables the ray continues to exist. It is thus essentially an imbedding technique. The calculations are simplified because the ray has to satisfy a variational equation (Fermat's principle). Specifically, a quasi-stationary approach to time variations is adopted and the refraction caused by the background ionosphere is assumed to be a first order perturbation of the free space ray. Then

$$P \sim P_o + (\delta_m P + \delta_r P) + \tfrac{1}{2}(\delta_m{}^2 P + 2\delta_m\delta_r P + \delta_r{}^2 P), \qquad (2.57)$$

where P is the phase path and P_o is the value of the phase path in the absence of the refracting ionosphere. $\delta_m P$ is the first m-variation of P which represents the first order (linear) contribution caused by the smooth background ionosphere and $\delta_r P$ represents the corresponding contribution caused by the irregularity. At the high frequencies used in satellite Doppler measurements the ionospheric refraction is small, and treating it as a first order perturbation is a good approximation.

Making use of the Doppler shift formula

$$\Delta f = - \frac{f}{c} \frac{dP}{dt} , \tag{2.58}$$

the following expression is obtained

$$\Delta f \approx - \frac{f}{c} \left(\frac{dP_o}{dt} + \frac{d}{dt} \delta_m P + \frac{d}{dt} \delta_r P + ... \right) . \tag{2.59}$$

In equation (2.59) the Doppler shift is represented as the sum of the free space contribution and the first order contributions of the smooth ionosphere and the irregularity. The second and third term in equation (2.59) may be evaluated using the general formula for the second variation of P (Bennet, 1969, 1973). Since we are interested in irregularities we will consider the third term only.

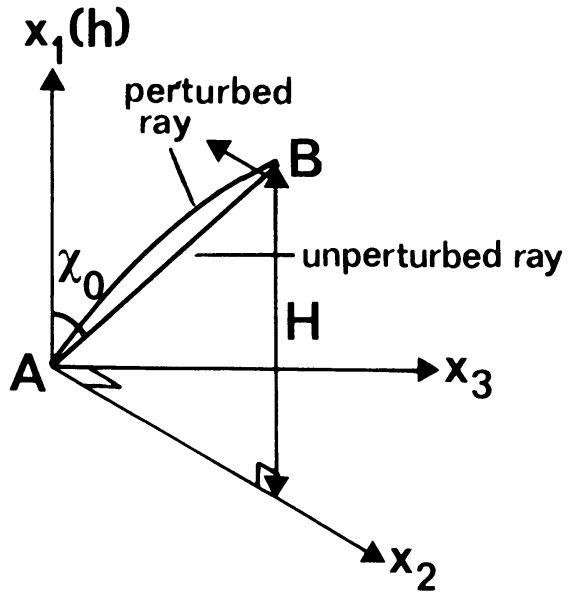

Figure 2.2 Coordinate system for observer at A and satellite at B.

Consider the simple case of a satellite passing the local zenith of the observer, as sketched in figure 2.2. Then the Doppler shift caused by the irregularity equals

$$\frac{d}{dt}(\delta_r P) = \frac{V_2}{\cos \chi_o} \frac{1}{H} \int_o^H h\mu_{rx_2} \delta r \, dh + \frac{V_2 \sin \chi_o}{H} \int_o^H \mu_r \delta r \, dh , \qquad (2.60)$$

where V_2 is the horizontal component of the velocity of the satellite. χ_o is the zenith angle and H is the height of the satellite. μ is the refractive index. x_2 is the horizontal coordinate. The variation in refractive index $\mu_r \delta r$ equals

$$\mu_r \delta r = - \frac{K}{2f^2} N_m \, \delta m \, w , \qquad (2.61)$$

where $K = \frac{e^2}{4\pi^2 m \, \varepsilon_o}$ and $N_m \, \delta m$ denotes the smooth electron density distribution and w is the perturbation of the background electron density.

The derivative term $\mu_{rx_2} \delta r$ reads

$$\mu_{rx_2} \delta r = - \frac{K}{2f^2} N_m \, \delta m \frac{\partial w}{\partial x_2} , \qquad (2.62)$$

where it is assumed that $N_m \delta m$ is independent of x_2.

As a simple model for the perturbation w of the background electron density distribution is taken:

$$w(x_1, x_2) = \exp \left\{ - \left(\frac{h - h_o}{H_s} \right)^2 \right\} \cos \left(2\pi \left(\frac{x_1}{\Lambda_1} - \frac{x_2}{\Lambda_2} \right) \right), \qquad (2.63)$$

where h_o is the height of the maximum percentage wave amplitude, H_s is the scale height of the wave and x_1, x_2 and Λ_1, Λ_2 are the vertical and horizontal coordinates and wavelengths respectively.

The first term on the right-hand side of equation (2.60), indicated from now on by I_{x_2}, is an integral over the horizontal gradient of the perturbation weighted by the height. The second term, indicated by I, is an integral over the perturbation.

From (2.60), (2.62) and (2.63) we obtain

$$I_{x_2} = - \frac{V_2 K \pi}{f^2 \cos \chi_o \Lambda_2 H} \int_o^H h N_m \delta m \; e^{-\left(\frac{h-h_o}{H_s}\right)^2} \sin \left(2\pi h \left(\frac{1}{\Lambda_1} - \frac{\tan \chi_o}{\Lambda_2}\right)\right) dh \; . \qquad (2.64)$$

Suppose that $N_m \delta m$ is independent of height in the region of interest, then (2.64) reads

$$I_{x_2} = - \frac{V_2 K \pi \; N_m \delta m}{f^2 \cos \chi_o \; \Lambda_2 H} \int_o^H h e^{-\left(\frac{h-h_o}{H_s}\right)^2} \sin \left(2\pi \; h \left(\frac{1}{\Lambda_1} - \frac{\tan \chi_o}{\Lambda_2}\right)\right) \; dh \; . \qquad (2.65)$$

With the substitution $h^1 = h - h_o$, the integral in (2.65) becomes

$$\int_{-h_o}^{H-h_o} (h^1 + h_o) \; e^{-\frac{h^{1\,2}}{H_s^2}} \sin (a \; h^1 + a \; h_o) \; dh^1 \; . \qquad (2.66)$$

where

$$a = 2\pi \left(\frac{1}{\Lambda_1} - \frac{\tan \chi_o}{\Lambda_2}\right) \; . \qquad (2.67)$$

Now $\left(\frac{H-h_o}{H_s}\right)^2 >> 1$ and $\left(\frac{h_o}{H_s}\right)^2 >> 1$.

Hence we may approximate the last integral by

$$\int_{-\infty}^{+\infty} h^1 \; e^{-\frac{h^{1\,2}}{H_s^2}} (\sin a \; h^1 \cos a \; h_o + \cos a \; h^1 \sin a \; h_o) \; dh^1 +$$

$$h_o \int_{-\infty}^{+\infty} (\sin a \; h^1 \cos a \; h_o + \cos a \; h^1 \sin a \; h_o) \; dh^1 \; . \qquad (2.68)$$

The non-zero contributions in (2.68) are

$$\cos a \; h_o \int_{-\infty}^{+\infty} h^1 \sin a \; h^1 \; e^{-\frac{h^{1\,2}}{H_s^2}} \; dh^1 + h_o \sin a \; h_o \int_{-\infty}^{+\infty} \cos a \; h^1 \; e^{-\frac{h^{1\,2}}{H_s^2}} \; dh^1 \qquad (2.69)$$

The integrals in (2.69) are given in Gradshteyn & Ryzhik (1965), p. 495 and 480 respectively. The result is

$$\sqrt{\pi}\, H_s \left(\frac{a^2 H_s^4}{4} + h_o^2\right)^{\frac{1}{2}} \sin(a h_o + \alpha)\, e^{-\frac{a^2 H_s^2}{4}}, \qquad (2.70)$$

where $\tan \alpha = \frac{a H_s^2}{2 H_o}$.

Using (2.70) I_{X_2} reads

$$I_{X_2} = -\frac{V_2 K}{f^2 \cos \chi_o \Lambda_2 H}\, \pi^{3/2} H_s N_m \delta m \left(\frac{a^2 H_s^4}{4} + h_o^2\right)^{\frac{1}{2}} \sin(a h_o + \alpha)\, e^{-\frac{a^2 H_s^2}{4}} . \qquad (2.71)$$

The integral in I can be approximated in the same way, resulting in:

$$I = -\frac{V_2 K}{2 f^2 H} N_m \delta m \sqrt{\pi} \frac{H_s}{H} e^{-\frac{a^2 H_s^2}{4}} \left[\sin \chi_o \cos a h_o + \right.$$

$$\left. \frac{\pi}{\Lambda_2 \cos \chi_o} \left(\frac{a^2 H_s^4}{4} + h_o^2\right)^{\frac{1}{2}} \sin(a h_o + \alpha)\right] . \qquad (2.72)$$

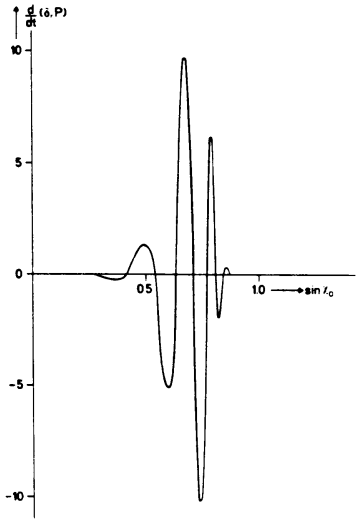

Figure 2.3 The frequency shift of a high-frequency signal as a function of the angle of incidence in the case of a wavelike perturbation in the electron density of the ionosphere.

In figure 2.3 the dependence of $\frac{d}{dt}(\delta_r P)$ as function of sin χ_o is drawn for $\Lambda_1 = \Lambda_2 = 100$ km, $H_s = 100$ km, $h_o = 250$ km and $H = 1000$ km. The shape of $\frac{d}{dt}(\delta_r P)$ agrees with commonly observed patterns in the Doppler shift measurements (see e.g. Stolp, 1985).

The largest values of $\frac{d}{dt}(\delta_r P)$ are found around a = 0, that is $\tan \chi_o = \frac{\Lambda_2}{\Lambda_1}$. That means that the zero-order ray is aligned with the phase fronts of the irregularity. The decay of the amplitude is exponential and depends on Λ_1, Λ_2 and H_s. The angles χ of the first maximum or minimum around χ_o are given by

$$\chi = \chi_o \pm \frac{\Lambda_2}{4h_o} \cos^2 \chi_o \ . \tag{2.73}$$

Hence, if the height of maximum density is known, the horizontal wavelength can be estimated.

More general cases can be discussed along the same lines. For example, instead of (2.61) a perturbation can be taken of the form

$$w(x_1,x_2,x_3) = \exp\left(-\left(\frac{x_1-h_o}{H_s}\right)^2\right) \cos\left(2\pi\left(\frac{x_1}{\Lambda_1} - \frac{x_2}{\Lambda_2} - \frac{x_3}{\Lambda_3}\right)\right). \tag{2.74}$$

Suppose the satellite has a horizontal velocity $(0, V_2, V_3)$. , then instead of (2.60) we obtain

$$\begin{aligned}\frac{d}{dt}(\delta_r P) = {} & \frac{V_2}{\cos \chi_o} \frac{1}{H} \int_o^h h \mu_{r x_2} \delta r \, dh + \\ & + \frac{V_2 \sin \chi_o}{H} \int_o^H \mu_r \delta r \, dh + \\ & + \frac{V_3}{\cos \chi_o} \frac{1}{H} \int_o^H h \, \mu_{r x_3} \delta r \, dh \ .\end{aligned} \tag{2.75}$$

This leads to the same kind of integrals as above and analytical expressions can be obtained in an analogue way.

2.6. References

Allis, W.P., Buchsbaum, S.J. & Bers, A., 1963 - Waves in anisotropic plasmas, M.I.T. Press Cambridge.

Appleton, E.V., 1928, 1930 - Some notes on wireless methods of investigating the electrical structure of the upper atmosphere. I. Proc. Phys. Soc. 41, 43-59. II. Proc. Phys. Soc. 42, 321-339.

Banks, P.M. & Kockarts, G., 1973 - Aeronomy, Part A and B, Academic Press, New York.

Bennet, J.A., 1969 - On the application of variation techniques to the ray theory of radio propagation. Radio Sci. 4, 667-678.

Bennet, J.A., 1973 - Variations of the ray path and phase path: A Hamilton formulation. Radio Sci., 8, 737-744.

Bennet, J.A. & Dyson, P.L., 1982 - The effect of fairly large-scale ionospheric irregularities on satellite Doppler shift. J. atm. terr. Phys. 44, 347-358

Bertel, L., 1969 - Effet ionosphérique de 1er ordre. Ann. Géoph., 25, 85-91.

Born, M. & Wolf, E., 1970 - Principles of Optics. Fourth Edn., Pergamon Press.

Budden, K.G., 1983 - Approximation in magnetoionic theory. J. atm. terr. Phys. 45, 213-218.

Chvojkova´, E., 1958 - Refraction of radiowaves in an ionised medium I, II and III. Bull. Astron. Inst. Czech. 9, 1-5, 6-9 and 133-138.

De Munck, J.C., 1982 - De ionosfeercorrectie bij afstandsmeting naar een buitenaards object. Geodesia, 429-433.

Ginzburg, V.L. 1961 - Propagation of Electromagnetic Waves in Plasmas. Gordon and Breach, New York.

Gradshteyn, I.S. & Ryzhik, I.M., 1965 - Table of Integrals, Series, and Products. Fourth Edn., Academic Press, New York.

Hartmann G.K. & Leitinger, R., 1984 - Range errors due to ionospheric and tropospheric effects for signal frequencies above 100 MHz. Bull.Geod. 58, 109-136.

Hartree, D.R., 1929 - The propagation of electromagnetic waves in a stratified medium. Proc. Camb. Phil. Soc. 25, 97-120.

Heading, J., 1984, - Approximations to the plasma refractive index. J. atm. terr. Phys. 46, 1169-1178.

Komesaroff, M.M., 1960 - Ionospheric refraction in radio astronomy. Austr. J. Phys. 13, 153-167.

Lassen, H., 1927 - Elektr. Nachr. Techn. 4, 324-334. Eq. 18.

Leitinger, R., 1974 - Der Einfluss ionosphärischer Ausbreitungsfehler bei der geodätische Anwendung von Navigationssatelliten. Kleinheubacher Berichte, 17, 321-335.

Leitinger, R. & Putz, E., 1978 - Die Auswertung von Differenz-Doppler-Messungen an den Signalen von Navigationssatelliten. Technischer Bericht, Institut für Meteorologie u. Geophysik, Universität Graz.

Lohmar, F.J., 1985 - Zur berechnung ionosphärischer Refraktions - Korrekturen für VLBI-Beobachtungen aus simultanen Dopplermessungen nach Satelliten. Mitteilungen aus den Geodätischen Instituten der Rheinischen Friedrich-Wilhelms-Universität, Berlin, nr. 67.

Ratcliffe, J.A., 1951 - The magnetoionic theory and its application to the ionosphere. Cambridge University Press.

Rawer, K. & Suchy, K., 1967 - Radio-observations of the ionosphere. Handbuch der Physik, band XLXX/2, Geophysik III, 1-546.

Saint-Etienne, J., 1981 - Erreur ionosphérique residuelle dans les systèmes de radiolocalisation spatiale bifréquences. Ann. Géoph. 37, 241-266.

Stolp, E.R., 1985 - Voortplanting van zwaartegolven in de atmosfeer en rond een kritieke laag. Scientific Report W.R. 85-1, KNMI, Netherlands.

Spoelstra, T.A.Th., 1983 - The influence of ionospheric refraction on radio astronomy interferometry. Astron. Astrophys. 120, 313-321.

Stix, T.H., 1962 - The theory of plasma waves. McGraw-Hill, New York.

Yeh, K.C. & Liu, C.H., 1972 - Theory of Ionospheric waves. Academic Press. international Geophysics Series, vol. 17.

CHAPTER 3

EXPERIMENTAL STUDY OF TRAVELLING IONOSPHERIC DISTURBANCES

3.1. Introduction

It is well known that the ionosphere is not a homogeneous medium. However, only during the last 30 years a more or less systematic search has been made for irregularities in the ionosphere. An important impetus to this study was given by the classic paper by Hines (1960) on the interpretation of irregularities in terms of internal gravity waves. Since then, the knowledge of internal gravity waves has considerably been extended. Several review papers and books are witnesses to this; see e.g. Hines (1974), Kato (1980). While looking for methods to correct radio-astronomy observations for ionospheric refraction, it was realized that a radio interferometer is also a sensitive tracer of ionospheric behaviour. High precision measurements made with the Westerbork Synthesis Radio Telescope (WSRT) in the Netherlands deliver information about the difference in total electron content at scales of the interferometer baselines ranging from 36 m to 2.7 km. In many observations strikingly clear wavelike patterns were present. According to the usual classification these are medium-scale travelling ionospheric disturbances (MS TIDs).

At De Bilt, the differential Doppler shift in the signals of satellites of the Navy Navigation Satellite System (NNSS) is determined. These satellites move in polar orbits around the earth. Hence the measurements contain mainly information on the north-south component of ionospheric disturbances. By following the satellite, a north-south section through the ionosphere of up to 4000 km long can be covered.

A general review of satellite beacon contributions to studies in the structure of the ionosphere can be found e.g. in Leitinger et al. (1975), Leitinger & Putz (1978) and Evans (1977).

Measurements with both techniques were combined. Climatological characteristics were determined. The direction of propagation of the medium scale TIDs seems to be mainly south - south-west. Fourier analysis of the amplitudes of the measured irregularities showed two periods of equal strength: one of 12h and the other of 24h. Given the characteristics of medium scale TIDs it seems probable that these are related to atmospheric tides.

3.2. Observations

From 8 January - 31 March, 1982 and from 24 December, 1982 - 16 March, 1983 the WSRT (52.9°N, 6.6°E) was used for observations at 608.5 MHz. The WSRT consists of an array of 14 steerable telescopes, each with a diameter of 25 m. They are situated along an east-west baseline. Ten of them occupy fixed positions at 144 m intervals. Four additional paraboloids are movable and serve as references against which the phase of the radiation received by the fixed antennas is measured. To do this, the fixed antennas are connected to the four additional paraboloids to form correlation interferometers. The back end of each correlation interferometer consists of a correlator system which measures the four complex correlation components necessary to characterize the polarization state of the radiation. The maximum baseline between two telescopes is 2.7 km.

A source is observed by tracking it in its diurnal rotation from 6 hours before to 6 hours after meridian transit, or over some fraction of this range. The array beam is continuously steered by proper phasing of the elements. This is done with very high precision.

Point sources are observed frequently for checking and calibrating the instrument. The standard integration time for these observations is 1 minute. For point sources, we know that the incident wavefront should in good approximation be perfectly flat as the distance to the astronomical source is large; deviations must be the result of propagation effects giving rise to path-length errors (e.g. Hamaker, 1978; Spoelstra, 1983).

These deviations are visible as phase errors, $\Delta\phi$, which are to first order proportional to the baseline between two interferometer elements.

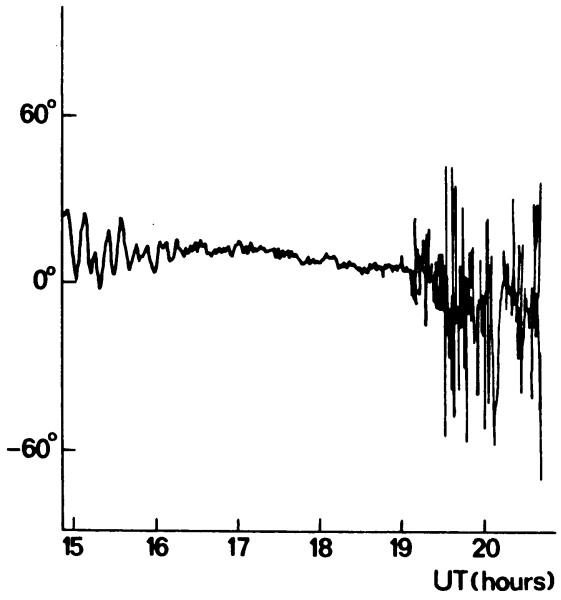

Figure 3.1 Example of the variation of the phase differences between two telescopes. The source is 3C147, the date is 21 February 1983. The used baseline was 2718 m. The frequency 608.5 MHz.

Many observations showed phase errors much larger than the internal accuracy of the system. These errors show often a wavelike behaviour as a function of the hour angle (= time). An example is given in figure 3.1. Their dependence on baseline and frequency of observation indicates that they are the result of ionospheric irregularities such as TIDs. For these wavelike patterns it is possible to derive the amplitudes as a function of interferometer baseline (the amplitude is half of the peak to peak of the variations of the phase errors), and as a function of the time of the day. Another quantity which can be determined is the apparent period of the wave.

The Navy Navigation Satellite System consists of five satellites moving in polar orbits at 1100 km altitude with a velocity of 7.5 km/s. They continuously transmit two coherent signals at frequencies of 150 and 400 MHz.

At the Royal Netherlands Meteorological Institute (KNMI) in De Bilt (52.1°N, 5.2°E) we use a receiver originally designed for geodetic Doppler observations. With this apparatus, a value is determined for the phase difference at each frequency, every 4.6 seconds, using the receiver's oscillator. The integration time of 4.6 sec means roughly a minimum horizontal scale of 12 km at 350 km height. As our aim was to study ionospheric irregularities we looked at the derivative of the phase difference, i.e. the frequency difference. An example is given in fig. 3.2. In total nearly 500 satellite passages have been analysed for these periods.

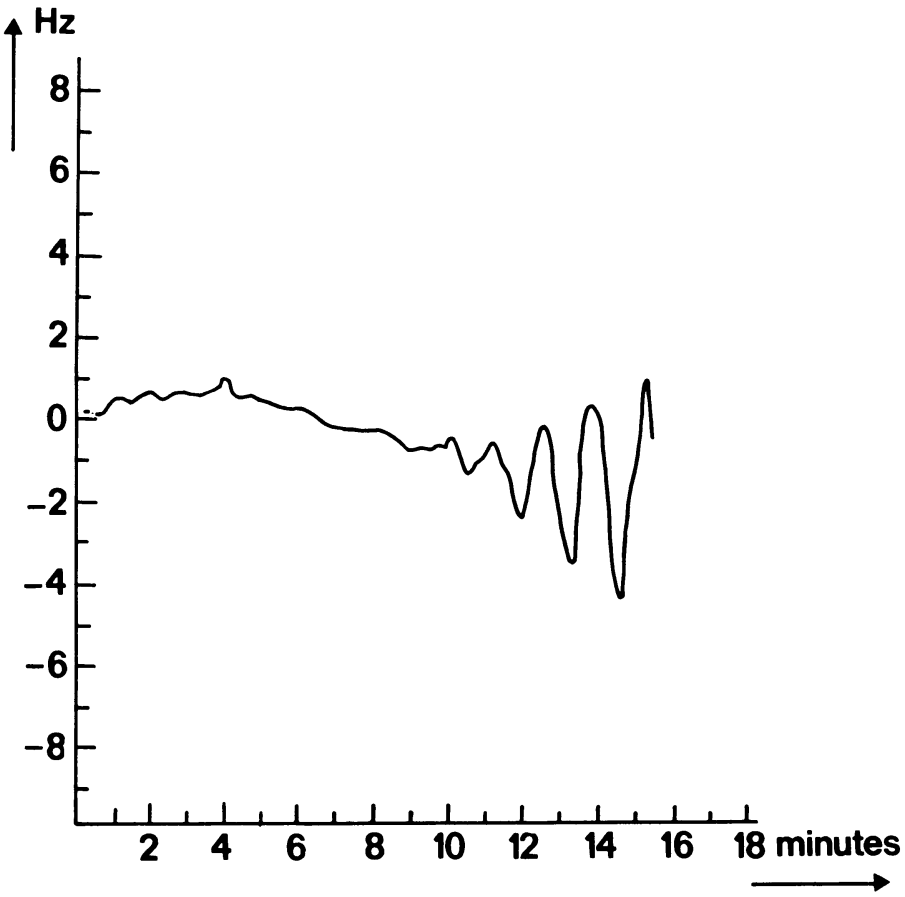

Figure 3.2 Example of the variation of the differential Doppler shift between the 150 and 400 MHz signals during a passage of the satellite. The measurement was carried out at 11 February 1985. The registration starts at 12h21m UT and the maximum elevation of the satellite is 46°.

In all cases, in which ionospheric irregularities were observed with the WSRT they were also present in simultaneous differential Doppler measurements. This does not necessarily imply that both instruments were looking for the same irregularities, but it strongly suggests that the domain in which irregularities occur may be several hundred kilometres wide. A suggestion also supported by the work of Stolp (1985) on measurements made simultaneously with two and three receivers of satellite signals.

3.3. Climatology

From the phase variations of the WSRT observations we determined the amplitude A and apparent period P of each irregularity as a function of time for the longest baseline (2.7 km). Figures 3.3 and 3.4 show A as a function of time (0 - 24 hours UT). The instrumental noise at 608.5 MHz is about 0.5° in phase. Noise in phase of up to about 4° may be caused by the presence of low level structure in some of the fields observed (e.g. 3C286). Figures 3.3 and 3.4 show a strikingly systematic pattern for the occurrence of ionospheric irregularities, a peak being reached around noon. With respect to noon the pattern is clearly asymmetric, for there is a sharp decrease after noon in A. When the sun is below the horizon significant irregularities are still observed. However, the apparent periods P are much shorter during the night. During the day a typical value for P is 16 minutes. During the night P is typically about 4 minutes or less. When very short phase variations occur, i.e. with $P < 1$ minute, the interferometer amplitude also shows serious disturbances, indicating that the variations occur within one integration time. After a careful check of the observations and of instrumental behaviour we conclude that ionospheric irregularities with time scales of less than 60 seconds are a common feature during the night. Figures 3.3 and 3.4 indicate that the amplitudes of the phase variations for these fast irregularities are up to twice as large than those for the slower ones. The linear horizontal distance which the line of sight travels per minute at an altitude of about 350 km, is 1.5 km. The primary beam for the WSRT at 608.5 MHz has linear dimensions of 6 km at this height. Thus the spatial dimensions of these fast irregularities are of this order of magnitude or less.

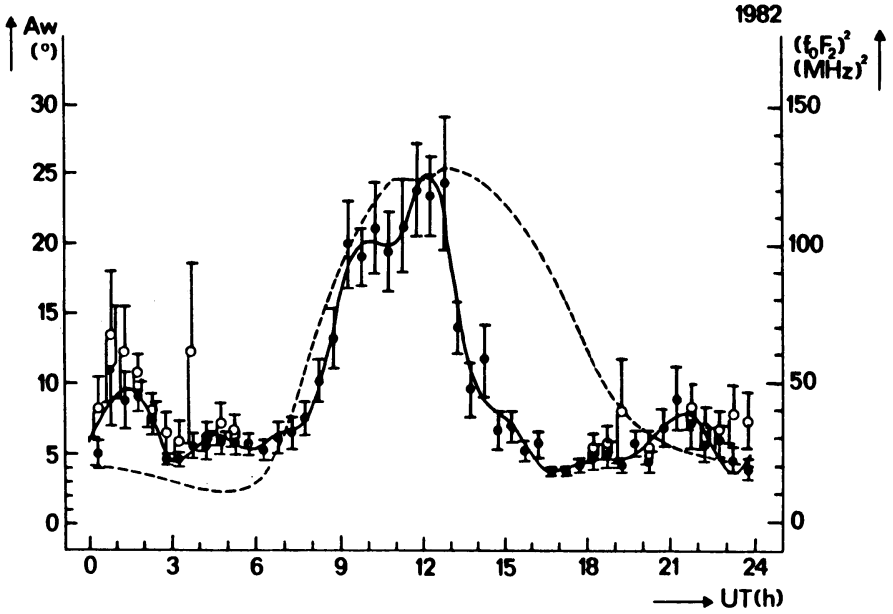

Figure 3.3 Variations in the amplitude of WSRT phase variations at 2.7 km baseline due to ionospheric irregularities as a function of time. Open symbols indicate that variations with periods less than 1 minute have been included. The dashed line represents the average $(foF2)^2$ variation with time. Local noon is at 11h34m UT.

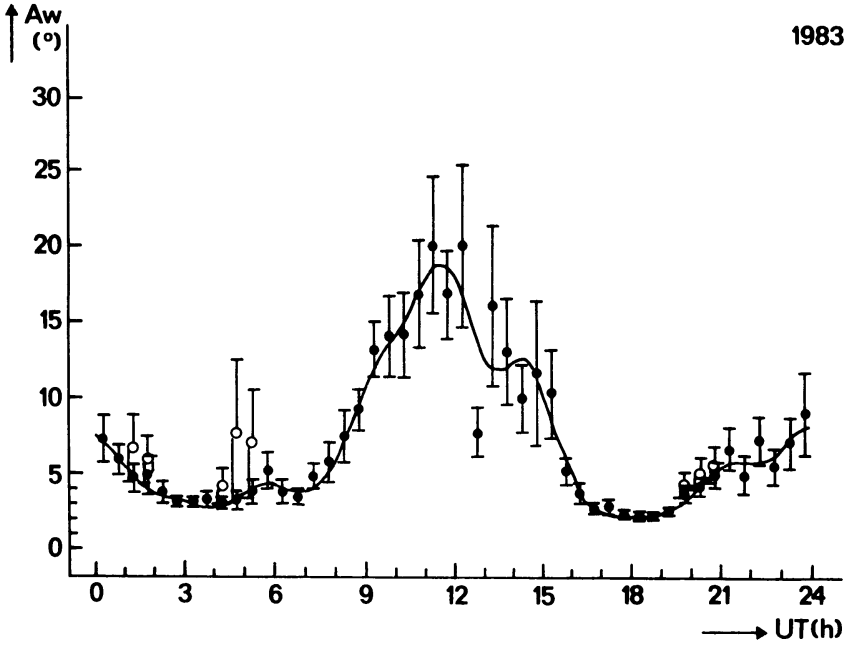

Figure 3.4 The mean amplitudes of the variation in phase differences during the first three months in 1983 as function of the time of the day.
The dots indicate data without short-time fluctuations (shorter than 1 minute), the open circles with these fluctuations included. The error bars are also given.

Differential Doppler observations of NNSS satellites in general show the same behaviour, although they are not completely comparable with the WSRT observations because the WSRT is mainly sensitive to the dependence on the east-west gradients of ionospheric parameters along the line of sight as a function of time. With the aid of satellites a latitudinal cross-section at discrete time intervals is obtained.

The most pronounced characteristics from the differential Doppler observations are:

(1) Nearly every day, waves occur during the daytime. They are observed most clearly south of our geographic latitude, because the radio path lies approximately in surfaces of constant gravity wave phase (Georges and Hooke, 1970; Davis, 1972; Sen Gupta and Nagpal, 1982).

(2) The amplitudes of the waves are a function of time and are larger in the morning than in the afternoon.

(3) Small-scale irregularities (with time scales < 4.6 sec, or spatial scales < 12 km), are present nearly every night north of our latitude. Their southern boundary varies with time and day.

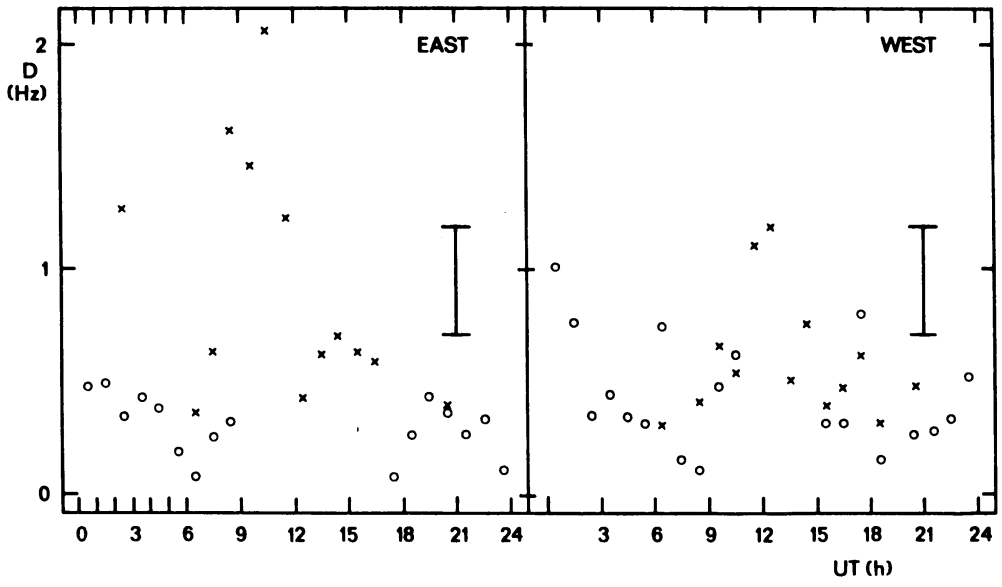

Figure 3.5 Hourly averages of the amplitudes of irregularities for satellite passage in the eastern as well as in the western sky. The crosses indicate irregularities with horizontal scales larger than 12 km. The circles the irregularities with shorter scales. The error bar indicates the mean spread per averaged data point.

Figure 3.5 presents the amplitudes of these waves as a function of time. The registrations for satellite passages east and west of the observing stations are shown separately. The amplitudes of the waves are about 50% larger in the eastern sky than in the western sky, which agrees with the results from the WSRT observations. The error bar in figure 3.5 indicates the spread in the data.

3.4. Mean diurnal behaviour of TID amplitudes

In order to compare the 1982 and 1983 data, the time-dependence of the amplitudes of the WSRT phase variations at 2.7 km baseline (as e.g. given in figure 3.4) was subjected to Fourier analysis. It is assumed that because of the large amount of data, observational biases depending on celestial coordinates are averaged out. With respect to the occurrence of TIDs, the scheduling of WSRT observations and of the satellite passages is random. The observations are spread over any part of the sky over the whole 24 hours of the day. We observed that the diurnal behaviour does not change significantly from one day to another. The total coverage of data as a function of time is not continuous,because of gaps in the schedule of the WSRT calibration observations and of the satellite passages. Therefore, we assumed that it is permissible to average all data in windows of 30 minutes to check the average diurnal behaviour, because TIDs occur as a function of time. Thus, all data were projected on a single day after averaging, and subjected to Fourier analyses (Smart, 1958, pp. 157f) in order to detect systematic variations of the amplitudes A as a function of time, i.e. periodicities in these variations. The results are shown in Table 3.1. In this table these amplitudes are represented by Aw. In Table 3.1 only those Fourier components that were above the mean errors of the solutions have been given. The results show a time-independent component, the zero level, and components with periods of 24 and 12h. These periodicities result primarily from the daytime observations. Taking errors into account, the 24h period has the same strength as the 12h period. This holds for 1982 as well as for 1983. The comparison between 1982 and 1983 shows that the amplitudes in 1983 are 20% lower than those of 1982. Within the errors the relative strengths of the 24 and 12h terms with respect to the zero level are equal. This means that the difference between the

diurnal variations of Aw in 1982 and 1983 is only a scaling factor. This factor, being the ratio of the amplitudes in 1983 to those in 1982, is 0.8 ± 0.1.

Table 3.1

Periodicities in TID amplitudes and foF2 data

year	data	units	zero level	24h period	12h period
1982	Aw	degrees	9.1 ± 0.4	5.8 ± 0.6	5.2 ± 0.6
	phase	degrees	-	347.1 ± 6.2	347.1 ± 6.9
	As	Hz	0.4 ± 0.1	0.5 ± 0.1	-
	phase	degrees	-	340.5 ± 12.9	-
	foF2	MHz	7.1 ± 0.1	3.9 ± 0.1	0.6 ± 0.1
	phase	degrees	-	32.3 ± 2.8	347.1 ± 10.1
1983	Aw	degrees	7.3 ± 0.3	4.4 ± 0.4	4.6 ± 0.4
	phase	degrees	-	357.5 ± 4.9	354.9 ± 4.7
	As	Hz	1.2 ± 0.2	0.9 ± 0.2	-
	phase	degrees	-	334.0 ± 20.9	-
	foF2	MHz	5.3 ± 0.2	3.4 ± 0.2	1.0 ± 0.2
	phase	degrees	-	24.9 ± 4.1	26.3 ± 14.4

We determined the phase of the different periods with respect to the meridian transit of the sun. These phases are also given in Table 3.1. Again, taking errors into account, the behaviour in 1982 and 1983 was identical.

The same analysis was done for the amplitudes of the TIDs detected by differential Doppler measurements and indicated by As. The data show only one specific period: 24 hours, see Table 3.1. Perhaps this is due to the fact that during the night no TIDs were observed and moreover the spread in the data is larger for the satellite observations than for the WSRT measurements. The phase with respect to the solar meridian transit is, taking errors into account, the same as that determined from WSRT observations.

In order to check for any relation between the amplitudes of medium scale TIDs and the variation in the total electron content we looked at the behaviour of foF2, the critical frequency of the F2 layer, which is roughly a measure of the total electron content. The same Fourier analysis was performed on the foF2 data of the observing periods. The results, given in Table 3.1 indicate that the time-dependence of foF2 differs from that of the Aw data. Although, apart from a zero level, a 24h period and a 12h period were detected, the relative strengths of these two periods are far from equal. Furthermore, the foF2 dependencies with respect to solar meridian transit are several tens of degrees out of phase with the equivalent WSRT data. We therefore conclude that the diurnal variations of medium-scale TID amplitudes show another time-dependency than that of the total electron content of the ionosphere.

3.5. Discussion

The medium scale TIDs reflect variations that are typically about 5% of the total electron content. The horizontal linear dimensions are between about 100 and 800 km. It is clear that the waves propagate neither perfectly east-west nor north-south. It may be that the suspected preference for one of these directions is based on observational selection effects: e.g. Bougeret (1981) made his observations mainly with an east-west radio interferometer and therefore his sensitivity to north-south irregularities was minimal.
There is controversy about the sources of the ionospheric irregularities, see

for a review e.g. Jones (1982). There is evidence for auroral generation see e.g. Hunsucker (1982). Bertin et al. (1975, 1978) suggest a tropospheric origin, i.e. instability of jet streams. They obtained the location of the sources by inverse ray tracing. On the basis of differential Doppler measurements Sizun et al. (1981) argue that single point-like sources are not sufficient to explain the irregularities. They prefer extended sources that are located at higher latitudes, rather than the local source suggested by Bertin et al. (1975, 1978). Herron and Donn (1973) using an array of continuous-wave Doppler sounders show that the direction of propagation of TIDs changes with local time: i.e. during the morning hours the preferred direction is south-east, at noon it is south, in the afternoon it is south-west. This pattern agrees with the observations made by Bertin et al. (1978). This effect might reflect directional filtering by the neutral wind. Herron and Donn (1973) and Bertin et al. (1978) made their observations mainly during daytime. Morton and Essex (1978) analysed gravity wave observations made at a southern hemisphere mid-latitude station. They also determined the speeds, azimuths, periods and times of occurrence of these disturbances. Although they tried to interpret their results in terms of directional coupling by neutral winds, their results for daytime observations basically show scatter diagrams only. However, night-time results seem to show some support for the hypothesis of directional coupling by neutral winds.

If the lines of sight of the WSRT and the Doppler receiver are aligned in the same direction in the sky, the observations can also be used to calculate the azimuth and speed of the observed TIDs (Kelder & Spoelstra, 1986). The calculations show that the azimuths of the direction towards which the irregularities propagate are mainly south - south-west. This differs from the observations made by Bertin et al. (1978) and by Herron and Donn (1973) and Mercier (1986). The reason for this difference is not necessarily seasonal, since the observations of Bertin et al. were made during the summer and those of Herron and Donn during the winter. Mercier's difference is maybe due to differences in data reduction but this will be pursued in the near future. In the current literature, different sources for medium-scale TIDs are mentioned. With respect to these sources a distinction must be made between discrete events and regular sources. On the basis of the observational data, the following must be explained:

a. the occurrence of TIDs is a regular phenomenon, since they can be observed continuously in any direction in the sky;
b. the amplitudes of the TIDs show 12- and 24-hour components of equal strength in the Fourier analysis (see Table 3.1);
c. TIDs occur worldwide.

The occurrence of 12h and 24h periods of equal strength in the data is, however, typical of upper atmospheric tides (Chapman and Lindzen, 1970). Our hypothesis is that atmospheric tides generate gravity waves continuously. Experimental evidence of a relationship between the intensity of internal gravity waves and tides has already been presented by Gavrilov et al. (1981). A possible mechanism could be as follows. Tidal waves, mainly excited in the lower atmosphere i.e. the stratosphere and troposphere, grow with height z as exp(z/2H), where H is the scale height of the atmosphere. Between 80 and 120 km altitude they can reach large amplitudes in the order of 150 m/s. At this height they travel through the coldest part of the atmosphere. Here are large temperature gradients and low values of the Brunt-Väisälä frequency. The Richardson number corresponding with the tidal field can become smaller than 0.25. Then, tidal waves become unstable and can generate internal gravity waves with horizontal phase velocities in the order of 150 m/s. For the theory of hydrodynamic instability see e.g. Drazin & Reid (1982).

The assumption that tides generate internal gravity waves can explain a, b and c (see above).
The consequences of the tidal generation of internal gravity waves are:
i. these internal gravity waves should show a latitudinal dependence;
ii. the TID parameters should be dependent on the season;
iii. Variation of the mesopause temperature profile shall modulate with the same period the internal wave amplitudes. This effect may be partly responsible for the observed spread in the amplitudes Aw and As.

The assumption that internal gravity waves are excited between 80 and 120 km does not conflict with the validity of the filtering mechanism of thermospheric tidal winds as proposed by e.g. Kalikhman (1978, 1980) and Waldock & Jones (1986). Both mechanisms should show up as a function of time of day. The conclusion derived by Bertin et al. (1978) that sources of gravity waves are located in the troposphere does not conflict with the hypothesis of

tides as sources. Instabilities in the troposphere will certainly generate gravity waves, but the troposphere may not be the regular source needed to explain the daily occurrence of such waves.

Most of this chapter has been published previously: Kelder & Spoelstra, 1984 a,b, 1986, and Spoelstra & Kelder, 1984.

3.6. References

Bertin, F., Testud, J. & Kersley, L., 1975 - Medium scale gravity waves in the ionospheric F-region and their possible origin in weather disturbances. Planet. Space Sci. 23, 493-507.

Bertin, F., Testud, J., Kersley, L. & Rees, P.R., 1978 - The meteorological jet stream as a source of medium scale gravity waves in the thermosphere: an experimental study. J. atmos. terr. Phys. 40, 1161-1179.

Bougeret, J.L., 1981 - Some effects produced by the ionosphere on radio interferometry: Fluctuations in apparent source position and image distortion. Astron. Astroph. 96, 259-271.

Chapman, S. & Lindzen R.S., 1970 -Atmospheric Tides. Reidel, Dordrecht.

Davis, M.J., 1972 - The integrated ionospheric response to internal atmospheric gravity waves. J. atm. terr. Phys. 35, 929-959.

Drazin, P.G.& Reid W.H., 1982 - Hydrodynamic Stability. Cambridge University Press.

Evans, J.V., 1977 - Satellite beacon contributions to studies of the structure of the ionosphere. Rev. Geophys. Space Phys. 15, 325.

Gavrilov, N.M., Kal'chenko, B.V., Kashcheyev, B.L. & Shved, G.M., 1981 - Discovery of a Relationship between the Intensity of Internal Gravity Waves in the Upper Atmosphere and the Tidal Phase. Izvestiya, Atmospheric and Oceanic Physics 17, 499-505.

Georges, T.M. & Hooke, W.H., 1970 - Wave-induced Fluctuations in Ionospheric Electron content: A Model Indicating Some Observational Biases. J. Geophys. Res. 75, 6295-6308.

Hamaker, J.P., 1978 - Atmospheric delay fluctuations with scale sizes greater than one kilometer, observed with a radio interferometer array. Radio Sci. 13, 873-885.

Herron, T.J. & Donn, W.L., 1973 - Diurnal variations of F-region waves. J. atm. terr. Phys. 35, 2163-2176.

Hines, C.O., 1960 - Internal Atmospheric Gravity Waves at Ionospheric Heights. Can. J. Phys. 38, 1441-1481.

Hines, C.O., 1974 - The Upper Atmosphere In Motion. Geophysical Monograph 18, American Geophysical Union.

Hunsucker, R.D., 1982 - Atmospheric Gravity Waves Generated in the High-Latitude Ionosphere: A Review. Rev. Geophys. Space Phys. 20, 293-315.

Jones, T.B., 1982 - Generation and propagation of acoustic gravity waves. Nature 299, 488-489.

Kalikhman, A.D., 1978 - Experimental proof of Directional Filtering of Medium-Scale Traveling Ionospheric Disturbances. Geomagnetism and Aeronomy 18, 238-239.

Kalikhman, A.D., 1980 - Medium-scale travelling ionospheric disturbances and thermospheric wind in the F-region. J. atm. terr. Phys. 42, 697-703.

Kato, S, 1980 - Dynamics of the upper atmosphere. Reidel, Dordrecht.

Kelder, H. & Spoelstra, T.A.Th., 1984 a - Multi-technique study of medium scale TID's. Kleinheubacher Berichte 27, 575-584.

Kelder, H. & Spoelstra, T.A.Th., 1984 b - Multi-technique study of ionospheric irregularities. Proc. Beacon Satellite Studies of the Earth's Environment. New Delhi, India. pp. 457-463.

Kelder, H. & Spoelstra, T.A.Th., 1986 - Medium scale TID's observed by radio interferometry and differential Doppler techniques. Accepted by J. atm. terr. Phys.

Leitinger, R., Schmidt, G. & Taurianen, A., 1975 - An evaluation method combining the Differential Doppler measurements from two stations that enables the calculation of the electron content of the ionosphere. J. Geophysics 41, 201-213.

Leitinger, R. & Putz, E., 1978 - Die Auswertung von Differenz-Doppler-Messungen and den Signalen von Navigationssatelliten. Technischer Bericht, Institut f. Meteorologie u. Geophysik, Universität Graz.

Mercier, C., 1986 - Observations of atmospheric gravity waves by radiointerferometry. J. atm. terr. Phys. 48, 605-624.

Morton, F.W. & Essex, E.A., 1978 - Gravity wave observations at a southern hemisphere mid-latitude station using the total electron content technique. J. atm. terr. Phys. 40, 1113-1122.

Sen Gupta, A. & Nagpal, O.P., 1982 - Interaction of medium scale gravity waves with ionization. Its application to electron content measurements of TID's. Indian J. Radio and Space Phys. 11, 45-59.

Sizun, H.L., Bertel, G. & De Maindreville, C., 1981 - Sur la localisation et l'étendue des sources des ondes de gravité observées dans la thermosphère. Ann. Géophys. 37, 423-434.

Smart, W.M., 1958 - Combination of observations, Cambridge University Press.

Spoelstra, T.A.Th., 1983 - The influence of ionospheric refraction on radio astronomy interferometry. Astron. Astroph. 120, 313-318.

Spoelstra, T.A.Th. & Kelder, H., 1984 - Effects produced by the ionosphere on radio interferometry, Radio Sci. 19, 779-788.

Stolp, E.R., 1985 - Voortplanting van zwaartegolven in de atmosfeer en rond een kritieke laag. Scientific Report W.R. 85-1, KNMI, Netherlands.

Waldock, J.A. & Jones, T.B., 1986 - HF Doppler observations of medium-scale ionospheric disturbances at mid-latitudes. J. atm. terr. Phys. 48, 245-260.

CHAPTER 4

CRITICAL LEVELS FOR INTERNAL GRAVITY WAVES IN A JET TYPE FLOW

4.1. Introduction

In their classic paper, Booker & Bretherton (1967) analysed the propagation of an internal gravity wave through a height-dependent wind field containing one critical level, i.e. a level where the wind velocity equals the horizontal phase velocity of the wave. They showed that the transmission of the wave depends only on the value of the Richardson number at the critical level. The Richardson number is defined as the ratio between the square of the Brunt-Väisälä frequency and the square of the vertical gradient of the wind. It is a measure for the dynamic stability of the wind field. They considered only background flows with a Richardson number larger than 0.25. Later on, Jones (1968) found that, for low values of the Richardson number at the critical level, overreflection, i.e. the absolute value of the reflection coefficient larger than 1, occurs.
These results have been confirmed by various authors and for different background flows, e.g. a broken-line profile (Eltayeb & McKenzie, 1975) and a hyperbolic-tangent profile (Van Duin & Kelder, 1982).
Viscosity and thermal conduction were introduced by Hazel (1967). He showed that for values of the Richardson number larger than 0.25 a large amount of wave energy is lost near the critical level. Not known is the fraction of energy lost by dissipation or by excitation of other modes. The transmission coefficient is the same as found by Booker & Bretherton in the dissipationless model. It was shown by Van Duin & Kelder (1986) that for all values of the Richardson number the reflection and transmission coefficients remain approximately unchanged at the limit of small molecular viscosity and thermal conduction. Geller, Tanaka & Fritts (1975) and Fritts & Geller (1976) studied instability in the vicinity of the critical level. They found that, in the linear theory, viscosity and heat conduction can play a stabilizing role with respect to convective instability. A numerical model was used by Fritts (1978,

1982) to compare the effects of viscosity, time dependence and nonlinear interaction. Time dependence was found to play only a minor role in stabilizing the critical level. Nonlinear effects can give rise to higher harmonics of the forcing wave, which develop large amplitudes near the critical level when viscous effects are small.

A nonlinear non-dissipative treatment was given by Brown & Stewartson (1980, 1982 a, b). They showed that for large values of the Richardson number the linear model is valid up to a certain time inversely proportional to the wave amplitude. After that time, the reflection and transmission coefficients change. Another very different approach to the nonlinear stationary problem was put forward by Teitelbaum & Sidi (1979). They showed that a contact discontinuity appears below the critical level in the absence of dissipative processes.

To summarize: nonlinearities might change the results of linear models to an extent that is neither theoretically well understood nor experimentally documented. In agreement with Lindzen (1973), we believe that, if dissipative damping occurs before amplitudes have grown to the point where nonlinear effects become important, the linear approximation is a good one.

The problem of one critical level has been extensively treated by Rosenthal & Lindzen (1983 a, b) and Lindzen & Rosenthal (1983) with regard to instabilities and the relation between instabilities and overreflection.

Within the framework of a linear and inviscid model with one critical layer Grimshaw (1980) added the effects of rotation and electrical conductibility. Although he mentioned the problem of two or more critical levels, he does not study this.

Propagation through background flows containing more than one critical level has attracted little interest. Drazin, Zaturska & Banks (1979) have done a calculation for a flow containing two critical levels. They modelled the flow by a broken-line profile. They showed that for large values of the Richardson number the transmission coefficient of the whole layer can be found by considering the transmission coefficients of the two critical levels independently, attributing to those two levels the same value of the Richardson number.

In this chapter propagation through two critical levels in a jet-type background flow is studied. Jet-type background winds are frequently observed in the atmosphere. Moreover short-period gravity waves see planetary and tidal

waves as stationary jets due to the large difference in the period and in the phase velocity.
As the mathematical treatment is more difficult than in the case of one critical level, we restrict ourselves to the linear non-dissipative case.
We have taken a symmetric jet-type background flow. This case can be solved analytically. The reflection and transmission coefficients are determined. The influence of the distances between the two critical levels is considered. Also, various values of the Richardson number are taken. The reflection and transmission coefficients are also calculated using a numerical approach, which gives the same results as the analytical one.
A large part of this chapter has been published previously (Teitelbaum & Kelder, 1985).

4.2. On the hydrostatic and the Boussinesq approximations

In the literature a lot can be found on approximations see e.g. Holton (1972) and Gill (1982). Here some elementary concepts are repeated in a rather concise form.
In the hydrostatic approximation the pressure at any point is approximated by the weight of a unit cross-section column of air above that point.
In formula

$$\frac{\partial p}{\partial z} = -\rho g \ . \tag{4.1}$$

The vertical momentum equation states that

$$\frac{dw}{dt} = -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \ . \tag{4.2}$$

The hydrostatic approximation is therefore equivalent to neglecting the vertical acceleration in comparison with the vertical pressure variation divided by the density and the acceleration of gravity. This will be further elucidated hereafter.
A scale analysis of atmospheric gravity waves leads to the following characteristic numbers:

Table 4.1

Scale analysis of atmospheric gravity waves

horizontal velocity scale U	10^2 m/s
vertical velocity scale W	10^{-1} m/s
horizontal length scale L_h	10^5 m
vertical length scale L_z	10^2 m
pressure scale p	10^3 Pa
time scale T	L_h/U 10^3 s
density scale ρ	10^{-2} kg m^{-3}

Define a standard pressure, $p_o(z)$, which is the average of the pressure over a large domain in time and in the horizontal plane. The corresponding standard density $\rho_o(z)$ is such that $p_o(z)$ and $\rho_o(z)$ are in exact hydrostatic balance:

$$\frac{1}{\rho_o}\frac{dp_o}{dz} = -g.$$

The total pressure and density fields may be written as

$$p(x,y,z,t) = p_o(z) + p'(x,y,z,t) ,$$

$$\rho(x,y,z,t) = \rho_o(z) + \rho'(x,y,z,t) .$$

where p' and ρ' are perturbations from the standard values of pressure and density. Substituting these expressions in the vertical momentum equation (4.2) and assuming that $\frac{\rho'}{\rho_o}$ and $\frac{p'}{p_o}$ are much less than unity in magnitude we obtain that

$$-\frac{1}{\rho}\frac{\partial p}{\partial z} - g \simeq -\frac{1}{\rho_o}\left[\rho' g + \frac{\partial p'}{\partial z}\right] .$$

Using the numbers from Table 4.1 we find that

$$\frac{1}{\rho_o}\frac{\partial p'}{\partial z} \sim 10^{-1}\ m/s^2 \quad , \quad \frac{\rho' g}{\rho_o} \sim 10^{-1}\ m/s^2$$

and

$$\frac{dw}{dt} \simeq \frac{W}{T} \sim 10^{-4}\ m/s^2.$$

We see that to a very good approximation the perturbation pressure field is in hydrostatic equilibrium with the perturbation density field such that

$$\frac{\partial p'}{\partial z} + \rho' g = 0.$$

In order to clarify further the validity of the hydrostatic and also the more commonly used Boussinesq approximation we consider the perturbations in a windless isothermal atmosphere in more detail. Assume again a state with an unperturbed pressure $p_o(z)$. Hydrostatic equilibrium is supposed hence

$$\frac{dp_o}{dz} = -\rho_o g. \tag{4.3}$$

Further it is assumed that $\frac{d}{dz} \ln \rho_o = -\frac{1}{H}$, where H is a constant scale height. This assumption is justified if the background temperature is not dependent of height.

If the perturbations are small enough the equations may be linearized and we get the following set (it is understood that quantities without indices are first order perturbations). Without loss of generality it is assumed that there is no y dependency.

$$\rho_o \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0, \tag{4.4}$$

$$\rho_o \frac{\partial w}{\partial t} + \rho g + \frac{\partial p}{\partial z} = 0, \tag{4.5}$$

$$\frac{\partial \rho}{\partial t} + w\frac{d\rho_o}{dz} + \rho_o \left(\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z}\right) = 0, \tag{4.6}$$

$$\frac{\partial p}{\partial t} + w\frac{dp_o}{dz} = c^2 \left(\frac{\partial \rho}{\partial t} + w\frac{d\rho_o}{dz}\right). \tag{4.7}$$

Assume a perfect gas, then the velocity of sound c equals $\sqrt{\frac{\gamma RT}{m}}$, where γ denotes the ratio of the specific heats $\gamma = c_p/c_v$, R is the gas constant and m the molecular weight.
With the help of (4.3) and (4.6) the equation (4.7) can be written as

$$\frac{\partial p}{\partial t} - \rho_o g w + \rho_o c^2 \left(\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z}\right) = 0 \tag{4.8}$$

The density perturbation ρ can be eliminated from this set of equations by using (4.5) resulting in:

$$\rho_o \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0 , \tag{4.9}$$

$$\frac{\partial p}{\partial t} - \rho_o g w + \rho_o c^2 \frac{\partial u}{\partial x} + \rho_o c^2 \frac{\partial w}{\partial z} = 0 , \tag{4.10}$$

$$\frac{\rho_o}{g} \frac{\partial^2 w}{\partial t^2} + \frac{1}{g} \frac{\partial^2 p}{\partial z \partial t} + \rho_o \frac{w}{H} - \rho_o \frac{\partial u}{\partial x} - \rho_o \frac{\partial w}{\partial z} = 0 . \tag{4.11}$$

Eliminating the horizontal velocity yields:

$$\frac{c^2 \rho_o}{g} \frac{\partial^2 w}{\partial t^2} + \rho_o (\gamma-1) g w + \frac{c^2}{g} \frac{\partial^2 p}{\partial z \partial t} + \frac{\partial p}{\partial t} = 0 , \tag{4.12}$$

$$\left(\frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2}\right) p - \rho_o \left(g \frac{\partial w}{\partial t} - c^2 \frac{\partial^2 w}{\partial z \partial t}\right) = 0 , \tag{4.13}$$

using that in a perfect gas $H = \frac{c^2}{\gamma g}$.

Eliminating the pressure perturbation p from the equations (4.12) and (4.13) a partial differential equation is obtained for the vertical velocity w

$$\frac{\partial^2}{\partial t^2} \left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2}\right)\right) w - \omega_B^2 \frac{\partial^2 w}{\partial x^2} + \frac{\partial^3 w}{\partial z \partial t^2} = 0, \tag{4.14}$$

where the Brunt-Väisälä frequency ω_B is defined as $\omega_B = \sqrt{\frac{(\gamma-1) g^2}{c^2}}$.

Define $\hat{w}$ by $w(x,z,t) = e^{z/2H} \hat{w}(x,z,t)$ then the equation for $\hat{w}$ reads

$$\frac{1}{c^2} \left(\frac{\partial^2}{\partial t^2} + \omega_a^2\right) \frac{\partial^2 \hat{w}}{\partial t^2} - \left(\frac{\partial^2}{\partial t^2} + \omega_B^2\right) \frac{\partial^2 \hat{w}}{\partial x^2} - \frac{\partial^4 \hat{w}}{\partial t^2 \partial z^2} = 0, \tag{4.15}$$

where $\omega_a^2 = \frac{\gamma^2 g^2}{4c^2}$.

The coefficients in equation (4.15) are constants hence solutions can be constructed with a normal mode analysis, i.e. suppose the form $e^{i(\omega t - kx - lz)}$, where ω is the frequency and k and l are the horizontal and vertical wave numbers respectively. From (4.15) the following dispersion relation is obtained

$$\frac{1}{c^2}(\omega_a^2 - \omega^2)\omega^2 + (\omega_B^2 - \omega^2)k^2 - \omega^2 l^2 = 0. \qquad (4.16)$$

A dispersion diagram is drawn in figure 4.1. An acoustic and gravity branch can be distinguished.

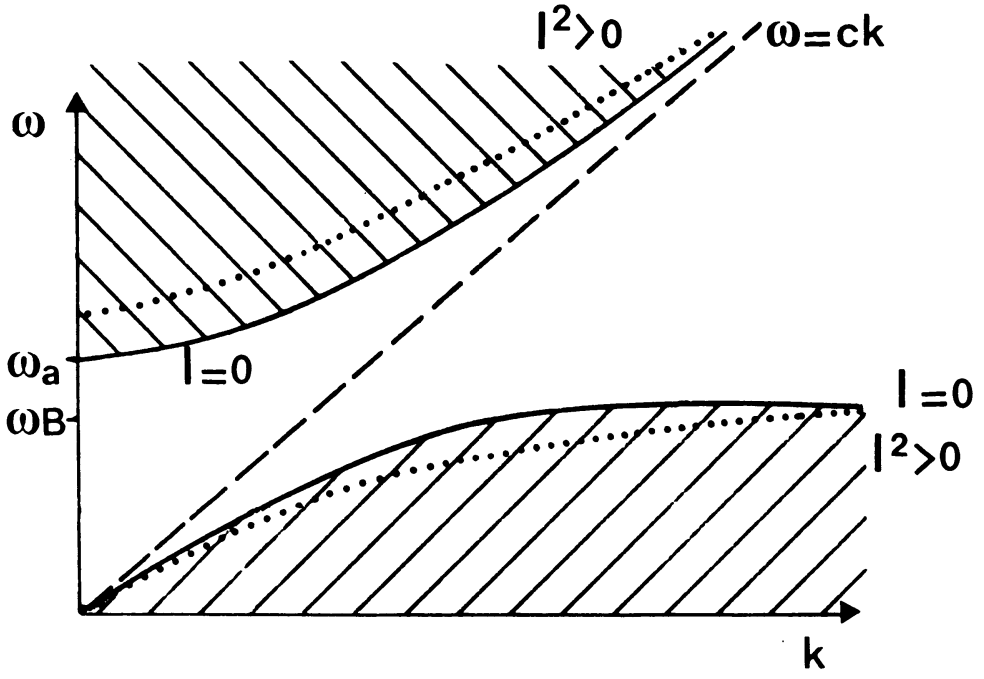

Figure 4.1 Dispersion diagram for constant vertical wavenumber l.

Assume now that the hydrostatic approximation is valid: i.e. the term $\rho_o \frac{\partial w}{\partial t}$ in the vertical momentum equation (4.5) is ignored with respect to the two other terms. Retracing the derivation of (4.14) with this in mind one ends up with

$$\frac{\partial^2}{\partial t^2}\left(-\frac{\partial^2}{\partial z^2}\right) w - \omega_B^2 \frac{\partial^2 w}{\partial x^2} + \frac{1}{H}\frac{\partial^3 w}{\partial z \partial t^2} = 0. \qquad (4.17)$$

Again the exponential growth of the amplitude can be split off i.e.

$$w(x,z,t) = e^{z/2H}\,\hat{w}(x,z,t).$$

The equation for $\hat{w}$ reads

$$\frac{1}{c^2}\,\omega_a^2\,\frac{\partial^2\hat{w}}{\partial t^2} - \omega_B^2\,\frac{\partial^2\hat{w}}{\partial x^2} - \frac{\partial^4\hat{w}}{\partial t^2\partial z^2} = 0 \qquad (4.18)$$

The hydrostatic wave equation (4.18) in comparison with wave equation (4.15) is lacking two terms i.e. $\frac{1}{c^2}\frac{\partial^4 w}{\partial t^4}$ and $\frac{\partial^4 w}{\partial t^2\partial x^2}$. Comparing (4.18) and (4.15) it is clear that the hydrostatic approximation is valid if

$$\frac{\partial^2}{\partial t^2} \ll \omega_a^2, \quad \text{and} \quad \frac{\partial^2}{\partial t^2} \ll \omega_B^2. \qquad (4.19)$$

Hence within the hydrostatic approximation only phenomena with time scales which are large with respect to $2\pi/\omega_a$ and $2\pi/\omega_B$ can be described. The hydrostatic approximation is essentially a low-frequency approximation.
In order to gain some insight into the values of ω_a and ω_B the dependency of these frequencies with height is drawn in figure 4.2.

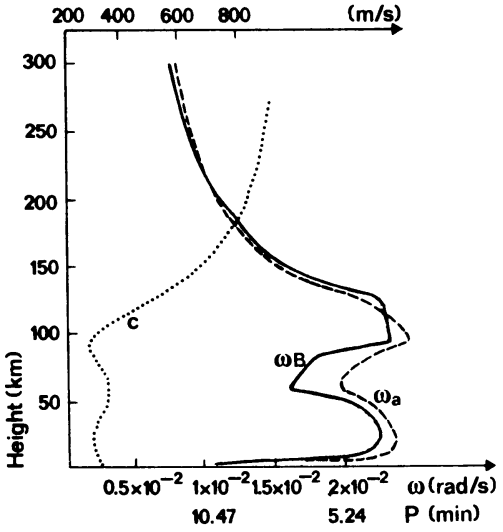

Figure 4.2 Height profiles of the velocity of sound c, the Brunt-Väisälä frequency ω_B and the acoustic cut-off frequency ω_a in a realistic atmosphere (after Tolstoy and Pan, 1970).

Remark that in the derivation of the equations ω_a and ω_B are assumed to be constant. For the lower atmosphere values of $\frac{2\pi}{\omega_a} \sim 270$ sec and $\frac{2\pi}{\omega_B} \sim 300$ sec are reasonable.

The hydrostatic dispersion relation is obtained by assuming in equation (4.18) a normal mode solution i.e. $\hat{w}(x,z,t) = e^{i(\omega t - kx - lz)}$:

$$\omega^2 \frac{\omega_a^2}{c^2} + \omega^2 l^2 - \omega_B^2 k^2 = 0. \qquad (4.20)$$

Figure 4.3 gives the dispersion diagram corresponding to (4.20). Notice that the acoustic branch is suppressed and that evanescent waves can be described within the hydrostatic approximation if $\frac{\omega_B^2}{\omega^2} k^2 < \frac{\omega_a^2}{c^2}$.

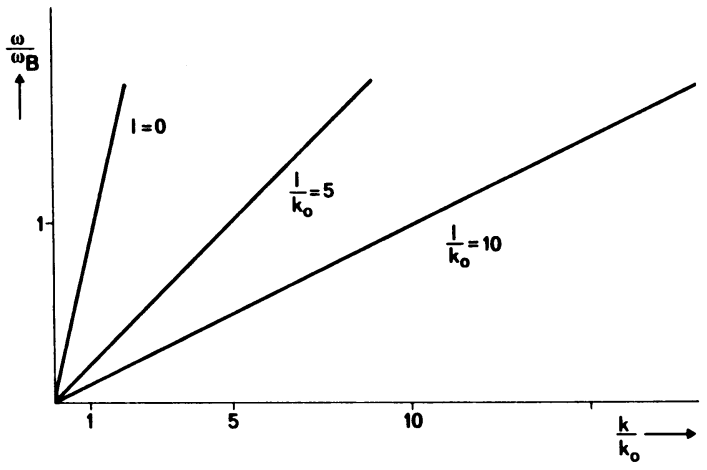

Figure 4.3 Dispersion diagram for constant l. Hydrostatic approximation applied.

In the Oberbeck-Boussinesq approximation (Oberbeck, 1879; Boussinesq, 1903) the fluid motion is assumed to be incompressible and the background density is treated as constant except where it is coupled to gravity in the buoyancy term

of the vertical momentum equation. Retracing the derivation of the wave equation (4.15) with this in mind we end up with

$$\frac{\partial^2}{\partial t^2}\left(-\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}\right)\right) w - \omega_B'^2 \frac{\partial^2 w}{\partial x^2} = 0, \tag{4.21}$$

where $\omega_B'^2 = \frac{\gamma g^2}{c^2}$.

In comparison with wave equation (4.15) the terms $\frac{1}{c^2}\frac{\partial^4 w}{\partial t^4}$ and $\frac{\partial^4 w}{\partial t^2 \partial x^2}$ are lacking and moreover no exponential growth has appeared.
The Oberbeck-Boussinesq approximation is valid if:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2} \ll \frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2},$$

and (4.22)

$$\frac{\omega_a^2}{c^2}\frac{\partial^2}{\partial t^2} \ll \omega_B'^2 \frac{\partial^2}{\partial x^2}.$$

From (4.22) it is clear that the Oberbeck-Boussinesq approximation is valid if the phase velocity is much smaller than the velocity of sound i.e.

$$\frac{\omega^2}{k^2} \ll c^2, \tag{4.23}$$

using that $\frac{\omega_a^2}{\omega_B'^2} = \frac{\gamma}{4} \sim O(1)$.

Hence the Oberbeck-Boussinesq approximation is essentially a low phase velocity approximation.
Substituting a normal mode solution i.e. $e^{i(\omega t - kx - lz)}$ in (4.21) the resulting dispersion solution reads

$$(\omega_B'^2 - \omega^2)\, k^2 - \omega^2 l^2 = 0. \tag{4.24}$$

From (4.16) it is clear that this Oberbeck-Boussinesq dispersion relation can also be obtained by taking the limit of the velocity of sound c tends to infinity. Note that in the Oberbeck-Boussinesq approximation evanescent waves can be described if $\omega > \omega_B'$. Figure 4.4 gives the dispersion diagram corresponding to (4.24). The Oberbeck-Boussinesq approximation suppresses the acoustic branch, changes considerably the value of Brunt-Väisälä frequency,

and finally neglects completely the exponential growth of the waves with height.

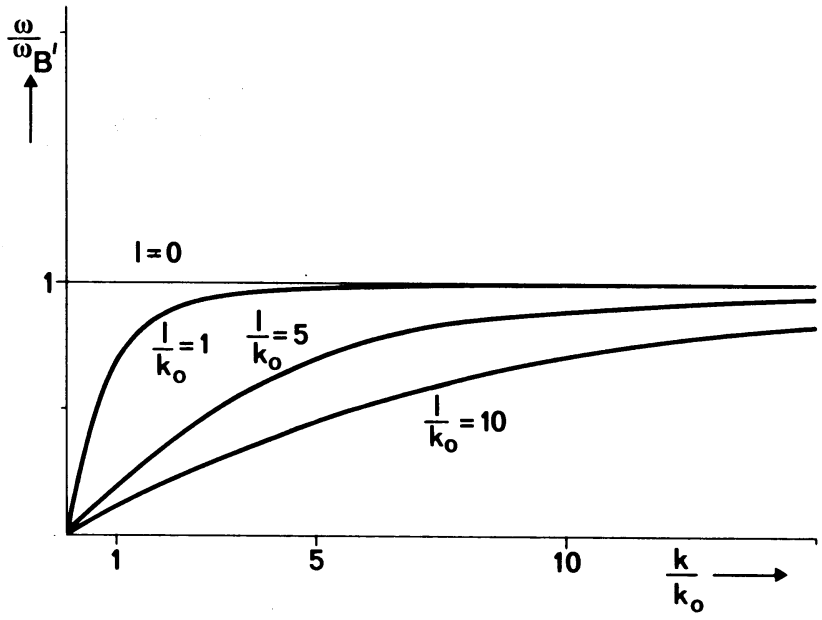

Figure 4.4 Dispersion diagram for the Boussinesq approximation.

4.3. On log-pressure coordinates

Pressure and log-pressure coordinates are treated in different textbooks see e.g. Holton, 1972 and Gill, 1982. For reasons of self-consistency we have gathered here some elementary concepts.

In the hydrostatic approximation it is sometimes advantageous to replace the z coordinate with another variable. In meteorology, pressure p is most commonly used and the variables are then known as isobaric coordinates. The advantage is that the density disappears from the equations. In log isobaric coordinates the logarithm of the pressure is used and the new vertical coordinate is then defined as

$$z^* := - H \ln \frac{p}{p_r} , \qquad (4.25)$$

where p_r is a standard reference pressure (usually taken to be 100 kPa), $H := \frac{RT_0}{g}$ is a constant scale height and R is the gas constant for dry air. In log isobaric coordinates, at variance with isobaric coordinates, the static stability parameter is almost constant with height.

4.3.1. The momentum equation

In the hydrostatic approximation only the horizontal momentum equation contains direct information on the velocity. That is

$$\rho \frac{du}{dt} + \frac{\partial p}{\partial x} = 0 \ . \tag{4.26}$$

In order to rewrite this equation it is useful to refer to figure 4.5 to elucidate the derivation

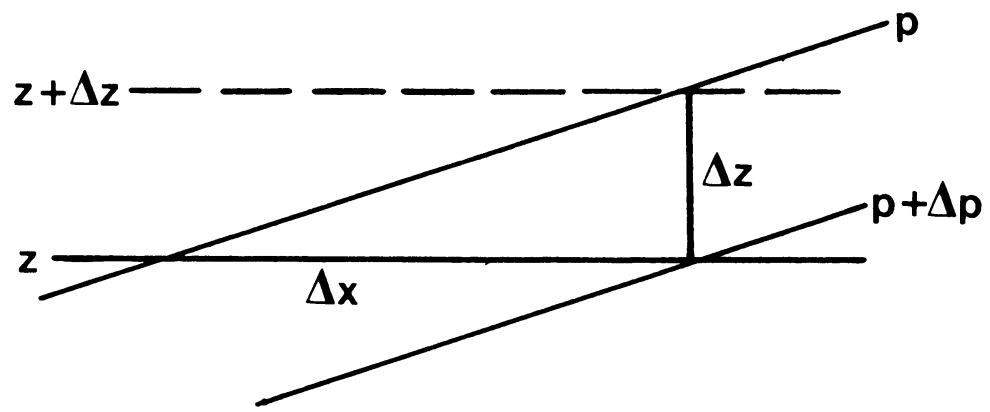

Figure 4.5 The relation between horizontal variations of pressure at a fixed level z and horizontal variations of level (or geopotential $\phi = gz$) at a fixed pressure. Two neighbouring pressure surfaces in the x,z plane such that the change in pressure in a horizontal distance Δx is Δp .

$$\left(\frac{\partial p}{\partial x}\right)_z = -\frac{\partial p}{\partial z}\left(\frac{\partial z}{\partial x}\right)_p = \langle \text{hydrostatic approximation} \rangle = \rho g \left(\frac{\partial z}{\partial x}\right)_p$$

The geopotential Φ is defined as the work required to raise unit mass from the surface of the earth to height z:

$$\Phi := \int_0^z g \, dz' \, . \tag{4.27}$$

The relation between variations in the pressure and the geopotential is illustrated in figure 4.6.

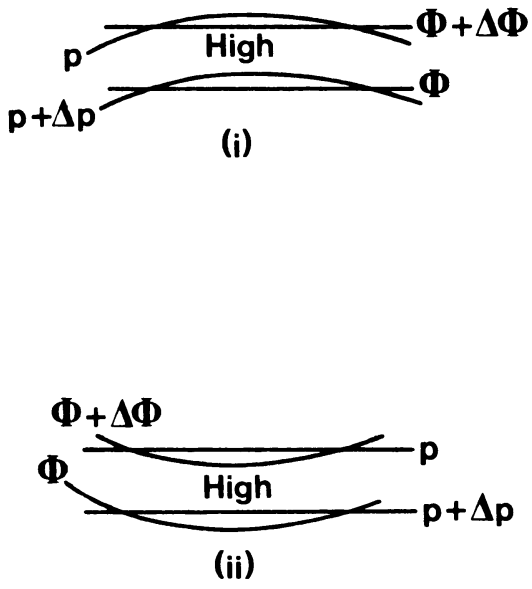

Figure 4.6 The relation between horizontal variations of pressure at a fixed level z and horizontal variations of level (or geopotential $\phi = gz$) at a fixed pressure. Representation of a high-pressure region in the x,z or x,ϕ plane (i).
In (ii) the same situation is redrawn in the x,p plane and the feature appears as a high in geopotential on an isobaric surface.

The horizontal pressure variation can be rewritten as

$$\left(\frac{\partial p}{\partial x}\right)_z = \rho g \left(\frac{\partial z}{\partial x}\right)_p = \rho \left(\frac{\partial \Phi}{\partial x}\right)_p = \rho \left(\frac{\partial \Phi}{\partial x}\right)_{z^*} .$$

The horizontal momentum equation can now be written as

$$\frac{du}{dt} + \left(\frac{\partial \Phi}{\partial x}\right)_{z^*} = 0 \ . \tag{4.28}$$

Suppose a horizontal background flow u in the x direction and a vertical velocity w^* defined by $w^* := \frac{dz^*}{dt}$.
The operator $\frac{d}{dt}$ now has the following form

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + w^* \frac{\partial}{\partial z^*} \ . \tag{4.29}$$

The horizontal momentum equation can then be expressed as:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + w^* \frac{\partial u}{\partial z^*} + \frac{\partial \Phi}{\partial x} = 0 \ . \tag{4.30}$$

4.3.2. The continuity equation

Consider a fluid element of mass ΔM and cross-section area $\Delta x \, \Delta y$, which is confined between pressure surfaces p and $p-\Delta p$ as shown in fig. 4.7.

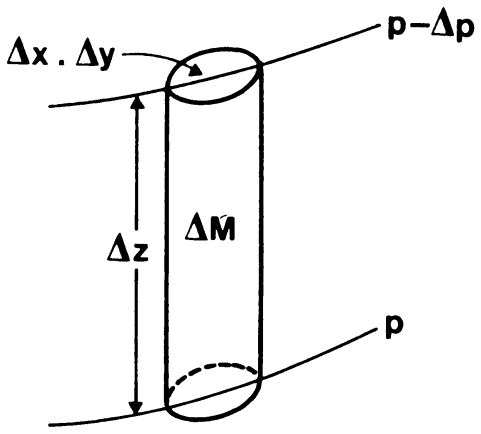

Figure 4.7 An air column of fixed mass ΔM confined between two isobaric surfaces.

As the problem is basically two-dimensional we remove Δy from the expression thus making ΔM a mass per unit length.
Applying the hydrostatic approximation we may write

$$\Delta M = \rho \Delta z \Delta x = \frac{\Delta x \Delta p}{g} .$$

Since the mass of the fluid element is conserved following motion, we obtain

$$\frac{1}{\Delta M} \frac{d}{dt} \Delta M = \frac{g}{\Delta x \Delta p} \frac{d}{dt}\left(\frac{\Delta x \Delta p}{g}\right) = 0 ,$$

or

$$\frac{\Delta}{\Delta x}\left(\frac{dx}{dt}\right) + \frac{\Delta}{\Delta p}\left(\frac{dp}{dt}\right) = 0 .$$

From definition (4.3) we have $p = p_o e^{-z^*/H}$ hence $\Delta p = -\frac{p}{H} \Delta z^*$ and

$$\frac{dp}{dt} = -\frac{p}{H} \frac{dz^*}{dt} = -w^* \frac{p}{H} .$$

This means that

$$\frac{\Delta}{\Delta p}\left(\frac{dp}{dt}\right) = \frac{\Delta}{\Delta p}\left(-w^* \frac{p}{H}\right) = \frac{-w^*}{H} - \frac{p}{H} \frac{\Delta w^*}{\Delta p} = \frac{-w^*}{H} + \frac{\partial w^*}{\partial z^*} .$$

The continuity equation in the log pressure coordinates finally becomes

$$\frac{\partial u}{\partial x} + \frac{\partial w^*}{\partial z^*} - \frac{w^*}{H} = 0 . \qquad (4.31)$$

4.3.3. The thermodynamic energy equation

Under the assumption that the motions are adiabatic, the first law of thermodynamics can also be written as

$$\rho c_p \frac{dT}{dt} - \frac{dp}{dt} = 0 . \qquad (4.32)$$

Remark that in order to verify this assumption the time scale of the motion has to be compared with the time scale referring to thermal conduction. As the thermal conduction varies considerably with height some doubt is justified

about the validity of the assumption of adiabatic motion in certain regions of the atmosphere.

The last term in equation (4.32) may be written as

$$\frac{dp}{dt} = \frac{-p}{H}\frac{dz^*}{dt} = \frac{-p}{H} w^* .$$

The time derivative $\frac{d}{dt}$ reads

$$\frac{d}{dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + w^*\frac{\partial}{\partial z^*} ,$$

hence

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + w^*\left(\frac{\partial T}{\partial z^*} + \frac{p}{\rho H c_p}\right) = 0 . \qquad (4.33)$$

Define the static stability Γ by $\Gamma := \frac{\partial T}{\partial z^*} + \frac{p}{\rho H c_p}$.

With this definition equation (4.33) becomes

$$\frac{\partial T}{\partial t} + u\frac{\partial T}{\partial x} + \Gamma w^* = 0 . \qquad (4.34)$$

The temperature T can be expressed in the geopotential as follows

$$\frac{\partial \Phi}{\partial z^*} = -\frac{\partial \Phi}{H \partial \ln p} = -\frac{p}{H}\frac{\partial \Phi}{\partial p} = \frac{p}{\rho H} = \frac{RT}{H} , \qquad (4.35)$$

or

$$T = \frac{H}{R}\frac{\partial \Phi}{\partial z^*} .$$

If we furthermore define the hydrostatic Brunt-Väisälä frequency N by $N^2 := \frac{R\Gamma}{H}$, the thermodynamic equation reads

$$\frac{\partial}{\partial t}\left(\frac{\partial \Phi}{\partial z^*}\right) + u\frac{\partial}{\partial x}\frac{\partial \Phi}{\partial z^*} + \left(N^2 w^*\right) = 0 . \qquad (4.36)$$

The following set of equations describes, within the hydrostatic approximation, the adiabatic variations of the fluid:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + w^* \frac{\partial u}{\partial z^*} + \frac{\partial \Phi}{\partial x} = 0 \ , \tag{4.37}$$

$$\frac{\partial u}{\partial x} + \frac{\partial w^*}{\partial z^*} - \frac{w^*}{H} = 0 \ , \tag{4.38}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \Phi}{\partial z^*}\right) + u \frac{\partial}{\partial x}\left(\frac{\partial \Phi}{\partial z^*}\right) + N^2 w^* = 0 \ . \tag{4.39}$$

We now expand on a static solution $\bar{U}(z^*)$ and $\Phi_o(z^*)$

$$\begin{aligned} u &= \bar{U} + u' \ , \\ w^* &= 0 + w \ , \\ \Phi &= \Phi_0 + \Psi \ , \end{aligned}$$

where

$u' << \bar{U}$ and $\Psi << \phi_o$.

Taking into account first order perturbations only and omitting the prime and the asterisk

$$\frac{\partial u}{\partial t} + \bar{U} \frac{\partial u}{\partial x} + \frac{\partial \Psi}{\partial x} + \frac{d\bar{U}}{dz} w = 0 \ , \tag{4.40}$$

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} - \frac{w}{H} = 0 \ , \tag{4.41}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial \Psi}{\partial z}\right) + \bar{U} \frac{\partial}{\partial x}\left(\frac{\partial \Psi}{\partial z}\right) + N^2 w = 0 \ . \tag{4.42}$$

From the last equation

$$w = - \frac{1}{N^2}\left(\frac{\partial}{\partial t} + \bar{U} \frac{\partial}{\partial x}\right) \frac{\partial \Psi}{\partial z} \ .$$

Substituting this in the first two equations yields

$$\left(\frac{\partial}{\partial t} + \bar{U} \frac{\partial}{\partial x}\right) u + \frac{\partial \Psi}{\partial x} - \frac{d\bar{U}}{dz} \frac{1}{N^2}\left(\frac{\partial}{\partial t} + \bar{U} \frac{\partial}{\partial x}\right) \frac{\partial \Psi}{\partial z} = 0 \ , \tag{4.43}$$

and

$$\frac{\partial u}{\partial x} - \frac{1}{N^2}\left(\left(\frac{\partial}{\partial t} + \bar{U}\frac{\partial}{\partial x}\right)\frac{\partial^2}{\partial z^2} + \frac{d\bar{U}}{dz}\frac{\partial^2}{\partial x \partial z}\right)\Psi + \frac{1}{N^2 H}\left(\frac{\partial}{\partial t} + \bar{U}\frac{\partial}{\partial x}\right)\frac{\partial \Psi}{\partial z} = 0 \ . \qquad (4.44)$$

The Brunt-Väisälä frequency N is assumed to be constant, which implies that the temperature is constant.
Multiplying equation (4.43) by $\frac{\partial}{\partial x}$, (4.44) by $\left(\frac{\partial}{\partial t} + \bar{U}\frac{\partial}{\partial x}\right)$ and subtracting, u is eliminated from these two equations and we end up with

$$\left(\frac{\partial}{\partial t} + \bar{U}\frac{\partial}{\partial x}\right)^2 \left(\frac{\partial^2 \Psi}{\partial z^2} - \frac{1}{H}\frac{\partial \Psi}{\partial z}\right) - N^2 \frac{\partial^2 \Psi}{\partial x^2} = 0 \ . \qquad (4.45)$$

This is the partial differential equation for the perturbation of the geopotential. Note that the derivatives of $\bar{U}$ have disappeared. This explains the simple structure of equation (4.45). We look for normal mode solutions:

$$\Psi(x,z,t) = \phi(z)\, e^{i(\omega t - kx)} \ .$$

Then the function $\phi(z)$ has to satisfy the linear second order ordinary differential equation

$$\frac{d^2\phi}{dz^2} - \frac{1}{H}\frac{d\phi}{dz} + \frac{N^2 k^2}{(\omega - k\bar{U})^2}\,\phi = 0 \ . \qquad (4.46)$$

With $\phi(z) := A(z) e^{\frac{z}{2H}}$, which means an exponential growth with height, equation (4.46) becomes

$$\frac{d^2 A}{dz^2} + \left[\frac{N^2 k^2}{\Omega^2} - \frac{1}{4H^2}\right] A = 0 \ , \qquad (4.47)$$

with the Doppler-shifted frequency Ω defined as $\Omega := \omega - k\bar{U}$.

Equation (4.47) has also been derived by Holton (1972), but he assumed a horizontal flow $\bar{U}$ independent of height, which - as is shown above - is not necessary. Equation (4.47) for A contains no derivatives of the background flow. Hence, applying the hydrostatic approximation and using log pressure as the vertical coordinate a relatively simple equation for the geopotential perturbation ϕ is obtained.

Notice that for a hyperbolic tangent wind profile, treated analytically by Van Duin & Kelder (1982), equation (4.47) can be transformed into the hypergeometric differential equation irrespective of the place of the critical level and the value of the Richardson number.

4.4. Critical levels in a jet-type flow

The background flow $\bar{U}(z)$ is taken to be

$$\bar{U}(z) = \frac{U_0}{1 + \frac{z^2}{D^2}} \tag{4.48}$$

This profile represents a symmetric jet-type flow. This type of flow occurs frequently in the atmosphere. For example in figure 4.8 the zonal wind structure is given as a function of height and more than one jet can be identified.

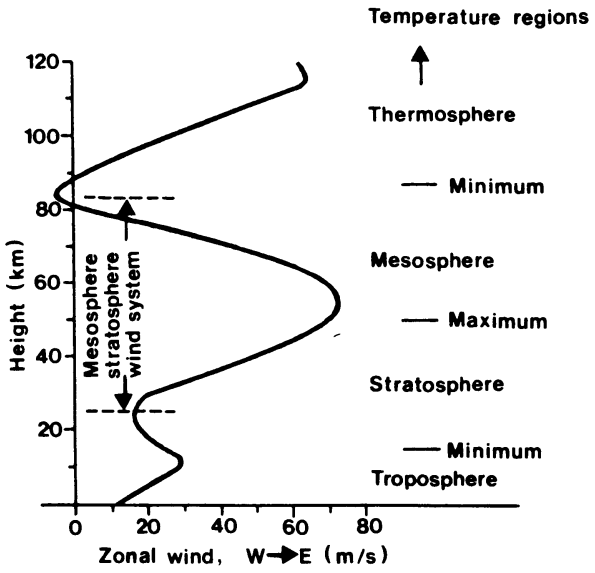

Figure 4.8 Magnitude of zonal winds in the upper atmosphere by the height profile for 45° N latitude in January (from COSPAR International Reference Atmosphere, 1972).

Substituting expression (4.48) in equation (4.47) yields

$$\frac{d^2A}{dz^2} + \left[\frac{k^2N^2}{\omega^2}\,\frac{\left(1+\frac{z^2}{D^2}\right)^2}{\left(\frac{z^2}{D^2}-\left(\frac{U_o}{c}-1\right)\right)^2} - \frac{1}{4H^2}\right] A = 0 \,, \qquad (4.49)$$

where $c := \frac{\omega}{k}$ is the horizontal phase velocity.

Three cases can be distinguished, namely:
$U_o < c$ no critical layer,
$U_o > c$ two critical layers,
$U_o = c$ one critical layer.

Below we consider the last two cases.

4.5. Two critical levels

As discussed above, if the horizontal phase velocity of the wave is smaller than U_o, the wave encounters two critical levels in the flow. This case will now be examined.
After introducing the independent variable ζ defined by $\zeta := \frac{z}{Dd}$, where

$d = \left(\frac{U_o}{c}-1\right)^{1/2}$, equation (4.49) becomes:

$$\frac{d^2A}{d\zeta^2} + \left(\frac{N^2D^2}{c^2d^2}\,\frac{(1+d^2\zeta^2)^2}{(\zeta^2-1)^2} - \frac{D^2d^2}{4H^2}\right) A = 0 \,. \qquad (4.50)$$

Furthermore, if we define the function $B(\zeta) := (\zeta^2-1)^{-1/2} A(\zeta)$, the equation for B reads:

$$(1-\zeta^2)\frac{d^2B}{d\zeta^2} - 2\zeta\frac{dB}{d\zeta} + \left(\lambda + \gamma^2(1-\zeta^2) - \frac{\mu^2}{1-\zeta^2}\right) B = 0 \,, \qquad (4.51)$$

where

$$\lambda := -2\,\frac{D^2 N^2 U_o}{c^3}\,, \qquad \gamma^2 := D^2 d^2 \left(\frac{N^2}{c^2} - \frac{1}{4H^2}\right),$$

and

$$\mu^2 := 1 - 4Ri_c \qquad \text{with} \qquad Ri_c := \frac{D^2 N^2 U_o^2}{4c^4 d^2}\,.$$

The Richardson number Ri is defined as $Ri := N^2/(\frac{d\bar{U}}{dz})^2$. It is easy to verify that Ri_c is the Richardson number at the critical level.
Equation (4.51) is known as the differential equation of spheroidal wave functions. The properties of its solutions are discussed in Meixner & Schäfke (1954) (hereinafter referred to as MS) and Erdelyi et al. (1953).
The equation has three singular points at $y = \pm 1$ and ∞. The points ± 1 are regular singularities, whereas the one at ∞ is an irregular one. The parameter μ is called the order of the wave function.
We need solutions which asymptotically become plane waves, the circuit relations between solutions of that form then give the reflection and transmission coefficients.

Solutions with these properties are $S_\nu^{\mu(3,4)}(\zeta;\gamma)$. . For $|\zeta| > 1$ they can be represented by convergent series of the following form

$$S_\nu^{\mu(3,4)}(\zeta;\gamma) = \frac{(\zeta^2-1)^{-\mu/2}\,\zeta^\mu}{A_\nu^\mu(\gamma^2)} \sum_{r=-\infty}^{\infty} a_{\nu,2r}^\mu(\gamma^2)\,\psi_{\nu+2r}^{(3,4)}(\zeta\gamma), \qquad (4.52)$$

where $A_\nu^\mu(\gamma^2) = \sum_{r=-\infty}^{\infty} (-1)^r a_{\nu,2r}^\mu(\gamma^2)$ and $\psi_{\nu+2r}^{(j)}$ are the spherical

Hankel functions, that is

$$\psi_\nu^{(3)}(\zeta) := (\tfrac{\pi}{2\zeta})^{1/2} H_{\nu+1/2}^{(1)}(\zeta) \text{ and } \psi_\nu^{(4)}(\zeta) := (\tfrac{\pi}{2\zeta})^{1/2} H_{\nu+1/2}^{(2)}(\zeta).$$

The coefficients $a_{\nu,2r}^\mu$ have to satisfy the following three term recurrency relation

$$\gamma^2 \frac{(\nu+\mu+2r+2)(\nu+\mu+2r+1)}{(2\nu+4r+3)(2\nu+4r+5)} a^{\mu}_{\nu,2r+2}(\gamma^2) +$$

$$[\lambda - (\nu+2r)(\nu+2r+1) + 2\gamma^2 \frac{(\nu+2r)(\nu+2r+1) + \mu^2-1}{(2\nu+4r-1)(2\nu+4r+3)}] a^{\mu}_{\nu,2r}(\gamma^2) \qquad (4.53)$$

$$+ \gamma^2 \frac{(\nu+2r-\mu)(\nu+2r-\mu-1)}{(2\nu+4r-3)(2\nu+4r-1)} a^{\mu}_{\nu,2r-2}(\gamma^2) = 0$$

The parameter ν is called the characteristic exponent of the spheroidal differential equation, and is a function of λ, μ and γ^2.
It is possible to expand λ in a power serie in γ^2, with coefficients depending on μ and ν(see MS) as

$$\lambda^{\mu}_{\nu}(\gamma^2) = \nu(\nu+1) - \tfrac{1}{2}\left[1 + \frac{(2\mu-1)(2\mu+1)}{(2\nu-1)(2\nu+3)}\right]\gamma^2 + \ldots \qquad (4.53a)$$

It is obvious that:

$$\lambda^{\mu}_{\nu}(0) = \nu(\nu+1).$$

Other solutions we need are $Ps^{\mu}_{\nu}(\zeta;\gamma^2)$ and $Qs^{\mu}_{\nu}(\zeta;\gamma^2)$. They may be represented in the following form

$$Ps^{\mu}_{\nu}(\zeta;\gamma^2) = \sum_{r=-\infty}^{\infty} (-1)^r a^{\mu}_{\nu,2r}(\gamma^2) \; P^{\mu}_{\nu+2r}(\zeta),$$

and

$$Qs^{\mu}_{\nu}(\zeta;\gamma^2) = \sum_{r=-\infty}^{\infty} (-1)^r a^{\mu}_{\nu,2r}(\gamma^2) \; Q^{\mu}_{\nu+2r}(\zeta), \qquad (4.54)$$

where P^{μ}_{ν} and Q^{μ}_{ν} are the Legendre functions of the first and second kind respectively. The series in (4.54) converge absolutely and uniformly in every compact domain with the exceptions of ± 1 and ∞.

The asymptotic behaviour of $S^{\mu(3)}_{\nu}(\zeta;\gamma)$ as $\zeta \to \infty$ is

$$S^{\mu(3)}_{\nu}(\zeta;\gamma) \sim (\zeta^2-1)^{-\mu/2} \zeta^{\mu} \frac{\exp\left[i(\gamma\zeta - \frac{\nu+1}{2}\pi)\right]}{\gamma\zeta} \left\{\sum_{s=0}^{q-1} \frac{A^{\mu}_{\nu,s}}{[-2i\gamma\zeta]^s} + O(\zeta^{-q})\right\}, \qquad (4.55)$$

for $-\pi + \varepsilon \leq \arg(\gamma\zeta) \leq 2\pi - \varepsilon$, $\varepsilon > 0$.

For $S_\nu^{\mu(4)}(\zeta;\gamma)$ the asymptotic form is:

$$S_\nu^{\mu(4)}(\zeta;\gamma) \sim (\zeta^2-1)^{-\mu/2}\,\zeta^\mu \frac{\exp\left[-i(\gamma\zeta - \frac{\nu+1}{2}\pi)\right]}{\gamma\zeta}\left\{\sum_{s=0}^{q-1}\frac{A_{\nu,s}^\mu}{[2i\gamma\zeta]^s} + O(\zeta^{-q})\right\}, \quad (4.56)$$

for $-2\pi + \varepsilon \leq \arg(\gamma\zeta) \leq \pi - \varepsilon \quad , \quad \varepsilon > 0.$

Here

$$A_\nu^\mu A_{\nu,s}^\mu = \frac{1}{s!}\sum_{r=-\infty}^{+\infty}\frac{\Gamma(\nu+2r+s+1)}{\Gamma(\nu+2r-s+1)}(-1)^r a_{\nu,2r}^\mu,$$

$$A_{\nu,0}^\mu = 1.$$

The $A_{\nu,s}^\mu$ have to satisfy a four terms recurrency relation which can be found in MS.

From (4.55) and (4.56), it follows that:

for $z \to \infty$

$$A(z) \sim \frac{\exp\left[\pm i(\ell z - \frac{\nu+1}{2}\pi)\right]}{\gamma} \qquad |\arg(\ell z)| < \pi\,, \quad (4.56)$$

where $\ell = \left[\frac{N^2}{c^2} - \frac{1}{4H^2}\right]^{1/2}$ is the vertical wave number without mean flow.

The relationship between A and ϕ means that ϕ tends exponentially to infinity as $z \to \infty$. This is a consequence of the density stratification. However, it must be taken into account that the wavelike form (4.56) is already a good approximation to the solution if the mean flow velocity becomes negligible with respect to the horizontal phase velocity of the wave.

The plus or minus signs in (4.56) correspond to the solutions $S_\nu^{\mu(3)}$ and $S_\nu^{\mu(4)}$ respectively.

The vertical wave energy flux, at least when the mean background flow is negligible, may be written as

$$F_w = \rho_o \langle \phi\ w \rangle = \frac{1}{2} \rho_o \quad \mathrm{Re}\,[\phi\ w^*],$$

where the brackets < > refer to the average over one cycle of the wave and ρ_o is the average state density. The asterisk means the complex conjugate.

Equation (4.39) with the temporal and horizontal dependence used here gives:

$$w = - i \frac{\Omega}{N^2} \phi_z \tag{4.57}$$

With (4.57) the vertical wave energy flux becomes:

$$F_w = \frac{\rho_o}{2} \mathrm{Re}\,\left[\frac{i\Omega}{N^2}\ \phi\ \phi_z^*\right],$$

and with the asymptotic form (4.56):

$$F_w = \pm\ \rho_{oo} \frac{\Omega\ \ell}{2N^2}\ , \tag{4.58}$$

where the plus or minus signs correspond to the plus or minus signs in (4.56). ρ_{oo} is defined by $\rho_o = \rho_{oo} e^{-z/H}$.
Thus in the upper half space, $S_\nu^{\mu(3)}$ tends asymptotically to an upward-, and $S_\nu^{\mu(4)}$ to a downward-propagating wave.

4.6. The reflection and transmission coefficients

Suppose there is a source emitting waves at $-\infty$. For $z \to +\infty$, we must only have an upward-propagating wave. This boundary condition may be satisfied as seen at the end of the previous section by $S_\nu^{\mu(3)}(\zeta;\Upsilon)$. It is then necessary to find the analytical continuation for values of $z \to -\infty$. The physically meaningful path of analytical continuation is discussed in Booker & Bretherton (1967), Baldwin & Roberts (1970), and Teitelbaum & Kelder (1985).

The connection path can be one of the equivalent paths shown in fig. 4.9. We followed the path indicated by a solid line, allowing us to use properties of the solutions, as we shall see later.

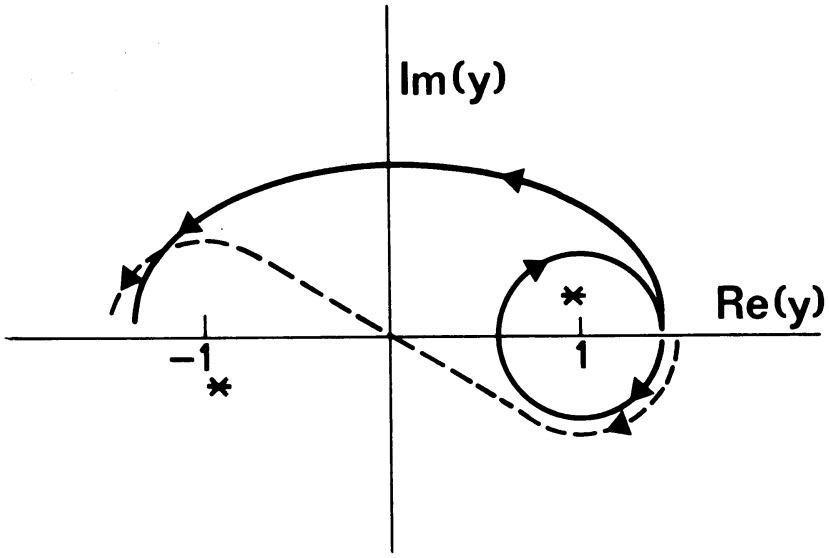

Figure 4.9 Complex y-plane showing the two equivalent paths for analytical continuation of the spheroidal wave functions: ---, direct path; ——, indirect path used in our calculation; *, singularities shifted as a consequence of the addition of a small dissipation.

The $S_\nu^{\mu(3,4)}$ are only defined for $|\zeta| > 1$. First we have to extend the functions into the unit circle. This may be done with the following relations (see MS):

$$S_\nu^{\mu(3)}(\zeta;\gamma) = \frac{1}{\cos\nu\pi}\left[-i\, V_{-\nu-1}^{\mu}\, \tilde{Q}s_\nu^{\mu}(\zeta;\gamma^2) + e^{-i\nu\pi}\, V_\nu^{\mu}\, \tilde{Q}s_{-\nu-1}^{\mu}(\zeta;\gamma^2)\right], \qquad (4.59)$$

$$S_\nu^{\mu(4)}(\zeta;\gamma) = \frac{1}{\cos\nu\pi}\left[e^{i\nu\pi}\, V_\nu^{\mu}\, \tilde{Q}s_{-\nu-1}^{\mu}(\zeta;\gamma^2) + i\, V_{-\nu-1}^{\mu}\, \tilde{Q}s_\nu^{\mu}(\zeta;\gamma^2)\right], \qquad (4.60)$$

where $\tilde{Q}s_\nu^{\mu}(\zeta;\gamma^2) = \frac{e^{-i\mu\pi}}{\Gamma(\nu+\mu+1)}\, Qs_\nu^{\mu}(\zeta;\gamma^2)$

and V_ν^μ is a constant, which for small values of γ^2 can be approximated by

$$V_\nu^\mu(\gamma) = \tfrac{1}{2}\left(\frac{\gamma}{4}\right)^\nu \frac{\Gamma(-\nu+\frac{1}{2})}{\Gamma(\nu+\frac{3}{2})}\,(1 - O(\gamma^2)) . \tag{4.61}$$

The connection path can be split into two parts.
First we turn around $\zeta = 1$ through -2π. In MS it is proved that:

$$Ps_\nu^\mu\left[1 + (\zeta-1)\,e^{-2i\pi}\right] = e^{\mu\pi i}\,Ps_\nu^\mu(\zeta) ; \tag{4.62}$$

accordingly

$$\tilde{Q}s_\nu^\mu\left[1+(\zeta-1)e^{-2\pi i}\right] = e^{i\nu\pi}\,\frac{\cos\mu\pi}{\cos\nu\pi}\,\tilde{Q}s_\nu^\mu(\zeta) - i\,\frac{\sin(\nu-\mu)\pi\;\Gamma(\mu-\nu)}{\cos\nu\pi\;\Gamma(\nu+\mu+1)}\,\tilde{Q}s_{-\nu-1}^\mu(\zeta). \tag{4.63}$$

Next we turn around $\zeta = 0$ through $+\pi$, which gives

$$\tilde{Q}s_\nu^\mu(\zeta\, e^{i\pi}) = e^{-(\nu+1)\pi i}\,\tilde{Q}s_\nu^\mu(\zeta). \tag{4.64}$$

Inserting (4.63) into (4.64) yields

$$\tilde{Q}s_\nu^\mu(-\zeta) = -\frac{\cos\mu\pi}{\cos\nu\pi}\,\tilde{Q}s_\nu^\mu(\zeta) - i\,\frac{\sin(\nu-\mu)\pi\;\Gamma(\mu-\nu)}{\cos\nu\pi\;\Gamma(\nu+\mu+1)}\,e^{i\nu\pi}\,\tilde{Q}s_{-\nu-1}^\mu(\zeta). \tag{4.65}$$

In the same way an expression for $\tilde{Q}s_{-\nu-1}^\mu(-\zeta)$ may be derived.
With the help of (4.59) and (4.60), we can go back to $S_\nu^{\mu(3,4)}$ and the result is the following circuit relation:

$$S_\nu^{\mu(3)}(-\zeta;\gamma) = \alpha\, S_\nu^{\mu(3)}(\zeta;\gamma) + \beta\, S_\nu^{\mu(4)}(\zeta;\gamma) , \tag{4.66}$$

with

$$\alpha = \frac{2\cos\mu\pi\,\sin\nu\pi + i\left[V^+ e^{-i\nu\pi} - V^- e^{i\nu\pi}\right]}{2\,i\,\cos^2\nu\pi} , \tag{4.67}$$

and

$$\beta = \frac{i\,e^{-i\nu\pi}\left[2\cos\nu\pi - V^+ e^{-2i\nu\pi} - V^- e^{2i\nu\pi}\right]}{2\,i\,\cos^2\nu\pi}, \tag{4.68}$$

where

$$V^{+} = \sin(\nu+\mu)\pi \frac{\Gamma(\mu+\nu+1)\, V_{\nu}^{\mu}(\gamma)}{\Gamma(\mu-\nu)\, V_{-\nu-1}^{\mu}(\gamma)}, \tag{4.69}$$

and

$$V^{-} = \sin(\nu-\mu)\pi \frac{\Gamma(\mu-\nu)\, V_{-\nu-1}^{\mu}(\gamma)}{\Gamma(\mu+\nu+1)\, V_{\nu}^{\mu}(\gamma)}. \tag{4.70}$$

If we let $\zeta \to \infty$ equation (4.66), with the asymptotic forms (4.55) and (4.56), represents a relation between plane waves.

If we take into account that for negative ζ, $S_{\nu}^{\mu(4)}$ represents the incident wave and $S_{\nu}^{\mu(3)}$ the reflected one, the coefficients of reflection R and transmission T become

$$R = \frac{\alpha}{e^{i\nu\pi}\,\beta}, \tag{4.71}$$

and

$$T = \frac{1}{e^{i\nu\pi}\,\beta}, \tag{4.72}$$

with α and β defined above.

The parameters are μ, γ and λ, but the number may be reduced using the fact that the horizontal phase velocity of the wave c is much smaller than the velocity of sound. A good approximation for the velocity of sound in the atmosphere is $V_s = 4\, N^2 H^2$, and with this approximation it is possible to write:

$$\lambda = \frac{-2J}{S^3}\ ;\ \mu^2 = 1 - \frac{J}{S^3(1-S)},\ \text{and}\ \gamma^2 \simeq \frac{J(1-S)}{S^3}, \tag{4.73}$$

where

$$J := \frac{N^2 D^2}{U_o^2} \quad \text{and} \quad S := \frac{c}{U_o}.$$

Note that this approximation, which can be interpreted as the incompressible fluid approximation ($H \to \infty$), has not been used elsewhere in this study. It is only adopted to simplify the presentation and interpretation.

Thus, essentially, only two parameters are involved: J and the ratio S between the phase velocity and the maximum velocity of the background flow.

Another useful parameter is the minimum Richardson number of the flow Ri_m, which corresponds to S = 0.75 and is equal to $(\frac{4}{3})^3$ J.

In figures 4.10 and 4.11 the variations of $|R|$ and $|T|$ are indicated as a function of S for some values of Ri_m. Note that the values of S for which the maximum of $|R|$ and $|T|$ are reached are greater than 0.75, i.e. $Ri_c > Ri_m$.

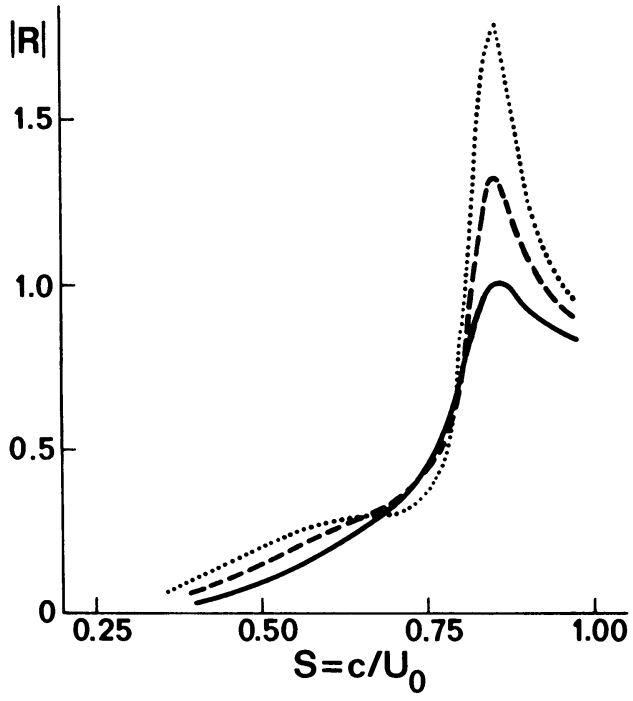

Figure 4.10 Variation of the reflection coefficient $|R|$ as a function of $S = c/U_0$, for three different values of the minimum Richardson number of the mean flow: ——, Ri_m = 0.143; ---, 0.1; ..., 0.07.

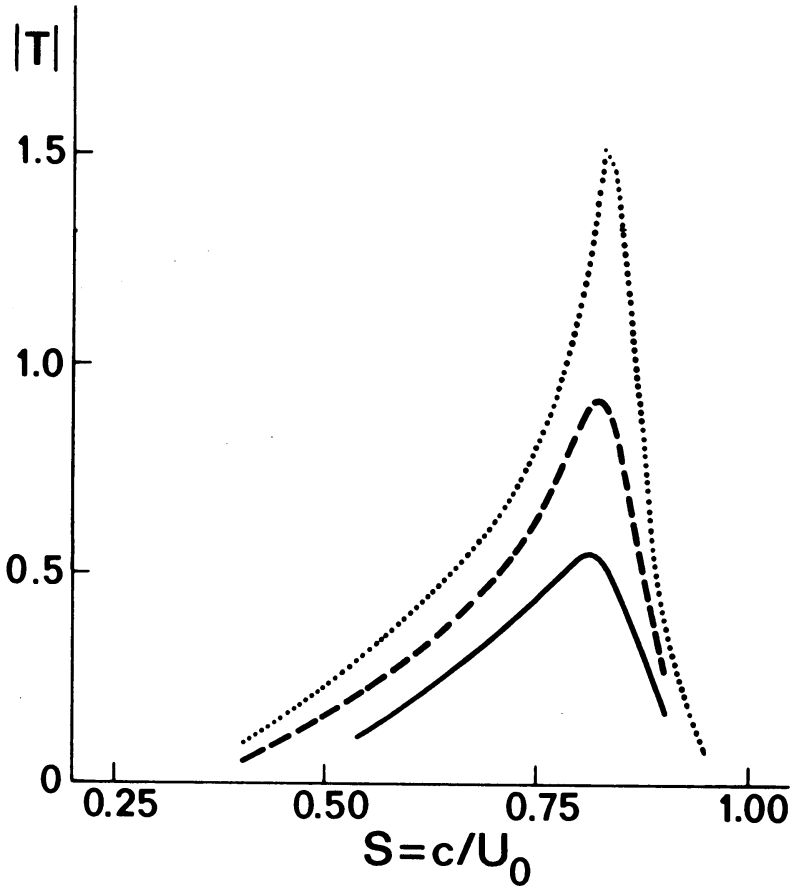

Figure 4.11 Variation of the transmission coefficient $|T|$ as a function of $S = c/U_o$, for three different values of the minimum Richardson number of the mean flow: ——, $Ri_m = 0.143$; ---, 0.1; ..., 0.07.

Overreflection starts with $Ri_m = 0.143$, while at the critical level $Ri_c = 0.169$. Lower values are found for one critical level. Jones (1968) calculated $Ri_c = 0.113$; Eltayeb & McKenzie (1975) obtained 0.115 and Van Duin & Kelder (1982) 0.132. The background mean flow considered by Jones and Eltayeb & McKenzie was formed by matching constant shear layers, which implies that the Richardson number is constant in each layer. The value given by Van Duin & Kelder for the hyperbolic tangent profile corresponds to $Ri_c = Ri_m$. The higher value found with two critical levels can be explained by the fact that, as we shall show, the upper critical level acts as a source of wave energy.

In regions where $\bar{U} \neq 0$, (4.58) does not represent the total vertical energy flux. Another term which represents the advection by the wave field of the kinetic energy of the mean flow is needed (Hines & Reddy, 1967; Lindzen, 1973) and this leads to:

$$F_E = \rho_o \bar{\phi}\bar{w} + \rho_o \bar{U}\bar{u}\bar{w} . \qquad (4.74)$$

In fact, what we need is a quantity whose flux is conserved across the jet except at critical levels. We could as well use other quantities such as horizontal wave momentum (Eliassen and Palm, 1962) or wave action (Bretherton, 1969; Andrews & McIntyre, 1978; Grimshaw, 1984). In the present case the flux of these quantities agrees with the flux of total energy as defined above up to a multiplicative constant.
The horizontal perturbation velocity (4.43) reads:

$$u = \frac{\bar{U}_z}{N^2} \phi_z + \frac{k}{\Omega} \phi ,$$

which allows us to write the vertical energy flux as follows

$$F_E = \tfrac{1}{2} \frac{\rho_o}{N^2} \text{Real} \left[i\, w\, \phi\, \phi_z^* \right] . \qquad (4.75)$$

Near the upper critical level where $\zeta - 1 > 0$ and $Ri_c > \frac{1}{4}$, $\phi(\zeta)$ can be written as

$$\phi = \exp\left[\frac{Dd\zeta}{2H}\right] \left[P(\zeta-1)^{\frac{1+i\mu}{2}} + Q(\zeta-1)^{\frac{1-i\mu}{2}} \right],$$

where P and Q are complex constants.

This last expression used in (4.75) gives

$$F_E^+ = \rho_{oo} \frac{\Omega\, \mu}{4Dd} \frac{1}{N^2} \left[|P|^2 - |Q|^2 \right] ,$$

where F_E^+ is the vertical energy flux above the critical layer.

The analytical continuation of ϕ for $\zeta-1 < 0$ through $-\pi$ is

$$\phi = \left[-i\, P\, e^{\mu\pi}\, (1-\zeta)^{\frac{1+i\mu}{2}} - i\, Q\, e^{-\mu\pi}\, (1-\zeta)^{\frac{1-i\mu}{2}}\right] \exp\left[\frac{Dd\zeta}{2H}\right],$$

which gives

$$F_E^- = \frac{\rho_{oo}\Omega\mu}{4dD}\,\frac{1}{N^2}\left[\,|Q|^2\, e^{-2\mu\pi} - |P|^2\, e^{+2\mu\pi}\right],$$

where F_E^- is the vertical energy flux evaluated below the critical level. With the condition of only an upward propagating wave above the upper critical level, F_E^+ has to be positive and this implies F_E^- negative: the upper critical level acts as a source of energy.

For $Ri_c < \frac{1}{4}$ it is not possible to prove the same statement, but numerical calculation shows that F_E^- always remains smaller than F_E^+ (Teitelbaum & Kelder, 1985). Hence there is again a netto positive flux of vertical energy at the upper critical level.

The propagation of an internal gravity wave through a jet type flow encountering two critical levels has also been solved numerically (Teitelbaum & Kelder, 1985). The results are in case of overlap consistent with those obtained analytically. We quote here some interesting results from these numerical calculations. If the profile which gives $|R| = 1$ ($Ri_c = 0.169$; $Ri_m = 0.143$) is modified in its upper part by taking a constant mean flow above its maximum, the upper critical level is suppressed. In that case, the reflection coefficient becomes $|R| = 0.91$.
The transmission coefficient depends on Ri_c and on the profile of the flow. If the mean shear of the flow below the critical levels is not very strong, the absolute value of the transmission coefficient $|T|$ can be closely approximated by $|T| = \exp\left[-2\pi(Ri_c - \frac{1}{4})^{\frac{1}{2}}\right]$. This is consistent with Booker & Bretherton's results in the sense that the two critical levels are apparently acting independently and each critical level is transmitting the wave with the Booker & Bretherton transmission factor $\exp\left[-\pi\,(Ri_c-\frac{1}{4})^{\frac{1}{2}}\right]$. If the shear is strong (low value of Ri_m), then even for large values of Ri_c, $|T|$ becomes lower than the product of the transmission coefficients because the wave is partially reflected below the critical level. As an example, the case

$Ri_m = Ri_c = 5$ gives $|R| = 2.9 \times 10^{-4}$, $|T| = 1.29 \times 10^{-6}$. Note that this value of $|T|$ corresponds to the product of two separated critical levels. With the same $Ri_c = 5$, but $Ri_m = 0.12$, the results are $|R| = 0.86$ and $|T| = 3 \times 10^{-7}$.
Comparison of our reflection coefficients with those found by Drazin et al (1979) for the triangular jet shows that our values are much lower. This can be explained by the reflection at the knees of their broken-line profile (see e.g. Jones, 1968; Eltayeb & McKenzie, 1975).

4.7. One critical level

When the horizontal phase velocity of the wave equals the maximum of the mean-flow velocity there is only one critical level. Mathematically, the two regular singularities of (4.51) are merging in one irregular singularity. However, when a small dissipation term is added the irregular singularity splits into two regular ones. Thus the case of one critical level has been reduced to the case of two merging critical levels.

Let us start with two critical levels and let $c \to U_o$, i.e. $S \to 1$. From (4.73) it is clear that $\gamma^2 \to 0$, $\mu \to i\infty$ and $\lambda \to -2J$. We calculated the values of R and T in this limit.
From Stirling's formula for gamma functions it follows that

$$\Gamma(\mu+\nu+1) \simeq e^{-\mu}\mu^{\mu+\nu+\frac{1}{2}}(2\pi)^{\frac{1}{2}} \tag{4.76}$$

and

$$\Gamma(\mu-\nu) \simeq e^{-\mu}\mu^{\mu-\nu-\frac{1}{2}}(2\pi)^{\frac{1}{2}} \tag{4.77}$$

for $\mu \to i\infty$. From (4.61) it may be inferred that for $\gamma \to 0$

$$\frac{V_\nu^\mu}{V_{-\nu-1}^\mu} \simeq \left(\frac{\gamma}{4}\right)^{2\nu+1} \left(\frac{\Gamma(-\nu+\frac{1}{2})}{\Gamma(\nu+\frac{3}{2})}\right)^2. \tag{4.78}$$

Moreover, if we put $\mu = iM$ then

$$\sin(\nu+\mu)\pi \simeq \tfrac{1}{2}ie^{(M-i\nu)\pi} , \tag{4.79}$$

$$\sin(\nu-\mu)\pi \simeq -\tfrac{1}{2}ie^{(M+i\nu)\pi} , \tag{4.80}$$

$$\cos \mu\pi \simeq \tfrac{1}{2}e^{M\pi}. \tag{4.81}$$

From (4.76) - (4.81) it can be derived that

$$V^{+} \sim \tfrac{1}{2}ie^{(M-i\nu)\pi} \left(\frac{\gamma\mu}{4}\right)^{2\nu+1} \left(\frac{\Gamma(-\nu+\frac{1}{2})}{\Gamma(\nu+\frac{3}{2})}\right)^{2} .$$

Taking into account that $\mu \simeq i\ ND/U_o d$ and $\gamma \to Dld$, we can write

$$V^{+} \sim -\tfrac{1}{2}\ e^{M\pi F} , \tag{4.83}$$

$$V^{-} \sim -\frac{e^{M\pi}}{2F} , \tag{4.84}$$

where

$$F = \left(\frac{ND^2l}{4U_o}\right)^{2\nu+1} \left(\frac{\Gamma(-\nu+\frac{1}{2})}{\Gamma(\nu+\frac{3}{2})}\right)^{2} . \tag{4.85}$$

Introducing (4.83) and (4.84) in (4.67) and (4.68) to calculate α and β , and using these results in (4.71) and (4.72), we obtain for the reflection and transmission coefficients

$$R = \frac{\sin \nu\pi + \frac{1}{2}i\left[F^{-1}e^{i\nu\pi} - F\ e^{-i\nu\pi}\right]}{i\left[1 + \frac{1}{2}(F^{-1}e^{2i\nu\pi} + F\ e^{-2i\nu\pi})\right]} , \tag{4.86}$$

$$T = \frac{e^{-M\pi}2\ \cos^2\ \nu\pi}{1 + \frac{1}{2}(F^{-1}e^{2i\nu\pi} + F\ e^{-2i\nu\pi})} . \tag{4.87}$$

In the limit $S \to 1$, that is $\gamma \to 0$, we can deduce from (4.53a) that

$$\nu \to -\tfrac{1}{2} - \tfrac{1}{2}(1 + 4\lambda)^{\frac{1}{2}} \simeq -\tfrac{1}{2} - \tfrac{1}{2}(1 - 8J)^{\frac{1}{2}}, \tag{4.88}$$

which shows that ν takes on a finite value in this limit.

From (4.87) it is clear that as $S \to 1$, the transmission coefficient $T \to 0$. The wave is not transmitted. This result is consistent with Booker & Bretherton's result, as, for $S \to 1$, $Ri_c \to \infty$.

Numerical calculation with (4.86) shows that $|R|^2 \leq 1$.
Hence a wave propagating through a jet with two merging critical levels will not be overreflected, but only partially reflected and not transmitted.
Figure 4.12 shows that $|R|$ decreases with increasing values of Ri_m.

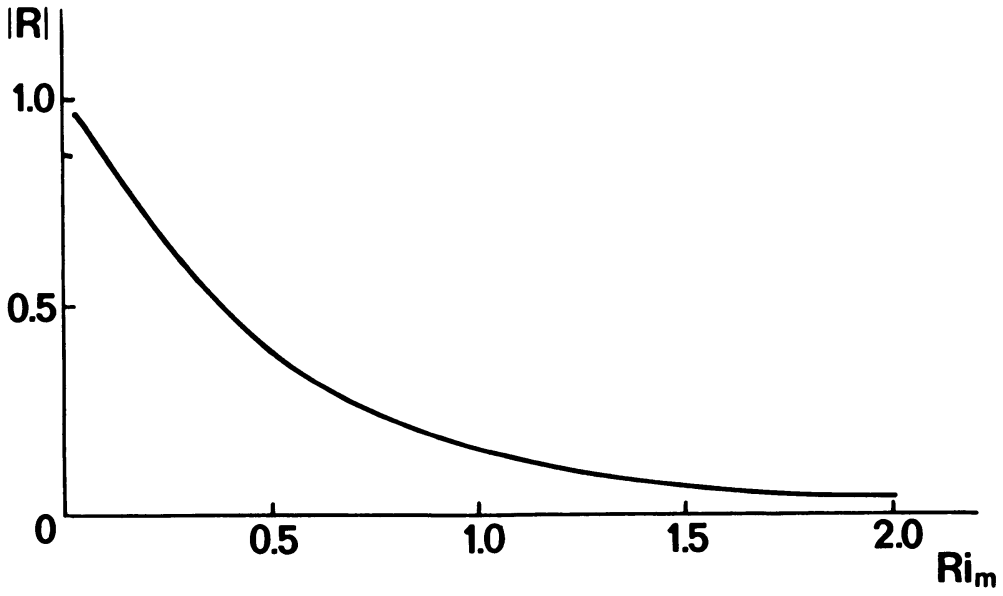

Figure 4.12 Variation of the absolute value of the reflection coefficient as a function of the minimum Richardson number of the mean flow in the case of two merging critical levels.

4.8. Conclusion

This work was motivated by the frequent observations of jet-type background winds in the atmosphere and the fact that planetary waves are seen as stationary jets by short-period gravity waves.

In such a background flow a gravity wave can have two critical levels or only one with specific characteristics. In this study we have shown that the two critical levels do not act independently to determine the behaviour of the travelling wave. In fact, some of the energy transmitted through the lower critical level can be reflected at the upper one, which acts as a source of wave energy. Then the downward energy flux is added to the energy already reflected at the lower critical level and can produce overreflection, with Ri_c higher than in the case of only one critical level. We found overreflection with $Ri_c = 0.169$, the critical levelt located at $S = 0.86$. On the other hand, even for values of Ri_c as low as 0.1, and with $Ri_c = Ri_m$ ($S = 0.75$), neither overreflection nor overtransmission occurs.

The transmission coefficient is different from the product of the transmission coefficients of the two critical levels when $0.25 \leq Ri_m < 1$ because the wave is partially reflected below the critical levels by the strong shear.

The results found for asymmetric jets show that the upper critical level can affect the reflection of the wave.

The case of a wave having only one critical level at the maximum of the background flow has been solved as the limit of two merging critical levels. In this case the transmission becomes zero and no overreflection can occur.

4.9. References

Andrews, D.G. & McIntyre, M.E., 1978 - On wave action and its relatives. J. Fluid Mech. 89, 647-664.

Baldwin, P. & Roberts, P.H., 1970 - The critical layer in stratified shear flow. Mathematika 17, 102-119.

Booker, J.R. & Bretherton, F.P., 1967 - The critical layer for internal gravity waves in a shear flow. J. Fluid Mech. 27, 513-529.

Boussinesq, J., 1903 - Théorie Analytique de la Chaleur, 2. Gauthier Villars, Paris.

Bretherton, F.P., 1969 - Momentum transport by gravity waves. Quart. J.R. Met. Soc. 95, 213-243.

Brown, S.N. & Stewartson, K., 1980 - On the nonlinear reflection of a gravity wave at a critical level. Part 1. J. Fluid Mech. 100, 577-595.

Brown, S.N. & Stewartson, K., 1982a - On the nonlinear reflection of a gravity wave at a critical level. Part 2. J. Fluid Mech. 115, 217-230.

Brown, S.N. & Stewartson, K., 1982b - On the nonlinear reflection of a gravity wave at a critical level. Part 3. J. Fluid Mech. 115, 231-250.

Drazin, P.G., Zaturska, M.B. & Banks, W.H.H., 1979 - On the normal modes of parallel flow of inviscid stratified fluid. Part 2. J. Fluid Mech. 95, 681-705.

Eliassen, A. & Palm, E., 1961 - On the transfer of energy in stationary mountain waves. Geofys. Publ. 22, 1-23.

Eltayeb, I.A. & McKenzie, J.F., 1975 - Critical-level behaviour and wave amplification of a gravity wave incident upon a shear layer. J. Fluid Mech. 72, 661-671.

Erdelyi, A., Magnus, W., Oberhettinger, F & Tricomi, F.G., 1953 - Higher Transcendental Functions. Vol. I, II and III. McGraw-Hill.

Fritts, D.C., 1978 - The nonlinear gravity wave critical-level interaction. J. Atmos. Sci. 35, 397-413.

Fritts, D.C., 1982 - The transient critical-level interaction in a Boussinesq fluid. J. Geophys. Res. 87, 7997-8016.

Fritts, D.C. & Geller, M.A., 1976 - Viscous stabilization of gravity wave critical level flows. J. Atmos. Sci. 33, 2276-2284.

Geller, M.A., Tanaka, H. & Fritts, D.C., 1975 - Production of turbulence in the vicinity of critical levels for internal gravity waves. J. Atmos. Sci. 32, 2125-2135.

Gill, A.E., 1982 - Atmosphere-Ocean Dynamics. Academic Press.

Grimshaw, R., 1980 - A general theory of critical level absorption and valve effects for linear wave propagation. Geophys. Astrophys. Fluid Dyn. 14, 303-326.

Grimshaw, R., 1984 - Wave action and wave-mean flow interaction, with application to stratified shear flow. Ann. Rev. Fluid Mech. 16, 11-44.

Hazel, P., 1967 - The effect of viscosity and heat conduction on internal gravity waves at a critical level. J. Fluid Mech. 30, 775-783.

Hines, C.O. & Reddy, C.A., 1967 - On the propagation of gravity waves through a region of wind shear. J. Geophys. Res. 72, 1015-1034.

Holton, J.R., 1972 - An Introduction to Dynamic Meteorology. Academic Press.

Jones, W.L., 1968 - Reflexion and stability of waves in stably stratified fluid with shear flow: a numerical study. J. Fluid Mech. 34, 609-624.

Lindzen, R.S., 1973 - Wave mean flow interaction in the upper atmosphere. Boundary-Layer Met. 4, 327-343.

Lindzen, R.S. & Rosenthal, A.J., 1983 - Instabilities in a stratified fluid having one critical level. Part III: Kelvin Helmholtz instabilities as overreflected waves. J. Atmos. Sci. 40, 530-542.

Meixner, J. & Schäfke, F.W., 1954 - Mathieusche Funktionen und Sphäroidfunktionen. Springer.

Oberbeck, A., 1979 - Ueber die Wärmeleitung der Flussigkeiten bei Berücksichtigung der Strömungen infolge von Temperaturdifferenzen. Ann. Physik 7, 271-292.

Rosenthal, A.J. & Lindzen, R.S., 1983 a - Instabilities in a stratified fluid having one critical level. Part I: Results. J. Atmos. Sci. 40, 509-520.

Rosenthal, A.J. & Lindzen, R.S., 1983 b - Instabilities in a stratified fluid having one critical level. Part II: Explanation of gravity wave instabilities using the concept of overreflection. J. Atmos. Sci. 40, 521-529.

Teitelbaum, H. & Sidi, C., 1979 - Discontinuity formation due to gravity wave propagation in a shear layer. Phys. Fluids 22, 209-213.

Teitelbaum, H. & Kelder, H., 1985 - Critical levels in a jet type flow. J. Fluid Mech. 159, 227-240.

Tolstoy, I. & Pan, P., 1970 - Simplified atmospheric models and the properties of long-period internal and surface gravity waves. J. Atmos. Sci. 27, 31-50.

Van Duin, C.A. & Kelder, H., 1982 - Reflection properties of internal gravity waves incident upon a hyperbolic tangent shear layer. J. Fluid Mech. 120, 505-521.

Van Duin, C.A. & Kelder, H., 1986 - Internal gravity waves in shear flows at large Reynolds number. J. Fluid Mech. 169, 293-306.

CHAPTER 5

PROPAGATION OF INTERNAL GRAVITY WAVES IN A ROTATING FLUID WITH SHEAR FLOW

5.1. Introduction

Over the past few decades many authors have studied the phenomenon of overreflection in a stratified fluid in shear motion. Overreflection is defined as an absolute value of the reflection coefficient larger than one. This phenomenon in relation to internal gravity waves in a non-rotating fluid was first treated by Jones (1968). He found that overreflection does not occur for values of the Richardson number at the critical level Ri_c larger than 0.115 at least when the transmitted wave is a propagating one. Eltayeb & McKenzie (1975), in an analytical study, obtained results for a broken-line wind profile, consisting of a constant shear layer imbedded in two layers of constant wind. Van Duin & Kelder (1982) later made an analytical study of a hyperbolic tangent or Epstein type wind profile, a model without the disadvantages of discontinuities in the derivatives of the wind field. They found that overreflection does not occur for values of Ri_c above 0.132. Acheson (1976) has given a review of the phenomenon of overreflection. Teitelbaum & Kelder (1985) have studied a jet-type wind profile in which the wave encounters more than one critical level. Their results show that in that case no overreflection occurs for values of the minimum Richardson number of the flow above 0.143 while Ri_c = 0.169. Instability as a consequence of multiple overreflection was investigated in a series of papers (Rosenthal & Lindzen part I, part II, 1983; Lindzen & Rosenthal part III, 1983 and Lindzen & Barker, 1985).

In a rotating fluid there appear to be three critical levels. These levels are encountered by the wave at the heights where the background wind is such that the Doppler-shifted frequency is equal to zero or to plus or minus the rotation frequency. Hence rotation causes a splitting up of critical levels analogous to the Zeeman splitting of spectral lines by a magnetic field. This case was also first studied by Jones (1967). He found that the vertical flux

of angular momentum is constant everywhere in the fluid except at the critical levels. By ignoring rotation the resulting solutions appear to be in error only at levels where the Doppler-shifted and Coriolis frequencies are comparable. He considered numerically only the case that the Richardson number is equal to 1.
Yamanaka & Tanaka (1984) made an analytical study of these so-called internal inertio-gravity waves (waves with frequencies near the rotation frequency) in an incompressible rotating flow with a constant shear layer. Tai (1983) has investigated the relation between overreflection and instability in a rotating fluid. In all the studies cited above the Boussinesq approximation was used. Our objective was to calculate overreflection and overtransmission in a rotating fluid and to analyse the sensitivity of the results for certain approximations.
A large part of this chapter has been published previously (Teitelbaum, Kelder & Van Duin, 1986).

5.2. The wave equations in the Boussinesq and hydrostatic approximations

5.2.1. The Boussinesq approximation

We shall adopt a model which was used by Jones (1967). The model is a planar one. Viscosity is ignored and the adiabatic equation of state is used. The fluid is supposed to be rotating around the vertical axis (the z-axis) with angular velocity f/2 where f is the so-called Coriolis parameter.
A geostrophically balanced mean flow $U_o(z)$ is directed along the x-axis and varies with z only

$$\frac{\partial p_o}{\partial y} = - f U_o \rho_o. \qquad (5.1)$$

This is valid only if $U_o/R < f$ where R is the radius of the earth.

The background pressure p_o is assumed to be in hydrostatic balance

$$\frac{\partial p_o}{\partial z} = - g \rho_o. \qquad (5.2)$$

The vertical structure of the background density ρ_o is given by

$$\frac{\partial \ln \rho_o}{\partial z} = -\beta. \tag{5.3}$$

From (5.1), (5.2) and (5.3) it follows that

$$\frac{\partial \ell n \rho_o}{\partial y} = \frac{f}{g}\left(\frac{dU_o}{dz} - \beta U_o\right). \tag{5.4}$$

Assuming that there are perturbations superimposed on the background state so small that the equations may be linearized, applying the Boussinesq approximation and looking for solutions of the form

$$A(x,y,z,t) = \hat{A}(z)\, e^{i(\omega t - kx - ly)},$$

Jones (1967) derived for the vertical velocity w the following ordinary differential equation:

$$\left[\Omega^2 - f^2\right]\frac{d^2w}{dz^2} + \left[2ifl - \frac{2f^2k}{\Omega}\right]\frac{dU_o}{dz}\frac{dw}{dz} +$$

$$\left\{\left[N^2 - \Omega^2\right](k^2 + l^2) + \frac{2ilkf}{\Omega}\left(\frac{dU_o}{dz}\right)^2 +\right.$$

$$\left.\left[\Omega k + ifl\right]\frac{d^2U_o}{dz^2}\right\} w = 0. \tag{5.5}$$

Here k and l are horizontal wave numbers, ω is the wave frequency and Ω is the Doppler-shifted frequency defined by $\Omega := \omega - k U_o$.
N^2 is the square of the Brunt-Väisälä frequency, in the Boussinesq approximation equal to $g\beta$.

In deriving equation (5.5) Jones apparently assumed that β is independent of y on the basis of the Boussinesq approximation. However, we shall show that this is a separate assumption based on an approximation which is independent of the Boussinesq one.

Differentiating equations (5.3) and (5.4) with respect to y and z respectively two expressions are obtained which must be equal i.e.

$$-\frac{\partial\beta}{\partial y} = \frac{f}{g}\left(\frac{d^2U_o}{dz^2} - \frac{\partial\beta}{\partial z}U_o - \beta\frac{dU_o}{dz}\right). \tag{5.6}$$

A Taylor expansion of $\beta(z,y)$ around y=0 yields

$$\beta(z,y) = \beta(z,o) + \frac{\partial\beta}{\partial y}(z,o)\,y + \ldots . \tag{5.7}$$

The y dependency of $\beta(z,y)$ may be neglected if

$$|\beta(z,o)| \gg \left|\frac{\partial\beta}{\partial y}(z,o)\,y\right| . \tag{5.8}$$

An expression for $\frac{\partial\beta}{\partial y}(z,o)$ is obtained from (5.6)

$$-\frac{\partial\beta}{\partial y}(z,o) = \frac{f}{g}\left(\frac{d^2U_o}{dz^2} - \frac{\partial\beta(z,o)}{\partial z}U_o(z) - \beta(z,o)\frac{dU_o}{dz}\right) . \tag{5.9}$$

Hence the condition (5.8) is satisfied if

$$\left|\frac{f}{g}\left(\frac{d^2U_o}{dz^2} - \frac{\partial\beta}{\partial z}(z,o)\,U_o(z) - \beta(z,o)\frac{dU_o}{dz}\right)y\right| \ll |\beta(z,o)| . \tag{5.10}$$

At a good approximation it is true that

$$\left|\frac{d^2U_o}{dz^2}\right| \ll \left|\beta(z,o)\frac{dU_o}{dz}\right|$$

and (5.11)

$$\left|\frac{\partial\beta}{\partial z}(z,o)\,U_o(z)\right| \ll \left|\beta(z,o)\frac{dU_o}{dz}\right|.$$

From (5.10) and (5.11) the condition that the y dependency of $\beta(z,y)$ may be neglected reads

$$\left|\frac{f}{g}\frac{dU_o}{dz}\,y\right| \ll 1 . \tag{5.12}$$

A reasonable value of $\frac{dU_o}{dz}$ is 10^{-4} s^{-1}, then $\left|\frac{f}{g}\frac{dU_o}{dz}\right|$ is smaller than 10^{-1} if y is smaller than 1000 km.

In the derivation of Jones it is assumed that the Coriolis frequency is constant, this is the so-called f-plane approximation. In fact f is a function of y and this approximation gives another restriction on the range of values of y.

5.2.2. The hydrostatic approximation

The hydrostatic approximation was discussed in chapter 4. Here we will include rotation. The derivation of the equations is only slightly different from that in chapter 4 and therefore the formulas will be given without comments. The symbols have the same meaning as before (cf. 4.1).

The momentum equations now reads

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w^*\frac{\partial u}{\partial z^*} + \frac{\partial \Phi}{\partial x} - fv = 0 \qquad (5.13)$$

and

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w^*\frac{\partial v}{\partial z^*} + \frac{\partial \Phi}{\partial y} + fu = 0. \qquad (5.14)$$

The continuity equation becomes

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z^*} - \frac{w^*}{H} = 0. \qquad (5.15)$$

Finally the thermodynamic energy equation reads

$$\left(\frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w^*\frac{\partial}{\partial z^*}\right)\frac{\partial \Phi}{\partial z^*} + N^2 w^* = 0\ , \qquad (5.16)$$

where N is the Brunt-Väisälä frequency.

Suppose a development around a static solution $U_o(z^*), \Phi_o(z^*, y)$ that is

$$\begin{aligned} U &= U_o + u^1 \\ v &= 0 + v^1 \\ w^* &= 0 + w \\ \Phi &= \Phi_o + \psi \end{aligned}$$

The static solution has to satisfy the geostrophic balance

$$\frac{\partial \Phi_o}{\partial y} = - f U_o(z^*) , \tag{5.17}$$

and the hydrostatic condition

$$\frac{\partial \Phi_o}{\partial z^*} = \frac{RT_o}{H} . \tag{5.18}$$

The first order perturbations have to satisfy the following set of equations (the prime and asterisk are from now on omitted)

$$\frac{\partial u}{\partial t} + U_o \frac{\partial u}{\partial x} + w \frac{dU_o}{dz} + \frac{\partial \psi}{\partial x} - fv = 0 , \tag{5.19}$$

$$\frac{\partial v}{\partial t} + U_o \frac{\partial v}{\partial x} + \frac{\partial \psi}{\partial y} + fu = 0 , \tag{5.20}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} - \frac{w}{H} = 0 , \tag{5.21}$$

$$(\frac{\partial}{\partial t} + U_o \frac{\partial}{\partial x}) \frac{\partial \psi}{\partial z} - f \frac{dU_o}{dz} v + N^2 w = 0 . \tag{5.22}$$

The only y-dependent coefficient is the Brunt-Väisälä frequency N, given by

$$N^2 = \frac{R}{H} (\frac{dT_o}{dz} + \frac{\kappa T_o}{H}) . \tag{5.23}$$

We will derive a condition under which it is only slightly dependent on y.

From (5.23) it is clear that N^2 depends on y through a dependency of T_o only. Combining the equations (5.17) and (5.18) yields

$$\frac{\partial To}{\partial y} = - \frac{H}{R} \frac{f dU_o}{dz} . \tag{5.24}$$

A Taylor expansion of $T_o(z,y)$ around y=0 yields

$$T_o(z,y) = T_o(z,o) + \frac{\partial T_o}{\partial z}(z,o)\, y + \ldots . \tag{5.25}$$

The y dependency of T_o may be ignored if

$$|\frac{\partial T_o}{\partial y}(z,o)\, y| \ll T_o(z,o). \tag{5.26}$$

Equation (5.24) yields an expression for $\frac{\partial T_o}{\partial y}(z,o)$:

$$\frac{\partial T_o}{\partial y}(z,o) = - \frac{H}{R} f \frac{dU_o}{dz}. \tag{5.27}$$

The scale height H was defined in chapter 4 and is taken in this case as $H = \frac{R\, T_o(z,o)}{g}$. This implies that condition (5.26) reads

$$\left|\frac{f}{g} \frac{dU_o}{dz} y\right| << 1. \tag{5.28}$$

With this condition the y-dependency of the Brunt-Väisälä frequency may be ignored.
Notice that the condition (5.28) is exactly the same condition as Jones used (cf. 5.2.1). Here the f-plane approximation has also been used implicitly. Hence the remark made after (5.12) is also valid here.
With this approximation solutions can be found of the form

$$f(x,y,z,t) = \hat{f}(z)\, e^{i(\omega t - kx - ly) + \frac{z}{2H}}.$$

If this form of solution is inserted into equations (5.19) - (5.22), an ordinary differential equation for the vertical velocity can be derived.

$$(\Omega^2 - f^2) \frac{d^2w}{dz^2} + \left[2ifl - \frac{2f^2k}{\Omega}\right] \frac{dU_o}{dz} \frac{dw}{dz}$$

$$+ \left\{ N^2 (k^2 + l^2) + \frac{2iflk \left(\frac{dU_o}{dz}\right)^2}{\Omega} + (ilf + \Omega k) \frac{d^2U_o}{dz^2} + \frac{k(\Omega^2 + f^2) \frac{dU_o}{dz}}{H\Omega} \right.$$

$$\left. - \frac{\Omega^2 - f^2}{4H^2} \right\} w = 0. \tag{5.29}$$

The difference between this and the Boussinesq equation (5.5) lies in the term between the parentheses. In (5.29) the term contains $N^2 (k^2 + l^2)$ instead of $(N^2 - \Omega^2) (k^2 + l^2)$ as in the Boussinesq case. As was argued in chapter 4 the hydrostatic approximation is in fact a low-frequency approximation that is $\Omega^2 << N^2$ and this explains the difference. Furthermore in equation (5.29) two extra terms appear: $\frac{k (\Omega^2 + f^2)}{H\Omega} \frac{dU_o}{dz}$ and $\frac{\Omega^2 - f^2}{4H^2}$.

These terms take into account the background variation and are ignored in the Boussinesq approximation (scale height H tends to infinity).

5.3. The singularities of the equation

For a background wind profile that is a monotonic function of height, Jones's equation and the hydrostatic equation both have three singularities, one for zero shifted frequency as in a non-rotating system and the other two where the Doppler-shifted frequency equals $\pm f$. The first singularity was studied by Booker & Bretherton (1967) and will be referred to hereinafter as the B.B. singularity. The other two were studied by Jones (1967) and will be referred to as the J. singularities.
When compressibility is taken into account two new singularities appear and, moreover, the B.B. singularity is split up into two as was shown by Eltayeb & Kandaswamy (1979). In the hydrostatic approximation compressibility is also taken into account but these singularities do not appear as the acoustic branch of the acoustic-gravity waves is suppressed (see chapter 4.1). Hence the acoustic waves cause these extra singularities.
It has been shown by Jones (1967) that in a rotating system the vertical flux of angular momentum is independent of height, except at the critical levels. A way to obtain such a fixed quantity is outlined below for a second-order linear differential equation in the Helmholtz form

$$\psi'' + q(z)\,\psi = 0 \,, \tag{5.30}$$

where ' denotes differentation to z.

The derivative of the Wronskian of the solutions ψ and its complex conjugate ψ^* have to satisfy the following equality

$$\frac{d}{dz}\left[\psi\psi'^{*} - \psi^{*}\psi'\right] = -2\,\mathrm{Im}\,q(z)\,|\psi|^2. \tag{5.31}$$

If the function q(z) takes on real values only and $|\psi|^2$ is finite then the Wronskian $W(\psi,\psi^*)$ is independent of height.

A differential equation of the form

$$\phi'' + a(z)\,\phi' + b(z)\,\phi = 0 \,, \tag{5.32}$$

can be transformed in the form (5.30) with the Liouville transformation

$$\phi(z) = \psi(z)\; e^{-\frac{1}{2}\int_0^z a(z')dz'} \,. \tag{5.33}$$

The analogue of (5.31) for equation (5.32) reads

$$\frac{d}{dz}\left[\; e^{\int_0^z \mathrm{Re}\; a(z')\,dz'} \left\{ -\,\mathrm{Im}\; a(z)\;|\phi|^2 + W(\phi,\phi^*)\right\}\right] =$$

$$-\,2\;\mathrm{Im}\left(b(z) - \frac{\dot{a}(z)}{2} - \frac{a^2(z)}{4}\right) e^{\int_0^z \mathrm{Re}\; a(z')dz'} |\phi|^2 . \tag{5.34}$$

Hence if

$$\mathrm{Im}\left(b(z) - \frac{\dot{a}(z)}{2} - \frac{a^2(z)}{4}\right) = 0 \,, \tag{5.35}$$

then the quantity G defined by

$$G: = e^{\int_0^z \mathrm{Re}\; a(z')dz'} \left\{-\mathrm{Im}\; a(z)\;|\phi|^2 + W(\phi,\phi^*)\right\} \,, \tag{5.36}$$

is independent of height.

Straightforward calculations show that equations (5.5) and (5.29) both satisfy condition (5.35) and the quantity G turns out to be the same. G can be written as

$$G = \mathrm{Re}\left[\frac{f\;l\;U'(z)}{\Omega^2}\;|w|^2 + i\left(1 - \frac{f^2}{\Omega^2}\right) w'w^*\right]. \tag{5.37}$$

Tai (1983) has proved that the quantity G is proportional to the vertical flux of angular momentum.
The changes in this flux across the critical levels have been calculated by Tai (1983) for Jones's equation.
This author has shown that for the J. critical levels there is a jump.
However, for the B.B.-critical level this is not always the case.

The B.B.-critical level at $z=z_c$ corresponds with a regular singularity in the differential equation. With the Frobenius method a solution can be found of the form

$$W = E_1 [\gamma_1 \log \zeta_c \, \zeta_c^3 \, W_{c_1}(\zeta_c) + W_{c_2}(\zeta_c)] + E_2 \, \zeta_c^3 \, W_{c_1}(\zeta_c) , \tag{5.38}$$

where $\zeta_c = z - z_c$, E_1 and E_2 are arbitrary constants, W_{c_1} and W_{c_2} are analytic functions which equal unity when $\zeta_c=0$, and

$$\gamma_1 = \frac{k^2+l^2}{3f^2} \left(2N \frac{dN}{dz} - \frac{dU_o}{dz} \frac{d^2U_o}{dz^2} - N^2 \frac{\frac{d^2U_o}{dz^2}}{\frac{dU_o}{dz}}\right) \Big|_{z=z_c} . \tag{5.39}$$

If $\gamma_1=0$, z_c is an apparent singularity. In Tai (1983) it is shown that, with the branch cuts taken as discussed there, the jump in the flux is equal to

$$G_{c_-} - G_{c_+} = \pi \left(1 + \frac{l^2}{k^2}\right) |E_1|^2 \gamma_2 , \tag{5.40}$$

where G_{c_-} is the value of G for $z < z_c$ and G_{c_+} is the value of G for $z > z_c$, and

$$\gamma_2 = - \operatorname{sgn}\left(k \frac{dU_o}{dz}\right) \left[\frac{2N\frac{dN}{dz}}{(\frac{dU_o}{dz})^2} - \frac{\frac{d^2U_o}{dz^2}}{\frac{dU_o}{dz}} - \frac{N^2\frac{d^2U_o}{dz^2}}{(\frac{dU_o}{dz})^3} \right] \Big|_{z=z_c} . \tag{5.41}$$

Hence, if $\gamma_2=0$ then there is no jump in the vertical flux of angular momentum at the B.B.-critical level.

In the hydrostatic equation the J. singularities are the same as in Jones's equation. In fact these singularities are determined by the coefficients of the first derivative and these are the same in both equations.

However, the condition for having an apparent singularity in the hydrostatic equation is not the same. Working analogously, we found this to be

$$2N \frac{dN}{dz} - \frac{dU_o}{dz} \frac{d^2U_o}{dz^2} - \frac{N^2 \frac{d^2U_o}{dz^2}}{\frac{dU_o}{dz}} + \frac{1}{H} \left(N^2 - (\frac{dU_o}{dz})^2\right) \Big|_{z=z_c} = 0. \tag{5.42}$$

Compared with (5.41) an extra term $\frac{1}{H}\,(N^2 - (\frac{dU_o}{dz})^2)$ appears. This is due to taking into account the background variation of the density.
It will be shown that the difference in the type of singularity can cause differences in the reflection and transmission coefficients.

5.4. The method of solution

To solve the hydrostatic or Jones's equation analytically is relatively difficult. Yamanaka & Tanaka (1984) were able to express solutions in terms of the hypergeometric functions for a special case with a constant wind shear and applying both the hydrostatic and the Boussinesq approximations. For a reflection and transmission problem these solutions are not useful because these solutions far from the critical levels cannot asymptotically be approximated by plane waves. Here we take a numerical approach. We have chosen a tangent hyperbolic background wind profile with the B.B.-critical level at the inflection point.
Accordingly, Jones's equation has there an apparent singularity while the hydrostatic equation has a logarithmic one.
We used two different numerical approaches. In the first we introduce a small imaginary component for the frequency, the singularities have then shifted from the real axis and integration along the real axis is possible. We start the calculations above the critical levels, far enough to give a good approximation of a constant background wind, with an upward-propagating plane wave. The equation is then integrated backwards by the method used by Bulrisch & Stoer (1966) down to a distance below the critical levels where the solutions are again closely approximated by plane waves. Only then can the solution be broken down to give an incident and reflected wave and the reflection and transmission coefficients calculated. The integration step length is automatically changed when the desired accuracy is not obtained. For example, the minimum step length near the critical levels can be as low as 10^{-10} of the total integration path with in total not more that 500 steps.
The second numerical method is similar to the first but instead of using an imaginary component for the frequency, a Frobenius expansion is made at each critical level to match the solutions by analytic continuation.

Owing to the accuracy and efficiency of the integration routine only two terms in the Frobenius expansion were needed. The difference between the results obtained by the two different methods turns out to be less than 1%.

5.5. Results

With Jones's equation overreflection and overtransmission were found even for values of Ri_c above 0.25. Ri_c is the Richardson number at the B.B.- critical level located at the inflection point and it is also the minimal value of the Richardson number for the tangent hyperbolic wind profile.
Tai (1983) found, by analytical methods, overreflection regardless of the value of the Richardson number when only two of the critical levels were present. In contradiction to our results the same author found by numerical methods that overreflection for layers with three critical levels is possible only for values of the Richardson number lower than 0.2402. This result was obtained subject to the condition that the transmitted wave was a propagating one. We suppose that the difference with Tai's result is due to the fact that he has not fully explored the influence of the different parameters.
We found that overreflection and overtransmission are very sensitive to the value of the ratio $\hat{f}$ between the Coriolis parameter f and the wave frequency ω . For certain values of $\hat{f}$ resonant overreflection occurs; i.e. the reflection and transmission coefficients tend to infinity and there is no incident wave. This resonant overreflection occurs even for values of Ri_c higher than 0.25. Figure 5.1 shows the values of $\hat{f}$ as a function of Ri_c for which resonant overreflection occurs.
The results obtained using the hydrostatic equation differ from the above. For example we found no resonant overreflection. Overreflection and overtransmission again occurred for values of Ri_c larger than 0.25.
Some of the results obtained using the two equations are illustrated in figure 5.2. For two values of the Richardson number at the critical level at the inflection point of the hyperbolic tangent wind profile, i.e. $Ri_c = 0.30$ and $Ri_c = 0.65$, the values of the reflection coefficient as a function of $\hat{f}$ are demonstrated for both approximations. For small $\hat{f}$'s the reflection coefficients tend to the same values.

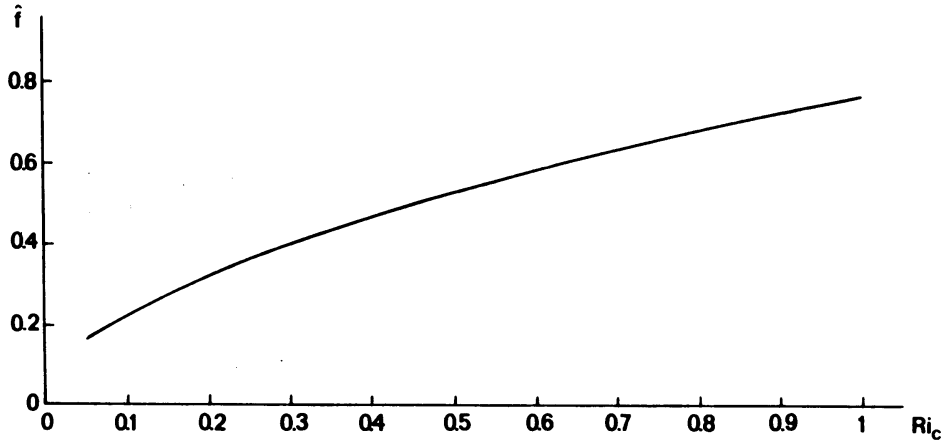

Figure 5.1 The values of $\hat{f}$, as a function of Ri_c, for which resonant overreflection occurs.

Hydrostatic approximation	———	$Ri_c = 0.30$
	- - - - -	$Ri_c = 0.65$
Boussinesq approximation	-.-.-	$Ri_c = 0.30$
	.....	$Ri_c = 0.65$

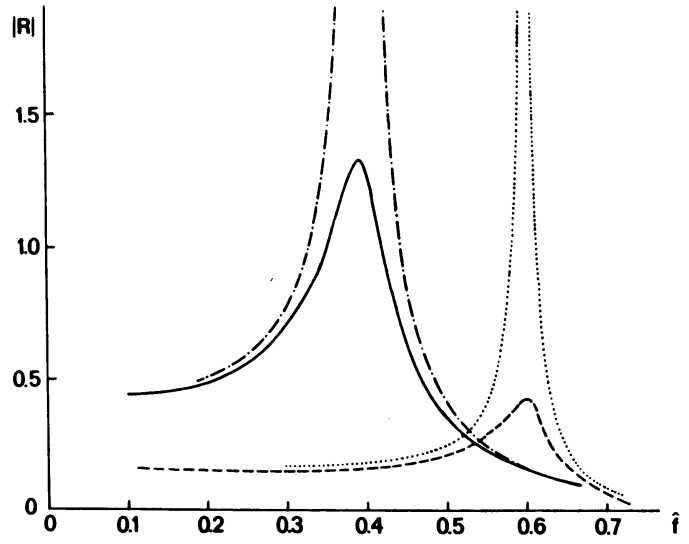

Figure 5.2 Variation of the absolute value of the reflection coefficient $|R|$, as a function of the ratio f between the angular velocity f and the wave frequency ω .Conventions as in the previous figure.

Calculating the reflection coefficients without rotation with the help of formulas given in Van Duin & Kelder (1982) values are obtained of $|R| = 0.43$ for $Ri_c = 0.30$ and $|R| = 0.13$ for $Ri_c = 0.65$. From figure 5.2 it is clear that these values correspond to the limiting values for $\hat{f} \to 0$. This is reasonable because this limit means that the rotation frequency becomes much smaller than the wave frequency.
For larger values of $\hat{f}$ the difference between the two approximations increases with a maximum at the value for which resonant overreflection occurs in the Boussinesq approximation. From figure 5.2 it can be inferred that the range of values of $\hat{f}$ for which overreflection occurs decreases with increasing Ri_c.
The difference in the results is due to the fact that, at the B.B. critical level, Jones's equation has an apparent singularity and the hydrostatic equation a logarithmic one. This difference is caused by the variation with height of the background density which is neglected in the Boussinesq approximation. This is confirmed by taking, in the hydrostatic approximation, the limit for very large values of the scale height. At this limit, expression (5.42) becomes identical to (5.41) and the singularity likewise becomes an apparent one. Numerically the results of Jones's equation are then obtained, as should be expected.

5.6. Conclusions

Overreflection and overtransmission of a gravity wave propagating in a layer with a sheared wind field in an inviscid, adiabatic, stratified and rotating fluid was studied. Our results show that, in contradiction to the non-rotating case, overreflection occurs for values of Ri_c above 0.25.
The transmission and reflection coefficients appear to be strongly dependent on the ratio between the rotation frequency and the wave frequency. For the range of this ratio for which overreflection occurs, different results are obtained using the Boussinesq and hydrostatic approximations. For example, resonant overreflection was found in the Boussinesq approximation only.

The results obtained here indicate that care must be taken in applying certain approximations to atmospheric problems. We are inclined to the opinion that the hydrostatic approximation is better in describing the propagation of gravity waves through a critical layer certainly if the rotation of the fluid is taken into account.

5.7. References

Acheson, D.J., 1976 - On over-reflexion. J. Fluid Mech. 77, 433-472.

Booker, J.R. & Bretherton, F.P., 1967 - The critical layer for internal gravity waves in a shear flow. J. Fluid Mech. 27, 513-539.

Bulrisch, R. & Stoer, J., 1966 - Numerical treatment of ordinary differential equations. Num. Math. 8, 1-13.

Eltayeb, I.A. & McKenzie, J.F., 1975 - Critical-level behaviour and wave amplification of a gravity wave incident upon a shear layer. J. Fluid Mech. 72, 661-671.

Eltayeb, I.A. & Kandaswamy, R., 1979 - Critical levels in a vertically rotating compressible atmosphere. Quart. J. Appl. Math. 32, 313-330.

Jones, W.L., 1967 - Propagation of internal gravity waves in fluids with shear flow and rotation. J. Fluid Mech. 30, 439-448.

Jones, W.L., 1968 - Reflexion and stability of waves in stably stratified fluid with shear flow: a numerical study. J. Fluid Mech. 34, 609-624.

Lindzen, R.S. & Rosenthal, A.J., 1983 - Instabilities in a stratified fluid having one critical level. Part III: Kelvin-Helmholtz instabilities as overreflected waves. J. Atmos. Sci. 40, 530-542.

Lindzen, R.S. & Barker, J.W., 1985 - Instability and wave over-reflection in stably stratified shear flow. J. Fluid Mech. 151, 189-217.

Rosenthal, A.J. & Lindzen, R.S., 1983 - Instabilities in a stratified fluid having one critical level. Part I: Results. J. Atmos. Sci. 40, 509-520.

Rosenthal, A.J. & Lindzen, R.S., 1983 - Instabilities in a stratified fluid having one critical level. Part II: Explanation of gravity wave instabilities using the concept of overreflection. J. Atmos. Sci. 40, 521-529.

Tai, C.K., 1983 - Overreflection and instability. Dyn. Atm. Oc. 7, 147-165.

Teitelbaum, H. & Kelder, H., 1985 - Critical levels in a jet-type flow. J. Fluid Mech. 159, 227-240.

Teitelbaum, H., Kelder, H. & Van Duin, C.A., 1986 - Propagation of internal gravity waves in a rotating fluid with shear flow. Accepted by J. atm. terr. Phys.

Van Duin, C.A. & Kelder, H., 1982 - Reflection properties of internal gravity waves incident upon a hyperbolic tangent shear layer. J. Fluid Mech. 120, 505-521.

Yamanaka, M.D. & Tanaka, H., 1984 - Propagation and Breakdown of Internal Inertio-Gravity Waves near Critical Levels in the Middle Atmosphere. J. Met. Soc. Japan, 62, 1-16.

CHAPTER 6

ON THE REFLECTION OF TIDES IN THE UPPER ATMOSPHERE

6.1. Introduction

Observations of winds in the upper atmosphere indicate the presence of strong tidal movements. The classical theory of atmospheric tides, that is without the inclusion of mean winds and meridional temperature gradients, is able to explain some major features of the observations (Chapman & Lindzen, 1969; Kato, 1980).
Some problems, however, are still not completely solved. One of them is the structure of the semidiurnal tide at and above 100 km height. Observations show a strong influence of the higher order modes (Fellous et al., 1975; Ahmed & Roper, 1983), whereas the theory predicts the dominance of the 2.2 mode (Chapman & Lindzen, 1970). One way to solve this discrepancy is indicated by Lindzen & Hong (1974). They introduce a zonal wind profile, and the interaction of the semidiurnal modes with the wind leads to a weakening of the 2.2 mode and an enhancement of the higher order modes.
Another way to explain the observations is to take into account the reflection of the semidiurnal modes, Fellous et al. (1975), Stening (1977), Stening et al. (1978). In this paper we will pursue the theory of reflection on a realistic temperature profile. The same problem was treated earlier, e.g. by Teitelbaum (1973), whose solution of the reflection problem leads, via the Mathieu equation, to qualitative results. However, the method followed here provides a basis for more comprehensive conclusions, because the results are quantitative. Vial et al. (1985) used a numerical method. They also discussed the influence of different standard profiles on the reflection of tidal waves.

The scheme of this chapter is as follows. In section 6.2 a short account is given on the derivation of the differential equation describing the vertical propagation of the semidiurnal modes.

In section 6.3 the calculation of the reflection coefficients for a class of refractive index profiles is discussed. In section 6.4 the method is applied to the reflection of the semidiurnal modes.

6.2. The differential equation for the vertical propagation

An extensive discussion of the derivation of the differential equation for tidal waves can be found in Chapman & Lindzen (1969) and Kato (1980). Here, only a sketch will be given. It is convenient to introduce the quantity

$$G(z, \theta, \phi, t) = - \frac{1}{\gamma p_o} \frac{dp}{dt} , \tag{6.1}$$

where p_o is the unperturbed pressure,
p is the tidal perturbation in pressure,
$\gamma = \frac{c_p}{c_v}$ is the ratio of the specific heats,
t is time,
z is height,
θ is colatitude ,
ϕ is longitude .

Suppose that the longitude and time dependency is periodic and, moreover, that the restriction is made to consider only tides migrating with the sun; then G can be written as

$$G(z, \theta, \phi, t) = \sum_{\sigma=1}^{\infty} G^{\sigma}(z, \theta) \, e^{i\sigma(\frac{2\pi}{24} t + \phi)} , \tag{6.2}$$

with $\sigma = 1, 2, 3, ...$ and t is in hours.

The function $G^{\sigma}(z, \theta)$ has to satisfy a linear second-order partial differential equation that can be solved by separation of variables. Therefore,

$$G^{\sigma}(z, \theta) = \sum_{n=1}^{\infty} L_n{}^{\sigma}(z) \, \Theta_n{}^{\sigma}(\theta) , \tag{6.3}$$

where each Θ_n^{σ} satisfies the Laplace's Tidal Equation:

$$F(\Theta_n^{\sigma}(\theta)) = - \frac{4a^2\omega^2}{gh_n^{\sigma}} \Theta_n^{\sigma}(\theta) , \tag{6.3a}$$

where

$$F = \frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \left(\frac{\sin\theta}{f^2-\cos^2\theta} \frac{\partial}{\partial\theta}\right) - \frac{1}{f^2-\cos^2\theta} \left(\frac{\sigma}{f} \frac{f^2+\cos^2\theta}{f^2-\cos^2\theta} + \frac{\sigma^2}{\sin^2\theta}\right),$$

a is the radius of the earth and h_n^{σ} is the separation constant, called the equivalent depth.

The solutions Θ_n^{σ} are known as Hough functions. The $L_n^{\sigma}(z)$ satisfy

$$H \frac{d^2 L_n^{\sigma}}{dz^2} + \left(\frac{dH}{dz} - 1\right) \frac{dL_n^{\sigma}}{dz} + \frac{1}{h_n^{\sigma}} \left(\frac{dH}{dz} + \kappa\right) L_n^{\sigma} = + \frac{\kappa J_n^{\sigma}}{\gamma g H h_n^{\sigma}} , \tag{6.4}$$

where H = scale height,

$\kappa = \frac{\gamma-1}{\gamma}$,

J_n^{σ} = thermal excitation of the tides.

The equation (6.4) is often called the vertical structure equation.
This equation can be transformed into a Helmholtz equation by introducing the reduced height x defined by

$$x = \int_o^z \frac{dz'}{H(z')}$$

and by using instead of L_n^{σ} the function y_n^{σ} defined by

$$y_n^{\sigma}(x) = e^{-\frac{x}{2}} L_n^{\sigma}(z(x)) .$$

The Helmholtz equation reads then

$$\frac{d^2 y_n^{\sigma}}{dx^2} + \left[\frac{1}{h_n^{\sigma}} \left(\kappa H + \frac{dH}{dx}\right) - \frac{1}{4}\right] y_n^{\sigma} = \frac{\kappa J_n^{\sigma} e^{-\frac{x}{2}}}{\gamma g h_n^{\sigma}} . \tag{6.5}$$

The particular solution of this equation can be written as

$$y_n^{\sigma}(x) = \frac{\kappa}{\gamma g h_n^{\sigma}} \int_0^{\infty} G_n^{\sigma}(x,x_o)\, J_n^{\sigma}(x_o)\, e^{-\frac{x_o}{2}}\, dx_o . \qquad (6.6)$$

where $G_n^{\sigma}(x,x_o)$ is the Green's function for the mode and the temperature profile in question.

The excitation, however, takes place predominantly below 50 km height (see fig. 6.1). Therefore, tidal waves above this height can be considered as freely propagating waves.

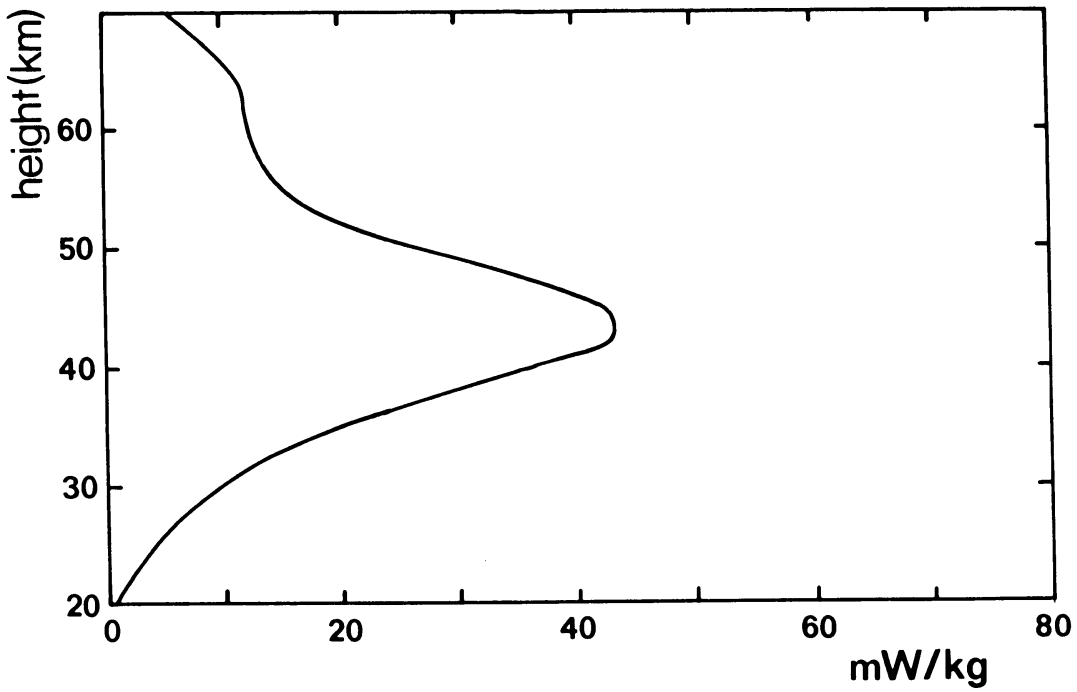

Figure 6.1 Height profile of the Hough component of the heating for the J_2^2 mode (Groves, 1982).

The propagation equation then takes the form

$$\frac{d^2 y_n^{\sigma}}{dx^2} + \left[\frac{1}{h_n^{\sigma}}\left(\kappa H + \frac{dH}{dx}\right) - \frac{1}{4}\right] y_n^{\sigma} = 0 . \qquad (6.7)$$

Let $\mu(x)$ be the invariant of differential equation (6.7), that is

$$\mu(x) = \frac{1}{h_n^{\sigma}} \left(\kappa H + \frac{dH}{dx}\right) - \frac{1}{4} . \tag{6.8}$$

μ can be interpreted as the square of the refractive index. If the refractive index is constant, then the differential equation (6.7) has two kinds of elementary solutions, namely if $\mu > 0$ then

$$y_n^{\sigma} = Ae^{i\mu^{1/2}x} + Be^{-i\mu^{1/2}x} , \tag{6.9}$$

and for $\mu < 0$

$$y_n^{\sigma} = Ce^{(-\mu)^{1/2}x} + Be^{(-\mu)^{1/2}x} . \tag{6.10}$$

If μ is not constant, the solutions of (6.7) generally cannot be expressed in elementary functions. In that case the WKB approximation can be a good substitute. Anyway, if strong reflections occur, the WKB method fails (Nayfeh, 1973). Because strong reflections can be expected we looked for a method not suffering from this restriction.

First we will discuss the behaviour of the refractive index for tidal waves in a model atmosphere. As most important model we have chosen the Mean Reference Atmosphere of the CIRA (1972). This model is suitable for heights between 30 and 500 km and is independent of season. It is considered to be representative for middle latitudes. The height dependence of μ in (6.8) appears in the term

$$\kappa H + \frac{dH}{dx} , \tag{6.11}$$

formed by the scale height and its derivative. In figure 6.2 the value of (6.11) is given as function of the reduced height for the Mean Reference Atmosphere. The flattening of the curve between 30 and 50 km and above 200 km is characteristic. From this curve the profiles of the refractive index for each of the tidal modes can be determined. We shall restrict ourselves to the semidiurnal modes. In figure 6.3 the graph of μ is drawn for the 2.2 mode.

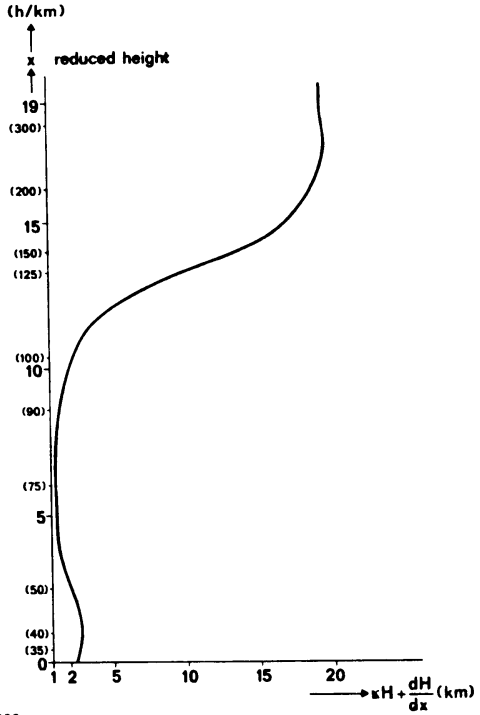

Figure 6.2. $\kappa H + \frac{dH}{dx}$ versus the reduced height for the Mean Reference Atmosphere of CIRA 1972. The real height is given in parentheses.

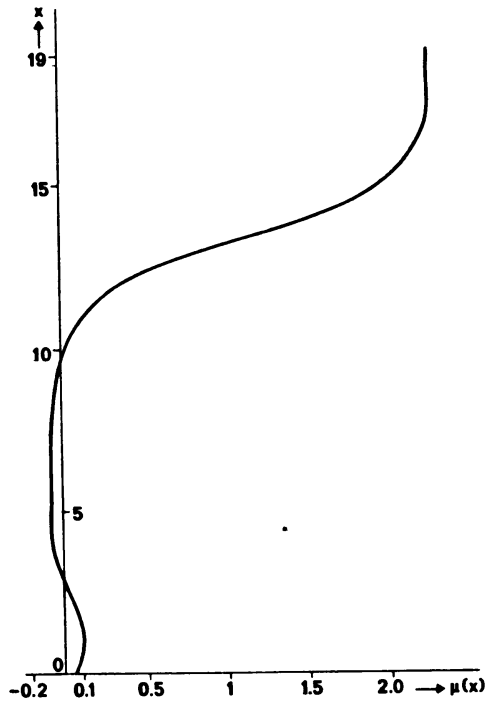

Figure 6.3. The value of the square of the refractive index for the 2.2 mode versus the reduced height.

The curve for the 2.2 mode shows negative values over a large height interval. At these heights only evanescent waves can exist, and strong partial reflection can be expected.
In the next section an analytical approach to describe reflection in an inhomogeneous medium will be outlined.

6.3. Reflection of waves in an inhomogeneous medium

An analytical description of wave propagation in an inhomogeneous medium is possible for a limited number of profiles of the refractive index, see e.g. Ginzburg (1961) and Brekhovskikh (1980). One of these is the Epstein-Eckart profile, Epstein (1930) and Eckart (1930), in formula

$$\varepsilon(x) = a_1 + a_2 \tanh \frac{x}{2l} + a_3 \cosh^{-2} \frac{x}{2l} , \tag{6.12}$$

where ε is the square of the refractive index, a_1, a_2 and a_3 are constants and l is a length scale.

There are two obvious special cases of this profile namely

$$a_3 = 0, \ \varepsilon(z) = a_1 + a_2 \tanh \frac{x}{2l} , \tag{6.13}$$

called the transitional Epstein profile and

$$a_2 = 0, \ \varepsilon(z) = a_1 + a_3 \cosh^{-2} \frac{x}{2l} , \tag{6.14}$$

called the symmetric Epstein profile.

A characteristic of the profiles is that for large values of x, ε takes on constant values, while in the neighbourhood of $z = 0$ the greatest change in value takes place in a domain with a width determined by l.
If the square of the refractive index is described by the Epstein profile then the wave equation can be solved analytically. Here, only a sketch of the solutions will be given; the complete treatment of the problem can be found in the original literature (Epstein, 1930; Eckart, 1930).

Epstein and Eckart start with

$$\frac{d^2y}{dx^2} + \varepsilon(x)\, y(x) = 0 \ . \tag{6.15}$$

Define k_1 and k_2 as

$$k_1^{\,2} = \lim_{x \to -\infty} \varepsilon(x) \ , \tag{6.16}$$

and

$$k_2^{\,2} = \lim_{x \to +\infty} \varepsilon(x) \ . \tag{6.17}$$

The equation (6.15) can be transformed into a hypergeometrical differential equation.
By putting

$$\rho = \frac{x}{l} \ , \ u = e^{\rho} \text{ and } y = (1 + u)^d\, u^a\, f, \tag{6.18}$$

one obtains

$$u(1 + u)\frac{d^2f}{du^2} + \left[(2d + 2a + 1)u + 2a + 1\right]\frac{df}{du} + (a+b+d)(a-b+d)f = 0 \ , \tag{6.19}$$

where $a = il\sqrt{a_1 - a_2}$,

$b = il\sqrt{a_1 + a_2}$,

$d = \tfrac{1}{2} - \tfrac{1}{2}\sqrt{1 + 16l^2a_3}$.

The second order linear differential equation in (6.19) has three regular singularities at 0, -1 and ∞ respectively and is known as the hypergeometrical differential equation. In the neighbourhood of the singularities solutions exist in the form of power series. Hence we obtain the following solutions f_1 to f_6:

around $u = 0$	f_1, f_2 valid for $\lvert u\rvert < 1$,
around $u = -1$	f_3, f_4 valid for $\lvert u+1\rvert < 1$,
"around" $u = \infty$	f_5, f_6 valid for $\lvert u\rvert > 1$.

Explicit expressions for the functions f_1 to f_6 can be found in e.g. Erdelyi et al. (1953).

For these solutions analytical continuation can be found, and solutions can be constructed that are everywhere defined. The way of analytical continuation has to be found on physical grounds. The starting point, however, was a second order linear differential equation, which can have no more than 2 independent solutions. Therefore, linear relations exist between the solutions. These relations, called circuit relations, have been calculated by Gauss (Epstein, 1930). One of these relations is

$$f_5 = Af_1 + Bf_2 \ , \tag{6.20}$$

where A and B are gamma functions and are given in Erdelyi et al. (1953).

To interpret relation (6.20) we will consider the asymptotic behaviour of the solutions involved. Let y_1, ..., y_6 be the solutions corresponding to f_1, ..., f_6 respectively. For very large negative values of x, that means, according to (6.18) for u approaching zero:

$$y_1(x) = (1+u)^d \, u^a \, f_1(u)(:) \, e^{ik_1x},$$

which can be interpreted as an incoming plane wave from below. The same asymptotic procedure applied to y_2 yields

$$y_2(:) \, e^{-ik_1x} \ ,$$

and this can be interpreted as a downward reflected plane wave.
If x, and according to (6.18) also u, take on very large positive values, then

$$y_5(:) \, e^{ik_2x} \ ,$$

which represents an upward propagating plane wave. Hence, equation (6.20) can be interpreted as a relation between the incoming, the reflected and the transmitted waves. This implies that a reflection coefficient R can be defined as the ratio between the asymptotic values of the reflected and incoming waves.

The result is:

$$R = \frac{\Gamma(2a)\,\Gamma(-a-b+d)\,\Gamma(-a-b-d+1)}{\Gamma(-2a)\,\Gamma(a-b+d)\,\Gamma(a-b-d+1)} \,. \tag{6.21}$$

We have obtained an expression for the reflection coefficient in terms of gamma functions, that are simpler than the hypergeometrical functions
The work of Epstein was later extended by many authors. One extension, that of Rawer (1939), is very suitable for describing tidal wave propagation. The profile of Rawer has the form

$$\varepsilon(x) = -\left(\frac{dp}{dx}\right)^2 \left[K_1 + \tfrac{1}{4} - K_2 \frac{\exp p(x)}{1+\exp p(x)} - K_3 \frac{\exp p(x)}{(1+\exp p(x))^2}\right] + \tag{6.22}$$

$$+ \tfrac{1}{2}\left[\frac{d^3p}{dx^3} \Big/ \frac{dp}{dx} - \frac{3}{2}\left(\frac{d^2p}{dx^2} \Big/ \frac{dp}{dx}\right)^2\right],$$

where K_1, K_2 and K_3 are constants
and

$$p(x) = x_0\, x + A_1 \frac{e^{sx}}{a_1+e^{sx}} + \frac{B_1}{s} \ln(1 + b_1 e^{sx}). \tag{6.23}$$

x_0, A_1, B_1, a_1, b_1 and s are also constants.

If $A_1 = B_1 = 0$ then equation (6.22) again represents the Epstein profile.

For the Rawer profile (6.22), the reflection coefficient equals

$$R = \frac{\Gamma(\gamma-1)\,\Gamma(1-\beta)\,\Gamma(1+\alpha-\gamma)}{\Gamma(1-\gamma)\,\Gamma(\gamma-\beta)\,\Gamma(\alpha)} \,, \tag{6.24}$$

with

$$\alpha = \tfrac{1}{2}\left\{1 + (1+4k_1)^{\frac{1}{2}} + (1+4K_3)^{\frac{1}{2}} - \left[1+4(K_1-K_2)\right]^{\frac{1}{2}}\right\} \,,$$

$$\beta = \tfrac{1}{2}\left\{1 + (1+4K_1)^{\frac{1}{2}} + (1+4K_3)^{\frac{1}{2}} + \left[1+4(K_1-K_2)\right]^{\frac{1}{2}}\right\} \,,$$

$$\gamma = 1 + (1+4K_1)^{\frac{1}{2}} \,.$$

Formulae (6.22) and (6.23) define a large class of profiles of the refractive index for which analytical solutions are possible. In the next section we will look for the best approximation of the refractive index profiles of the tidal waves by (6.22) and then apply the results of this chapter.

6.4. Results

The approximation of the refractive index profiles of the semidiurnal tidal modes by the functions described in (6.22) and (6.23) requires some calculations. The approximation is done in least squares sense. Two examples of the curves obtained are shown in figures 6.4 and 6.5. In figure 6.4 the square of the refractive index of the 2.2 mode and its approximation are drawn.

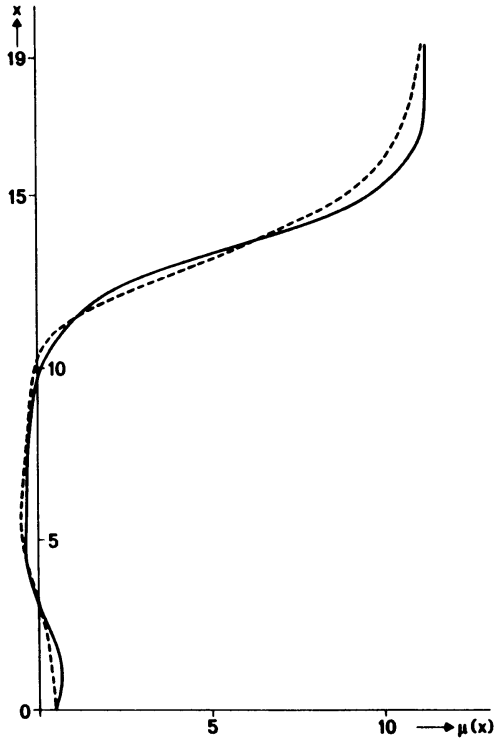

Figure 6.4. Solid line: the square of the refractive index of the 2.2 mode. Dashed line: the approximation by Rawer's function.

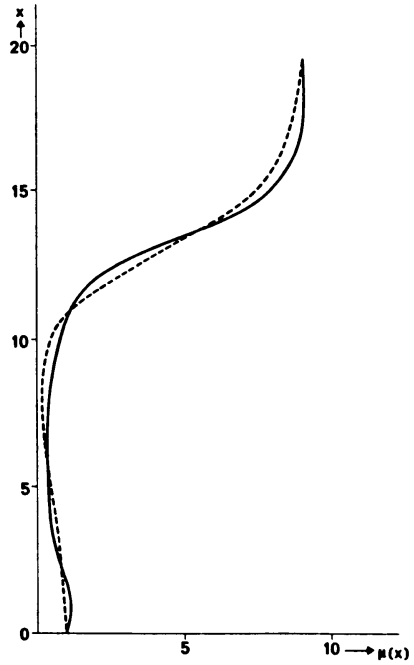

Figure 6.5 Solid line: the square of the refractive index for the 2.4 mode. Dashed line: the approximation by Rawer's function.

In figure 6.5 the same has been done for the 2.4 mode. The agreement between the approximations and the profiles computed from the Mean Reference Atmosphere is rather good. Hence it is permissible to substitute the approximating functions into the Helmholtz equation for the propagation of the tidal modes. The reflection coefficients can then be calculated. The absolute values of some reflection coefficients are given in Table 6.1:

Table 6.1

mode	|R|
2.2	0.98
2.3	0.31
2.4	0.08

These values indicate that for this temperature profile the 2.2 mode is nearly completely reflected, that the 2.3 mode is also strongly reflected, but the 2.4 mode is only slightly reflected. Higher order modes are so weakly reflected that their reflected wave can be ignored.

The temperature profile, however, is dependent on season, latitude and exospheric temperature. To test these dependencies we calculated the temperature profiles for July and January from the CIRA 1972, for 50°N and an exospheric temperature of 1000 K.
The absolute values of the corresponding reflection coefficient are summarized in Table 6.2.

Table 6.2

mode	|R|	
	July	January
2.2	0.99	0.89
2.3	0.88	0.42
2.4	0.74	0.19

Hence, in summer the reflection, especially of the higher modes, is substantially higher. The influence of the latitude was determined by calculating the values of $|R|$ for January and 1000 K exospheric temperature for 0° (the equator) too. The values of $|R|$ were somewhat higher: 0.96, 0.55 and 0.30 for the 2.2, 2.3 and 2.4 modes respectively. Finally, we varied the exospheric temperature and calculated the reflection coefficient for January for 50°N and an exospheric temperature of 1500 K. We found 0.81, 0.06 and 0.02 for the 2.2, 2.3 and 2.4 modes respectively.

Hence, a higher exospheric temperature leads to lower values of the reflection coefficient, especially for the higher order modes.

6.5. Conclusions

In this paper analytical solutions has been given for the equation describing the vertical propagation of the semidiurnal modes through the mesosphere and lower thermosphere.
These analytical solutions offer the possiblity to calculate the wave function for every height. We restricted ourselves to the calculation of the reflection coefficients. It was shown that the reflection coefficients can be expressed in gammafunctions.
The Mean Reference Atmosphere of the CIRA 1972 was used as a model. For this model it turns out that the 2.2 semidiurnal tidal mode is nearly completely reflected whereas the 2.3 mode is reflected for 31% and the 2.4 mode for only 8%. The reflection of the higher order modes appears to be negligible. This implies that with increasing height the character of the semidiurnal tidal wave gradually shifts to the higher order modes. In the mesosphere interference by a superposition of tidal modes will be important. This agrees with observations see e.g. Stening (1977, 1978) and Ahmed & Roper (1983), but is not predicted by the classical theory of atmospheric tides (Chapman & Lindzen, 1970). The results are valid for the Mean Reference Atmosphere that is thought to be representative for middle latitudes and is independent of season.
However, the dependence on season, latitude and exospheric temperature proved to be considerable as was demonstrated by calculating the value of $|R|$ for other reference atmospheres. Consequently, this sensitivity of the reflection for the temperature profile causes a large variability in tidal structure in the mesosphere and thermosphere.

6.6. References

Ahmed, M.I. & Roper, R.G., 1983 - The diurnal and semidiurnal oscillations in meteor winds over Atlanta. J. atm. terr. Phys. 45, 181-192.

Brekhovskikh, L.M., 1980 - Waves in Layered Media. Academic Press.

Chapman, S. & Lindzen, R.S. 1970 - Atmospheric Tides. Reidel, Dordrecht, Holland.

CIRA, 1972 - COSPAR International Reference Atmosphere. Akademie Verlag, Berlin.

Eckart, C., 1930 - The penetration of a potential barrier by electrons. Phys. Rev. 35, 1303-1309.

Epstein, P.S., 1930 - Reflection of waves in an inhomogeneous absorbing medium. Proc. Nat. Ac. Sci. 16, 627-630.

Erdelyi, A., Magnus, W., Oberhettinger, F. & Tricomi, F.G., 1953 - Higher Transcendental functions. Vol. I, McGraw-Hill, New York.

Fellous, J.L., Bernard, R., Glas, M., Massebeuf, M. & Spizzichino, A, 1975 - A study of the variation of atmospheric tides in meteor zone. J. atm. terr. Phys. 37, 1511-1524.

Ginzburg, V.L., 1961 - Propagation of Electromagnetic Waves in Plasma. Gordon & Breach, New York.

Groves, G.V., 1982 - Hough components of ozone-heating. J. atm. terr. Phys. 44, 111-121.

Kato, S., 1980 - Dynamics of the upper atmosphere. Developments in Earth and Planetary Sciences 01, Reidel, Dordrecht.

Lindzen, R.S. & Hong, S., 1974 - Effects of mean winds and horizontal temperature gradients on solar and lunar semidiurnal tides in the atmosphere. J. Atmos. Sci. 31, 1421-1466.

Nayfeh, A., 1973 - Perturbation Methods. Wiley & Sons, New York.

Rawer, K., 1939 - Elektrische Wellen in einem Geschichteten Medium. Ann. der Physik 35, 385-416.

Stening, R.J., 1977 - Analysis of contributions to ionospheric dynamo currents from e.m.f.'s at different latitudes. Planet. Space Sci. 25, 587-594.

Stening, R.J., Meek, C.E., Manson, A.H. & Stephenson, D.G., 1978 - Winds and Wave Motions to 110 km and Midlatitudes. VI Tidal, Gravity and Planetary Waves, 1976. J. Atmos. Sci. 35, 2194-2204.

Teitelbaum, H., 1973 - Réflexions des marées induites par le profile de température entre 50 et 110 km d'altitude. Ann. Géophys. 29, 395-397.

Vial, F., Fellous, J.L. & Teitelbaum, H., 1985 - Tidal vertical structure and temperature profiles in the lower thermosphere. Ann. Geophysicae 3, 313-318.