

On the interaction between thermal tides and gravity waves in the middle atmosphere

Dissertation

zur Erlangung des Doktorgrades
in den Naturwissenschaften

vorgelegt beim Fachbereich 11
Geowissenschaften / Geographie
der Johann Wolfgang Goethe - Universität
in Frankfurt am Main

von

Fabian Senf

aus Leipzig

Frankfurt (2012)

(D30)

vom Fachbereich Geowissenschaften / Geographie

der Johann Wolfgang Goethe - Universität als Dissertation angenommen

Dekan: Prof. Dr. Andreas Junge

Gutachter: Prof. Dr. Ulrich Achatz (Universität Frankfurt)

Prof. Dr. Erich Becker (Universität Rostock)

Datum der Disputation:

Zusammenfassung: Über die Wechselwirkung von thermischen Gezeiten und Schwerewellen in der mittleren Atmosphäre

Einleitung

Die mittlere Atmosphäre ist ein besonderer Teil des Klimasystems. Sie ist nicht im direkten Kontakt mit der Erdoberfläche und hat doch großen Einfluss auf uns. Der chemische Bestandteil Ozon, welcher in der mittleren Atmosphäre maximale Konzentrationen hat, sorgt zum Beispiel für eine Absorption von gefährlicher UV-Strahlung und bildet somit einen Schutzschild für uns Menschen. Aber auch dynamische Mechanismen, die zur Kopplung zwischen der unteren und mittleren Atmosphäre führen, sind von besonderem wissenschaftlichen Interesse. Atmosphärische Wellen, welche über einen großen Bereich von Raum- und Zeitskalen existieren, breiten sich von der unteren zur mittleren Atmosphäre aus und tragen in einem bedeutendem Maße zum Austausch von Energie und Impuls zwischen beiden Bereichen bei. Das Spektrum der Wellenaktivität reicht dabei von planetaren Wellen, wie thermischen Gezeiten, bis hin zu mesoskaligen Phänomenen, wie brechenden Schwerewellen. Doch gerade die enorme Spannweite von Skalen, involviert in die Dynamik der mittleren Atmosphäre, macht ihre detaillierte Beschreibung zu einer anspruchsvollen Aufgabe.

Im zeitlichen Mittel hat die mittlere Atmosphäre eine besondere Temperaturstruktur. Ausgehend von der Tropopause bei ca. 10 km bis 15 km nimmt die Temperatur in der darüber liegenden Luftschicht, der Stratosphäre, durch die Absorption von solarer Strahlung zu. Die Stratopause bei ca. 50 km bis 60 km markiert mit mehr als 10°C die Höhe des Temperatur-Maximums. Oberhalb davon, in der Mesosphäre, sinkt die Lufttemperatur wieder, da die Konzentration der mit der Solarstrahlung wechselwirkenden Konstituenten abnimmt. Besonders außergewöhnlich ist die Temperaturabnahme in der Mesosphäre während der Sommersonnenwenden. In der polaren Sommermesopause bei ca. 85 km, also gerade dort, wo während des polaren Tages permanent die Sonne scheint, werden die kältesten Temperaturen von unter -140°C erreicht. Dieses bemerkenswerte Phänomen führt zur Bildung von nachleuchtenden Eiswolken in dieser Höhe, die während des Sommers von mittleren und hohen Breiten aus beobachtet werden können.

Die globale residuelle Zirkulation, welche den Sommer- mit dem Winterpol verbindet, trägt wesentlich zur thermischen Struktur der mittleren Atmosphäre bei. Im Mittel steigt die Luft über der tropischen Tropopause auf und wird in der Stratosphäre zum Winterpol transportiert. Diese großräumigen Luftbewegungen bewirken anomale Ozonkonzentrationen in hohen Breiten und sind bekannt als Brewer-Dobson-Zirkulation. In der Mesosphäre reicht die residuelle Zirkulation bis zur Sommerhemisphäre und führt zu einem Aufsteigen der Luft über dem Sommerpol und einem Absinken über dem Winterpol. Durch die Druckabnahme dehnen sich aufsteigende Luftpakete aus und werden somit adiabatisch gekühlt, während absinkende Luftpakete sich durch die wirkende Kompression erwärmen. So entwickeln sich eine dynamisch induzierte, kalte Sommermesopause und eine warme Winterstratopause.

In der Mesosphäre kann die globale residuelle Zirkulation hauptsächlich auf das Brechen von Schwerewellen zurückgeführt werden. Diese Wellen, bei denen Luftpakete in einer stabil geschichteten Atmosphäre um ihre Gleichgewichtslage schwingen, werden überwiegend durch Gebirge, Gewitterwolken, Fronten und dynamische Instabilitäten in der Troposphäre angeregt. Sie transportieren Energie und Impuls aufwärts und beeinflussen entscheidend die mittlere Energie- und Impulsbilanz in den Regionen, wo sie durch Instabilitäten in kleinere turbulente Strukturen zerfallen. Die Vielzahl von brechenden Schwerewellen-Paketen erzeugt eine mittlere Kraft und treibt so die globale

residuelle Zirkulation an.

Ein weiterer wichtiger Kopplungsmechanismus zwischen unterer und mittlerer Atmosphäre sind thermische Gezeiten. Diese großskaligen atmosphärischen Wellen werden hauptsächlich durch die Erwärmung der Luft auf der Tagseite der Erde verursacht. Entscheidend sind dabei die Absorption von solarer Strahlung von Ozon und Wasserdampf sowie die Freisetzung von latenter Wärme. Thermische Gezeiten sind periodische, großräumige Fluktuationen in Temperatur und Wind. Sie sind harmonische Oszillationen mit Perioden von 24 h, 12 h und höheren Vielfachen und dominieren mit Amplituden von bis zu 30 K und 50 m/s die Tages-Variabilität in der oberen Mesosphäre.

Die Wechselwirkung zwischen Schwerewellen und Gezeiten wird in der Mesosphäre wichtig. Zum einen verändern die periodischen Strömungsmuster der Gezeiten die Ausbreitungsbedingungen der Schwerewellen. Schwerewellen-Pakete werden in ihren Eigenschaften zeitlich verändert, destabilisiert und in ihrem Brechungsvorgang beeinflusst. Zum anderen entstehen durch die periodische Modulation von Schwerewellen periodische Kräfte und Heizraten, welche auf die thermischen Gezeiten zurückwirken. Dieser Effekt ist besonders in den Regionen wichtig, wo Schwerewellen-Pakete durch ihren turbulenten Zerfall, Impuls und Energie auf die großskalige Strömung übertragen. Dort erhöht sich zusätzlich der Einfluss der kleinskaligen turbulenten Mischung, welcher sich destruktiv auf die Gezeitenwellen auswirkt.

Die vielen Raum- und Zeitskalen, die bei der Wechselwirkung von Schwerewellen und Gezeiten wichtig sind, machen es unmöglich, alle involvierten Prozesse mithilfe von detaillierten numerischen Simulationen zu ergründen. In der Vergangenheit wurden deshalb die großskaligen Gezeitenwellen mithilfe von linearen und nicht-linearen Modellen untersucht, die kleinskaligen Prozesse aber, wie Schwerewellen und Turbulenz, durch Parametrisierungen nur näherungsweise beschrieben. Es konnte gezeigt werden, dass eine durch Gezeitenwinde periodisch modulierte Schwerewellen-Kraft die Phasenlage der Gezeiten verschiebt und so die vertikale Wellenlänge der Gezeiten verringert. Allerdings wurden verschiedene Effekte auf die Gezeitenamplitude beobachtet und auf unterschiedliche Formulierungen in den Anfangsspektren und Dissipationsmechanismen der parametrisierten Schwerewellen zurückgeführt. Da ein Großteil der bisherigen Studien konventionelle Schwerewellen-Parametrisierungen mit sehr restriktiven Annahmen benutzt, ist es jedoch äußerst fraglich, inwieweit eine solche Beschreibung für das Ergründen der Schwerewellen-Gezeiten-Wechselwirkung hinreicht. Für diese Arbeit wurde deswegen das Ziel gesetzt, die Defizite der konventionellen Beschreibung der Schwerewellen-Ausbreitung in realistischen Gezeiten zu quantifizieren.

Die Methodik: „Ray Tracing“

Die detaillierte mathematische Beschreibung eines Schwerewellen-Feldes in einer sich verändernden Hintergrundströmung ist für realistische Fälle sehr aufwendig. Eine klare Interpretation der dynamischen Mechanismen für Veränderungen im Wellenfeld ist schwierig. Wie in dieser Arbeit ausführlich diskutiert, kann aber mithilfe der sogenannten multiplen Skalen-Asymptotik eine Reduktion auf die wesentliche Dynamik der Wellen vollzogen werden. Es werden nur lokal-monochromatische Wellenpakete betrachtet, denen man eine Frequenz, einen Wellenvektor und eine Amplitude als einzigste Eigenschaften zuordnet. Die kleinskaligen Fluktuationen der Wellenphasen werden nicht betrachtet. Aus dem Wellenpaket wird eine Art Wellenteilchen, das die genannten Eigenschaften besitzt und sich mit seiner Gruppengeschwindigkeit fortbewegt. Durch den Pfad des Wellenteilchens entsteht ein Strahl, woraus sich der Name Strahlmodell oder „Ray Tracing“ ableitet. Entlang eines Strahls können sich die Eigenschaften des Wellenfeldes

abhängig von den Symmetrien der Hintergrundströmung ändern. Analog zu konventionellen Schwerewellen-Parametrisierungen, die meist als vertikale Säulen aufgebaut sind, basiert die „Ray Tracing“-Methode auf einschneidenden Näherungen. Zum Beispiel können Reflexionen, Überlagerungen und nichtlineare Wechselwirkungen von mehreren Wellenfeldern, und der turbulente Zerfall nur unzureichend beschrieben werden. Im Gegensatz zu konventionellen Schwerewellen-Parametrisierungen werden aber mit der „Ray Tracing“-Methode sowohl die Effekte durch zeitliche Variabilität als auch die Effekte durch horizontale Inhomogenitäten in der Hintergrundströmung berücksichtigt. Deshalb kann ein Strahlmodell wichtige Einblicke in den Ablauf dynamischer Prozesse während der Schwerewellen-Gezeiten-Wechselwirkung bieten und kann die Fehler durch die strengeren Annahmen in den konventionellen Parametrisierungen quantifizieren.

Die „Ray Tracing“-Methode wurde in der Vergangenheit benutzt, um klein- bis mesoskalige Schwerewellen-Pakete in einfachen periodischen Hintergrundströmungen zu untersuchen. Es stellte sich heraus, dass die zeitliche Veränderung der großskaligen Strömungen eine Modulation der Schwerewellen-Frequenz verursacht, während die räumliche Änderung auf die Komponenten des Wellenvektors wirkt und eine Refraktion der Welle hervorruft. In der vorliegenden Arbeit wird die „Ray Tracing“-Methode auf die Problemstellung der Schwerewellen-Gezeiten-Wechselwirkung angewendet. Vom Autor wurde das globale Strahlmodell RAPAGI¹ entwickelt, welches im besonderen Maße den Einfluss der Gezeitenwellen auf nach oben propagierende Schwerewellen-Pakete berücksichtigt. Realistische Gezeitenfelder werden aus dem Zirkulationsmodell HAMMONIA² entnommen. Dieses am Max-Planck-Institut für Meteorologie in Hamburg entwickelte Modell ist eines der führenden Klimamodelle der mittleren Atmosphäre. Ganztages-Schwingungen in Wind und Temperatur aus HAMMONIA werden als Ausbreitungsbedingung für Schwerewellen-Pakete festgelegt und in das globale Strahlmodell RAPAGI übergeben.

Im Folgenden werden drei verschiedene Experimente für RAPAGI beschrieben, um den Einfluss von zeitlicher Periodizität und horizontaler Inhomogenität der Gezeitenwellen auf die Schwerewellen-Dynamik zu untersuchen. Für alle Experimente wird ein Schwerewellen-Ensemble aus 14 Komponenten benutzt, bei denen die einzelnen Anteile homogene und stationäre Anregungsbedingungen in einer Höhe von ca. 20 km haben. Das Ensemble wird so gewählt, dass zum einen der mittlere Schwerewellen-Effekt im Mesopausen-Bereich hinreichend gut repräsentiert ist, zum anderen jedoch die Interpretation der zeitlichen und räumlichen Variabilität in den Schwerewellen-Eigenschaften vereinfacht wird. Da konventionelle Parametrisierungen von Schwerewellen typischerweise keine zeitlichen Änderungen und auch nur vertikale Gradienten der großskaligen Strömungen berücksichtigen, wird ein globales „Ray Tracing“-Experiment erstellt, welches sowohl die zeitliche als auch die horizontale Abhängigkeit der HAMMONIA-Felder ignoriert und somit den konventionellen Ansatz nachstellt („TS“-Experiment). Im zweiten Experiment werden hauptsächlich horizontale Gradienten des Hintergrunds vernachlässigt und die horizontale Propagation der Schwerewellen verhindert („noREF“-Experiment). Wellenpakete breiten sich nur vertikal aus, haben aber im Gegensatz zum „TS“-Experiment eine endliche Ausbreitungsgeschwindigkeit und spüren demnach die zeitliche Veränderung der Gezeitenwellen. Im dritten Experiment wird das Strahlmodell RAPAGI ohne Approximationen benutzt („full“-Experiment). Es dient als Referenz. Die Modulation der Schwerewellen-Frequenz und die endliche vertikale Informationsausbreitung tritt in den beiden Experimenten „noREF“ und „full“ auf, während die Refraktion

¹ „**RAy** **PA**rameterization of **G**ravty-wave **I**mpacts“

² „**HA**mbug **MO**del of the **N**eutral and **I**onized **A**tmosphere“

der horizontalen Schwerewellenzahlen und die horizontale Ausbreitung der Wellenpakete nur im Experiment „full“ möglich ist.

Aus den globalen „Ray Tracing“-Simulationen wird für das erwähnte Schwerewellen-Ensemble eine zeitlich mittlere und eine periodische Kraftwirkung abgeschätzt. Sie ist besonders dann wichtig, wenn der turbulente Zerfall der Wellen einsetzt. Aber gerade dieser Prozess wird weder in konventionellen Parametrisierungen noch in der „Ray Tracing“-Methode realistisch beschrieben. Eine einfache Abschätzung der turbulenten Dämpfung wird durch das Sättigungs-Schema gegeben, welches trotz seiner Defizite bei konventionellen Parametrisierungen verbreitet ist und auch in der vorliegenden Arbeit verwendet wird. Darin wird das ungehinderte Anwachsen der Schwerewellen-Amplituden beschränkt, wenn die Schwelle zur konvektiven Instabilität innerhalb des betrachteten Wellenfeldes überschritten wird.

Ergebnisse

Die vorliegende Arbeit besteht aus zwei wichtigen Teilbereichen. Zum einen wird eine konsistente, theoretische Basis für die Schwerewellen-Gezeiten-Wechselwirkung geschaffen und zum anderen werden numerische Simulations-Experimente zur Quantifizierung der Mechanismen während der Wechselwirkung durchgeführt.

Im ersten Teil wird die großskalige Dynamik aus den fundamentalen Erhaltungssätzen abgeleitet und schrittweise wichtige Näherungen und ihre zugrundeliegenden Annahmen erläutert. Die mittleren dynamischen und thermodynamischen Effekte durch subskalige Phänomene, welche Schwerewellen und die durch sie erzeugte Turbulenz beinhalten, werden detailliert hergeleitet. Es wird gezeigt, dass die Formulierung der Kräfte und Heizraten von der Art der Filterung der großskaligen Dynamik abhängt. Sowohl die ortsgebundene Eulersche Mittelung als auch die mitbewegte Lagrangesche Mittelung werden untersucht und die wichtigsten Unterschiede zwischen beiden herausgestellt. Als unmittelbare Folge der Unterschiede zwischen beiden Formulierungen ergibt sich, dass die Art der Subskalen-Parametrisierung von der Interpretation der aufgelösten, mittleren Strömung abhängt. Mithilfe der multiplen Skalen-Asymptotik und der WKB-Theorie³ wird für die Dynamik von Schwerewellen ein System von Gleichungen abgeleitet, welches die großskalige Veränderung von Schwerewellen-Eigenschaften wie zum Beispiel Frequenz und Wellenvektor beschreibt. Bei der Ableitung dieser „Ray Tracing“-Gleichungen werden die benutzten Annahmen intensiv diskutiert. Speziell kann eine nahtlose Beschreibung von Schwerewellen vom nicht-hydrostatischen Grenzfall bis hin zur Trägheits-Schwerewelle nur durch eine künstliche Hybridisierung der zu Grunde liegenden approximativen Gleichungen erreicht werden. Dies folgt aus dem unterschiedlichen Skalenverhalten der verschiedenen Schwerewellen-Regime und wurde in der bisherigen Literatur nicht hinreichend betrachtet.

Im zweiten Teil werden mithilfe der „Ray Tracing“-Methode realistische Experimente zur Schwerewellen-Gezeiten-Wechselwirkung durchgeführt. In diesem Rahmen wird zum ersten Mal die Gültigkeit der Annahmen von konventionellen Schwerewellen-Parametrisierungen untersucht. Zwei wichtige Effekte der thermischen Gezeiten und des klimatologischen Strömungsfeldes werden herausgestellt. Zum einen verursachen die zeitlich periodischen Schwankungen der Gezeitenwinde eine Modulation der ortsgebundenen Frequenz eines Schwerewellen-Paketes. Zum anderen werden durch horizontale Gradienten in der gesamten Hintergrundströmung Schwerewellen-Pakete horizontal abgelenkt. Die entstehende Refraktion der horizontalen Wellenzahl kann die horizontale Struktur des Schwerewellen-Feldes entscheidend verändern und zu einer Drehung sowie zur Stauchung

³Wentzel-Kramers-Brillouin-Theorie

bzw. Streckung der Welle führen. Bei der horizontalen Refraktion ist der Einfluss des zonal gemittelten, klimatologischen Zonalwindes besonders wichtig. Durch beide Effekte, Frequenz-Modulation und horizontale Refraktion, werden Schwerewellen-Pakete so modifiziert, dass sich auch die Kräfte und Heizraten, die bei ihrem turbulenten Zerfall entstehen, signifikant verändern. Eine genaue Erläuterung wird im Folgenden gegeben.

Durch die Frequenz-Modulation erfolgt ein Austausch von Energie zwischen der Hintergrund-Strömung und dem Wellenpaket. Im gleichen Maße wie die Frequenz wird auch die horizontale Phasengeschwindigkeit der Schwerewellen, die der Quotient aus Frequenz und horizontaler Wellenzahl ist, verändert. Die Modulation der Phasengeschwindigkeit hat bedeutende Auswirkungen auf das Verständnis der Schwerewellen-Gezeiten-Wechselwirkung. Im konventionellen Bild werden eine Momentaufnahme des vertikalen Verlaufs des Hintergrundwindes und eine als konstant angenommene horizontale Phasengeschwindigkeit gegenüber gestellt. Schwerewellen, die gegen den Wind laufen, haben beste Ausbreitungsbedingungen. Nähert sich die Geschwindigkeit des Hintergrundwindes an die Phasengeschwindigkeit der Wellen an, schrumpft das betrachtete Wellenpaket vertikal bei gleichzeitigem Amplitudenwachstum. Setzt sich dieser Vorgang fort, wird das Wellenpaket destabilisiert und bricht an einer sogenannten kritischen Schicht, an der Wind und Phasengeschwindigkeit den gleichen Wert haben. In der Brechungshöhe tritt die maximale durch Schwerewellen induzierte Kraftwirkung und Heizrate auf. Im Bereich der oberen Mesosphäre verändern aber die periodischen Schwankungen der Gezeiten signifikant die Hintergrundströmung. Die konventionelle Theorie der Schwerewellen-Gezeiten-Wechselwirkung sagt dann voraus, dass sich transiente, kritische Schichten bilden, die mit der Gezeitenphase nach unten wandern und somit eine stark lokalisierte, periodische Kraft entsteht. Diese Abschätzung ist aber problematisch, da hier weder die endliche vertikale Ausbreitungsgeschwindigkeit der Schwerewellen noch der Einfluss der Zeitabhängigkeit der Gezeiten auf Schwerewellen-Eigenschaften in Betracht gezogen werden. Die periodischen Schwankungen modulieren die horizontale Schwerewellen-Phasengeschwindigkeit so, dass sie dem Gezeitenwind folgt. Es entsteht eine Art Ausweichbewegung, die die kritische Filterung verringert und die Kraftwirkung auf einen größeren Höhenbereich verteilt. Die periodische Schwerewellen-Kraft, und demnach auch der Einfluss der durch Gezeiten modulierten Schwerewellen zurück auf die Gezeitenwellen selbst, ist in Wirklichkeit kleiner als von der konventionellen Abschätzung erwartet. Mithilfe des „noREF“-Experiments wird für die realistischen Strömungsfelder aus dem Modell HAMMONIA im Vergleich zum „TS“-Experiment eine Verringerung der periodischen Schwerewellen-Kraft im Mittel um 30 % gefunden. Lokal wird die Kraft aber um bis zu 90 % reduziert.

Ebenso wird die Wirkung von horizontalen Inhomogenitäten der Hintergrundströmung auf Schwerewellen in konventionellen Parametrisierungen ausgeblendet. Auch diese Effekte werden mithilfe der „Ray Tracing“-Methode untersucht und Veränderungen in den Schwerewellen-Eigenschaften sowie in den durch Schwerewellen induzierten mittleren und periodischen Kräften quantifiziert. Die horizontale Refraktion verändert die Richtung und Länge des horizontalen Wellenvektors der Schwerewellen und zusammen mit der horizontalen Wellenausbreitung auch die Verteilung der Wellenpakete im Raum. Vornehmlich haben die meridionalen Gradienten des mittleren zonalen Windes Einfluss auf die Wellen-Refraktion. Wellenpakete, die gegen die Hintergrundströmung laufen, werden in der Stratosphäre in das Maximum der Wind-Jets hineingeführt. Durch dieses Verhalten wird analog zum Fermatschen Prinzip der geometrischen Optik die Laufzeit der Schwerewellen in der mittleren Atmosphäre minimiert. Es entsteht eine Fokussierung von Schwerewellen-Feldern, bei gleichzeitiger Zunahme der meridionalen Wellenzahl. Durch diesen Effekt wird der Betrag des horizontalen Wellenvektors im Experiment

„full“ im Mittel um ca. 10 % vergrößert. Dies hat zur Folge, dass sich die horizontale Phasengeschwindigkeit im Mittel um den selben Betrag verringert und infolgedessen Schwerewellen-Pakete früher destabilisiert werden und zerfallen. Außerdem wird der Schwerewellen-Impulsfluss reduziert und die mittlere Kraft auf die Hintergrundströmung im „full“-Experiment im Mittel um ca. 20 % bis 30 % verkleinert. Die Reduktion der zeitlich mittleren Schwerewellen-Kraft bewirkt unmittelbar auch eine Verringerung der periodischen Schwerewellen-Kraft um einen vergleichbaren Betrag. Die horizontale Wellenausbreitung führt im „full“-Experiment zu räumlichen Verschiebungen der Wellenfelder. Typische Distanzen zwischen den Positionen von Wellenpaketen bei 20 km und 80 km Höhe betragen 20° bis 30° sowohl in Länge als auch in Breite.

Die periodische Kraft durch Schwerewellen-Dämpfung wirkt zwar zurück auf die Gezeitenwellen, doch die dynamische Rückkopplung zwischen beiden wird in dieser Arbeit nicht vollzogen. Vielmehr werden äquivalente Rayleigh-Koeffizienten berechnet, um die Rückwirkung der Schwerewellen-Kräfte auf die Gezeiten abzuschätzen. Für diese komplexen Koeffizienten wird die periodische Kraft auf den Gezeitenwind und dessen Tendenz projiziert, da die effektive Wirkung von der Phasenlage zwischen Kraft und Wind abhängt. Der Realteil des äquivalenten Rayleigh-Koeffizienten beschreibt den Anteil der Schwerewellen-Kraft, der genau in Phase oder Gegenphase zum Wind liegt. Der Imaginärteil des äquivalenten Rayleigh-Koeffizienten jedoch resultiert aus der Kraftkomponente, die um 90 Grad zum Gezeitenwind phasenverschoben ist. Für das „TS“-Experiment, welches die Annahmen bei konventionellen Parametrisierungen nachstellt, werden positive Realteile und negative Imaginärteile der Rayleigh-Koeffizienten gefunden. Erstere weisen auf eine Abschwächung der Gezeitenamplituden hin und sind ein typisches, aber in der Literatur kontrovers diskutiertes Ergebnis der Sättigungsannahme für brechende Schwerewellen. Zweitere zeigen eine Verkleinerung der vertikalen Gezeiten-Wellenlänge an. Im Experiment „full“ verursachen sowohl Frequenz-Modulation als auch horizontale Refraktion eine Verringerung der äquivalenten Rayleigh-Koeffizienten. Der Imaginärteil der Rayleigh-Koeffizienten bleibt vorrangig negativ. Der Betrag seiner Minima im Bereich der Mesopause wird auf der Winterhemisphäre um den Faktor vier reduziert, während er sich um einen vergleichbaren Betrag auf der Sommerhemisphäre erhöht. Der Realteil der Rayleigh-Koeffizienten wird drastisch verringert. Die Maxima im Mesopausen-Bereich reduzieren sich um einen Faktor zwei bis vier. In Polnähe und in der unteren Thermosphäre ist so gut wie keine Dämpfung der Gezeiten durch Schwerewellen im „full“-Experiment zu erwarten. Da die Rayleigh-Koeffizienten nur eine erste Abschätzung für den Einfluss vom brechenden Schwerewellen auf Gezeiten bieten, wird in Zukunft eine interaktive Kopplung zwischen beiden Wellenphänomenen benötigt, um die Erkenntnisse über deren Wechselwirkung weiter zu vertiefen. Des weiteren muss über eine Schwerewellen-Anregung durch realistische Erzeugungsmechanismen in der Troposphäre nachgedacht werden. Die Schwerewellen-Gezeiten-Wechselwirkung bleibt also eine anspruchsvolle, aber spannende Aufgabe.

Contents

Zusammenfassung	V
List of Figures	XIII
List of Tables	XV
Notation	XVII
1. Introduction	1
1.1. On middle atmosphere dynamics	1
1.2. Interaction of gravity waves and thermal tides	2
1.3. On self-consistency in subgrid-scale parameterizations	4
1.4. Outline	4
2. A hierarchy of flow phenomena	7
2.1. From compressible to sound-proof dynamics	7
2.1.1. Basic compressible flow	7
2.1.2. On available energetics of the compressible dynamics	8
2.1.3. Discussion of sound-proof approximations	9
2.1.4. The anelastic approximation	10
2.2. On scale-selective filtering of subgrid-scale processes	12
2.2.1. Introduction	12
2.2.2. On turbulent structures	13
2.2.3. On filtering of gravity waves	15
2.3. Summary	22
3. Overview of diurnal tides	23
3.1. HAMMONIA model	23
3.2. Diurnal heating rates	24
3.3. Diurnal tides	26
3.4. Discussion	30
3.5. Summary	30
4. Dynamics of gravity waves	31
4.1. Discussion of gravity-wave sources	31
4.2. Propagation of linear gravity waves in a resting, isothermal atmosphere	32
4.3. Propagation of linear gravity waves in a general moving basic state . . .	36
4.3.1. Possible impacts of background wind and stability	36
4.3.2. Multiple-scale asymptotics	38
4.3.3. Wentzel-Kramers-Brillouin theory for gravity waves	46
4.3.4. Gravity-wave propagation and the ray-tracing method	50
4.4. Eliassen-Palm theorems	55
4.5. Summary	57

5. Modulation of gravity waves in thermal tides	59
5.1. Introduction	59
5.2. Global ray-tracing simulations with RAPAGI	59
5.2.1. Gravity-wave ensemble	59
5.2.2. Experimental setup	60
5.3. Gravity-wave frequency and phase-speed modulation	61
5.3.1. Motivation	61
5.3.2. The mechanism of frequency modulation	62
5.3.3. Simple energetics of frequency modulation	64
5.3.4. The impact on saturated gravity-wave trains	68
5.3.5. Vertical column thinking and phase velocity modulation in realistic flows	69
5.4. Horizontal propagation and refraction of gravity-wave fields	71
5.4.1. Motivation	71
5.4.2. Mechanisms of mean horizontal refraction	72
5.4.3. The impact on saturated gravity-wave trains	74
5.4.4. Horizontal refraction in realistic flows	74
5.5. Gravity-wave forces on the tide	79
5.5.1. Mean gravity-wave forces	79
5.5.2. Periodic forces due to wave stresses	81
5.5.3. Direct effects of horizontal refraction on the diurnal forcing	84
5.5.4. Equivalent Rayleigh friction coefficients	87
5.6. Summary	89
6. Conclusions and Outlook	91
6.1. Conclusions	91
6.2. Outlook	93
A. Appendix	95
A.1. Approximations of the Lagrangian-mean anelastic dynamics	95
A.2. Non-dimensionalization of linear gravity-wave dynamics	97
A.2.1. Linear continuity equation	97
A.2.2. Linear momentum balance	97
A.2.3. Linear thermodynamics	99
A.3. Relations between gravity-wave variables	100
A.3.1. Wave energy and fluxes	100
A.3.2. Wave action and pseudo-momentum	101
A.4. Conservation of wave action	102
A.5. Conservation of GW angular momentum in a resting, isothermal basic state	105
A.6. RAPAGI: the numerical implementation	105
Bibliography	107

List of Figures

1.1. Climatological zonal-mean flow	1
2.1. Actual and mean trajectory in GLM theory	18
3.1. Geographical distribution of heating rates at 5 km and 50 km	24
3.2. Amplitudes of diurnal heating rates	25
3.3. Geographical distribution of diurnal perturbations at 110 km	27
3.4. Vertical cross-section of diurnal zonal wind at different latitudes	28
3.5. Zonally-averaged tidal amplitudes	29
4.1. Structure of a simple gravity-wave packet	35
4.2. Variations of zonal wind and buoyancy frequency	37
5.1. Counter-propagating gravity waves in monochromatic tidal winds	63
5.2. GW packet in tidal background	64
5.3. Induced flow of a GW packet in tidal background	66
5.4. Distribution of travel times	69
5.5. Vertical profiles of u and c at subsequent times for the “noREF” experiment	70
5.6. Longitude cross-section of c_h and u_h for the “full” experiment	71
5.7. Schematic view of the mean refraction of the horizontal wavenumber vector	72
5.8. GW rays in stationary zonal-mean January wind	75
5.9. GW rays and planetary waves in the winter stratosphere	76
5.10. Mean GW streaming in a background flow including tides	77
5.11. Temporally averaged \mathbf{k}_h at 80 km	78
5.12. Distribution of meridional displacements	79
5.13. Average zonal GW force	81
5.14. Temporal behavior of the zonal GW force	82
5.15. Amplitudes of the zonal diurnal GW force	83
5.16. Amplitudes of the meridional diurnal GW force	84
5.17. Direct GW force contributions due to horizontal refraction and dissipation	86
5.18. Horizontal averages of GW forces due to horizontal refraction and dissipation	86
5.19. Real part of the equivalent Rayleigh friction coefficient	88
5.20. Imaginary part of the equivalent Rayleigh friction coefficient	89

List of Tables

4.1. Sign convention for gravity-wave wavenumbers	35
4.2. Typical background scales	39
4.3. Typical background variables and mean flow numbers	40
4.4. Typical gravity-wave scales for the medium-frequency regime	40
4.5. Distinguished limits for different gravity-wave regimes	44
5.1. Initial values for the gravity-wave ensemble	60
5.2. Overview of the different simulation experiments	61

Notation

General remarks

scalars	italic, e.g. f
vectors	boldface, e.g. \mathbf{f}
tensors	upright, sans-serif, e.g. \mathbf{F}
turbulent perturbations	double prime, e.g. \mathbf{v}''
Eulerian gravity-wave perturbations	single prime, e.g. \mathbf{v}'
Lagrangian gravity-wave disturbances	superscript ℓ , e.g. \mathbf{v}^ℓ
lifted variables	superscript ξ , e.g. \mathbf{v}^ξ
relevant scales	hat, upright, e.g. \hat{H}
dimensionless quantities	asterisk and uppercase for background variables, e.g. u^* and U^*
wave amplitudes	hat, e.g. \hat{u} (some exceptions $\hat{\omega}$, \hat{z} , \hat{c} , ...)
zonal and temporal average	bracket and overbar, e.g. $[\bar{u}]$
average over turbulent and gravity-wave scales	angle brackets with subscript t and w , e.g. $\langle u \rangle_t$ and $\langle u \rangle_w$ (from chapter 3, angle brackets are omitted for background quantities)
Lagrangian mean	brackets with superscript L , e.g. $\langle u \rangle_w^L$
reference state	subscript r , e.g. ρ_r

Partial list of symbols

$(\mathbf{e}_\lambda, \mathbf{e}_\varphi, \mathbf{e}_z)$	unit vectors in zonal, meridional and vertical direction
λ, φ	longitude and latitude
$(x, y, z), \hat{z}, \eta$	set of Cartesian coordinates, horizontally mean geopotential height, hybrid vertical coordinate
$\partial_t, \partial_\lambda, \partial_\varphi, \partial_z, \nabla, \nabla_h$	partial derivatives, full and horizontal gradient
D_t, d_t	advective derivatives with mean flow and with the group flow, i.e. along ray
c_v, c_p, R, g	specific heat capacity at constant volume and pressure, ideal gas constant and gravitational acceleration
f, β	Coriolis parameter, beta-term
$\mathbf{v}, (\mathbf{u}, w)$	full 3D, horizontal and vertical wind

p, ρ, T, π, θ	pressure, mass density, temperature, Exner pressure and potential temperature
b, ϕ	buoyancy and normalized pressure
$p_r, \rho_r, T_r, \pi_r, \theta_r$	hydrostatic reference values of pressure, mass density, temperature, Exner pressure and potential temperature
N, c_s, H_ρ	buoyancy frequency, speed of sound and density scale height derived from reference state
e, h, s, Φ	specific internal energy, enthalpy and entropy, and force potential
Π_{an}, Π_{el}	available potential and elastic energy per mass
h_{an}, s_{an}	anelastic enthalpy and entropy per mass
$\mathbf{T}, \mathbf{F}, \mathbf{R}_t, \mathbf{R}_w$	general stress tensor and tensors of molecular, turbulent and gravity-wave stresses
ν, ν_t, τ	molecular and turbulent viscosity, gravity-wave damping rate
$\mathbf{q}, \mathbf{J}_t, \mathbf{J}_w$	molecular, turbulent and gravity-wave fluxes
ϵ, E	dissipation, direct gravity-wave heating
ξ	Lagrangian flow displacement
Θ	gravity-wave phase
k, l, m, k_h	zonal, meridional and vertical wavenumber and horizontal wavenumber of gravity waves
\mathbf{k}, \mathbf{k}_h	full and horizontal wavenumber vector
$\omega, \hat{\omega}$	observed (ground-based) and intrinsic frequency
$\mathbf{c}_g, \hat{\mathbf{c}}_g, c_{gh}, c_{gz}$	vector of observed and intrinsic group velocity, horizontal and vertical group velocity
$\mathbf{c}, \hat{\mathbf{c}}, c, c_h, u_h$	vector of observed and intrinsic phase velocity, zonal, horizontal and vertical phase velocity, background wind in wave direction
M, C, Ω	vertical wavenumber and phase velocity of tide, diurnal frequency
$E_w, \mathcal{E}, A, \mathcal{P}$	wave energy, pseudo-energy, wave action and pseudo-momentum (all per volume)
F_{EP}, F, F_h	Eliassen-Palm flux, vertical flux of zonal and horizontal pseudo-momentum
$\mathbf{f}_{gw, h}^L, f_\lambda$	horizontal and zonal GW force
γ_R, γ_I	real and imaginary part of equivalent Rayleigh friction coefficient

1. Introduction

1.1. On middle atmosphere dynamics

The middle atmosphere is a fascinating part of the climate system. It is not directly in contact with surface of the earth but it strongly influences our life. For instance, in the middle atmosphere important chemical species such as ozone are present. It absorbs energy from then sun and protects us from dangerous UV radiation. Recently, the dynamical coupling between the middle atmosphere and the troposphere has become the focus of scientific interest (e.g the SPARC initiative ¹). Waves excited in the troposphere encounter various propagation conditions on their way up and can act back in different ways on the lower atmosphere. The spectrum of wave activity ranges from planetary structures such as solar thermal tides to mesoscale phenomena like gravity waves. The huge range of scales involved makes the dynamics of the middle atmosphere a challenging task.

The special vertical structure of the middle atmosphere is shown in fig. 1.1. Above

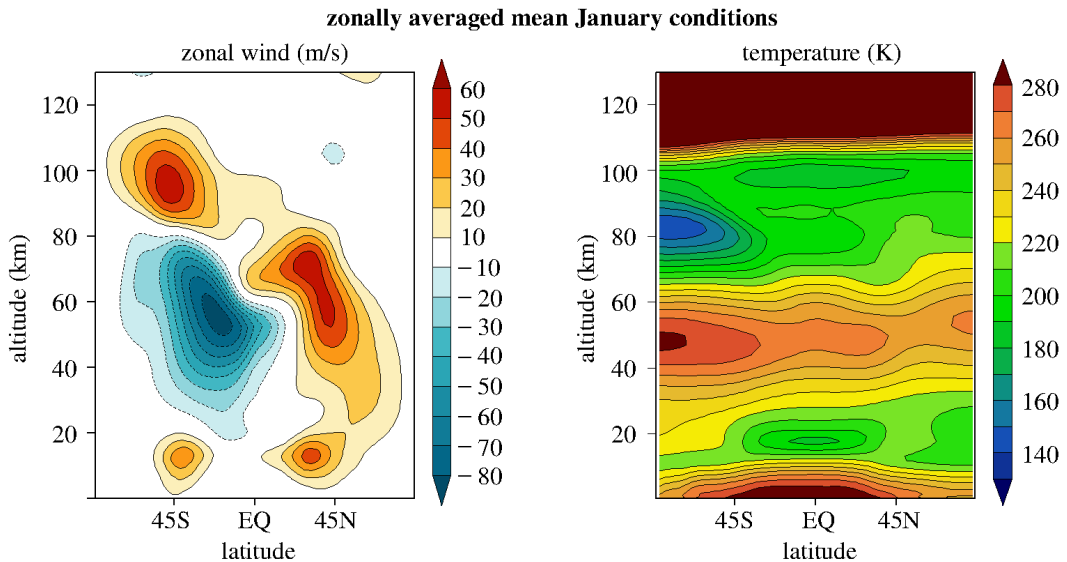


Figure 1.1.: The zonal-mean zonal wind $[\bar{u}]$ (left) and the zonal-mean temperature $[\bar{T}]$ (right) in the lower and middle atmosphere for mean January conditions obtained from HAMMONIA data (see section 3.1).

the cold tropopause, from about 10 km to 15 km, the mean temperature rises. In the summer hemisphere, ozone heating is responsible for that increase, but in the winter, dynamical effects are responsible. The mesosphere starts above the warm stratopause, located between about 50 km and 60 km. Here, the concentration of radiatively active air constituents strongly decreases, and the average temperature diminishes accordingly. Contrary to our intuition, however, the coldest region in the middle atmosphere is a place where the sun continuously shines, the polar summer mesopause. With average temperatures below 130 K, it is the coldest place in the whole atmosphere. Two wind

¹Information is available at <http://www.atmosp.physics.utoronto.ca/SPARC/index.html>

systems dominate the middle atmosphere. As shown in fig. 1.1, the west-wind vortex controls motion in the winter hemisphere and the east-wind jet the summer hemisphere with wind speeds up to 80 m/s. Above the mesopause region, the zonal wind reverses.

The mean residual circulation connecting the summer to the winter pole causes the extraordinary temperature structure of the middle atmosphere. On average, air rises above the tropical tropopause and is transported to the winter pole in the stratosphere. These large-scale motions, known as the Brewer-Dobson circulation, explain the anomalous ozone concentrations in high winter latitudes [Andrews *et al.*, 1987]. In the mesosphere, this circulation extends to the summer hemisphere and induces ascending motion at the summer pole and descending motion at the winter pole with an inter-hemispheric coupling between the two. Due to the enormous decrease in background pressure with altitude, rising air packets expand and descending air packets are compressed. Adiabatic cooling results in the first case and adiabatic warming in the second. A dynamically induced cold summer mesopause and warm winter stratopause develop as a result.

It is believed that mesoscale gravity waves excited by e.g. orography, deep convection and dynamical instability processes are the main cause of the residual circulation in the mesosphere [Fritts and Alexander, 2003]. Since the pioneering work of Lindzen [1981], Dunkerton [1982] and Holton [1982], the mechanisms of wave driving have been well established. The breaking of mesoscale gravity waves exerts a mean force in the mesopause region that differs between the two hemispheres. In the winter west-wind jet, preferentially westward propagating gravity waves reach the mesopause and deposit their negative zonal pseudo-momentum². A negative zonal force results in the winter mesosphere/lower thermosphere (MLT). The opposite happens in the summer hemisphere where there is a corresponding positive force. Climatologically, the mean gravity-wave forces are approximately balanced by a mean Coriolis torque which pushes air northward on both hemispheres in winter. Hence, the inter-hemispheric pole-to-pole circulation and the temperature structure far from radiative equilibrium are the consequences of the impact of mesoscale gravity waves on middle atmosphere dynamics.

Another important coupling mechanism between lower and middle atmospheres are solar thermal tides. They are large-scale atmospheric waves which are excited mainly by solar heating of ozone in the stratosphere and by heating of water vapor and latent heat release in the upper troposphere [Chapman and Lindzen, 1970]. Thermal tides were first investigated with respect to surface pressure, where their impact is small [Lindzen, 1990]. But as tidal waves propagate upward, the extreme decrease in density leads to increasing perturbations in wind and temperature as less mass is transported by tidal oscillations. In the MLT region, the amplitudes are considerably large and dominate the diurnal variability.

1.2. Interaction of gravity waves and thermal tides

As the Earth's atmosphere is continuously heated by the sun, tidal waves are always present in the middle atmosphere, even though their magnitude may undergo large variations [Lieberman *et al.*, 2007; Liu *et al.*, 2007]. Hence, the background conditions for gravity-wave propagation are periodically modulated by tidal perturbations. These impacts on gravity waves are far from being small. Therefore, the understanding of

² As discussed by McIntyre [1981], the physical momentum is not necessarily a property which is locally attached to the waves. For gravity waves, the term "pseudo-momentum" (see also Bühler [2009] and section 4.4) is more appropriate to describe a wave quantity which appears in the mean momentum budget.

the interaction of gravity waves and thermal tides is important for an understanding of many processes in the middle atmosphere.

For the interaction between gravity waves and thermal tides, one might discriminate between the impact of tides on gravity waves and the impact of gravity waves on tides. Clearly, the two are non-linearly connected, but this type of linear thinking can illuminate several important mechanisms of the interaction process. First, as mentioned, thermal tides change the background conditions for gravity-wave propagation. It is mainly the tidal winds that are responsible for major changes in gravity-wave properties. Thermal tides can destabilize gravity-wave packets and modify the gravity-wave breaking process. Second, due to the periodic modulation of gravity-wave fields, periodic gravity-wave forces and heating rates result especially when gravity waves deposit their pseudo-momentum and energy in the wave-breaking process. The enhanced turbulent diffusion due to the decay of globally-distributed gravity-wave packets can also strengthen the turbulent mixing of tidal properties. Finally, the induced tidal flows and temperature perturbations can act back on the gravity waves and change their impact on the tidal fields in a very non-linear way. Hence, a multitude of ways exists in which gravity wave and thermal tides might affect each other and this had initiated many investigations in the past.

The study of the interactions between gravity waves and thermal tides began with the work of *Fritts and Vincent* [1987], *Forbes et al.* [1991] and *Miyahara and Forbes* [1991], with simple qualitative interpretations of the impact of gravity waves based on Lindzen's saturation hypothesis [*Lindzen*, 1981] and with simplified linear modeling of tidal waves. Subsequently, refinements of all aspects of the models and discussions of the effect of different gravity-wave parameterizations were done within linear tidal models [*McLandress*, 1997; *Meyer*, 1999; *Ortland and Alexander*, 2006] and non-linear circulation models of the middle atmosphere including the diurnal cycle [*Mayr et al.*, 1999; *Norton and Thuburn*, 1999; *Mayr et al.*, 2001; *Akmaev*, 2001; *McLandress*, 2002]. A weakness of all the mentioned modeling studies is that they applied gravity-wave parameterizations with very restrictive assumptions.

Gravity-wave parameterizations typically work in vertical columns in which temporal and horizontal dependence of the background medium is neglected [*McLandress*, 1998]. Gravity-wave fields are assumed to adjust instantaneously to any change in the whole vertical column. As tidal waves induce large and fast variations of the flow, it is questionable whether the stationarity assumption is still valid. It is discussed in this study how time-dependence and horizontal gradients can have a significant impact on gravity-wave propagation and dissipation.

The more realistic behavior of small-scale gravity waves in a temporally changing large-scale wave was investigated by several authors [*Broutman*, 1984; *Broutman and Young*, 1986; *Zhong et al.*, 1995; *Eckermann and Marks*, 1996; *Sonmor and Klaassen*, 2000; *Sartelet*, 2003]. This problem is of interest in both the atmospheric and the oceanographic context. With the help of the ray-tracing method, motion of gravity-wave packets were directly described. Changes in gravity-wave properties were found which in some circumstances significantly differ from the estimate of the conventional parameterization. Most of the mentioned studies have in common that the ray tracing is applied in the presence of very simplified background waves. This motivated the use of the ray-tracing method for mesoscale gravity waves in a background of more realistic thermal tides obtained from climate model simulations. The comparison of diurnal gravity-wave forces retrieved from the ray tracing simulations to the conventional approach is one of our major concerns.

In contrast to the parameterization approach, high-resolution global simulations have

been performed very recently by *Watanabe and Miyahara* [2009]; *Becker* [2012]. Although dynamical processes which lead to the excitation of resolved gravity waves are well captured in these investigations, the intermittent excitation of gravity waves by deep moist convection is still not sufficiently well reproduced. Hence, one major gravity-wave source in the tropics is still not available to complex simulation experiments. Additionally, global models take only hydrostatic gravity waves into account. Furthermore, the inherent complexity of such models complicates the interpretation of the results and the formulation of explanations of the gravity wave-tidal interaction. Hence, models of intermediate complexity as presented here are needed for a deeper understanding of middle atmosphere dynamics.

1.3. On self-consistency in subgrid-scale parameterizations

Models are abstractions of real-world systems. In most situations, there is an inherent impossibility to describe all essential parts of the dynamics sufficiently due to incomplete knowledge about the whole system or other, e.g. computational, restrictions. To overcome this problem, unresolvable processes are approximately parameterized. In the best case, parameterizations are based on physical understanding of the corresponding process and satisfactorily constrained. In the worst case, parameterizations are just a set of tuning parameters adjusted to meet the modeler's needs.

In atmospheric modeling, parameterizations are mainly needed to describe small-scale processes which are beyond the spatial and temporal model resolution threshold. For instance, these include turbulent eddies and mesoscale gravity waves for middle-atmosphere climate models. In recent years, the self-consistent formulation of subgrid-scale parameterizations received large scientific attention. Several studies have been devoted to the angular momentum budget of turbulence parameterizations [*Becker*, 2001; *Burkhardt and Becker*, 2006; *Becker and Burkhardt*, 2007] and gravity-wave parameterizations [*Shepherd and Shaw*, 2004; *Shaw and Shepherd*, 2007; *Shaw and Shepherd*, 2009], turbulent, dissipative heating [*Boville and Bretherton*, 2003; *Becker*, 2003a; *Burkhardt and Becker*, 2006; *Becker and Burkhardt*, 2007] and direct gravity-wave heating [*Becker and Schmitz*, 2002; *Becker*, 2004; *Akmaev*, 2007]. It was emphasized that parameterizations of dynamical processes must not conflict with the underlying conservation law structure of the flow dynamics. Conservation of angular momentum and total energy and the 2nd law of thermodynamics are important restrictions which should be respected in the formulation of a subgrid-scale parameterization.

1.4. Outline

This thesis consists of two main parts: In the first, the theoretical concept of the interaction between thermal tides, gravity waves and induced turbulence is developed. A major theme of the current study is parameterization self-consistency. Hence, the terms in which the subgrid-scale gravity waves interact with the tides are derived here with special care. It is one of the major aims of this thesis to provide the theoretical framework for the interaction process. Assumptions in and restrictions to the current study are lucidly discussed. The second part is devoted to the description and interpretation of results obtained from ray-tracing simulations of gravity waves in thermal tides, where the comparison between the complex ray simulations and the conventional parameterization reveals important insights into the impact of thermal tides on gravity-wave propagation and dissipation.

The impact of small-scales on the large-scale flow is derived in chapter 2 using anelastic theory. Thermal tides are explained in chapter 3 with the help of a monthly mean daily cycle obtained from the climate model HAMMONIA. Chapter 4 is devoted to gravity waves, their structure, and their excitation mechanisms. Multiple-scale asymptotics and Wentzel-Kramer-Brillouin (WKB) theory is applied to the linear dynamics of gravity waves within a general moving basic state, and the ray-tracing method is derived for the propagation of gravity waves in thermal tides. In chapter 5, global ray-tracing simulations performed using the HAMMONIA data as background for wave propagation are discussed. The impacts of time-dependence and horizontal structure of the background conditions on the gravity-wave fields are studied. Three different ray-tracing experiments with differing assumptions are presented, and the diurnal forcing due to gravity waves in each is compared. The deficiencies of the conventional parameterization approach are evaluated. A summary and outlook is given in chapter 6 and supplementary material is collected together in the appendix.

Some of the theoretical considerations were inspired by a work on multiple-scale asymptotics by *Achatz et al.* [2010] to which the author contributed. A shortened version of the discussion of the ray-tracing results appeared in the Journal of Geophysical Research [*Senf and Achatz*, 2011]. Global ray-tracing simulations with a background of thermal tides and the calculation of the corresponding diurnal forces have not been done before to this extent. As most of the former studies of gravity wave-tide interaction were based on conventional gravity-wave parameterizations, the careful evaluation of the assumptions made in these parameterizations is of large scientific interest. The present study was able to contribute significantly to this issue and to address at least some of the problems appearing in vertical column parameterizations when temporal and horizontal dependence of the background flow is taken into account. The work of the author was financed as a project in the priority program CAWSES of the German Science Foundation (DFG). A corresponding overview about the project activities was given in *Achatz et al.* [2012]. Lastly, a derivation of generalized sound-proof dynamics and its underlying variational structure is in preparation [*Senf and Achatz*, 2012].

2. A hierarchy of flow phenomena

In this chapter, the theoretical basis of the interaction between gravity waves (GWs) and thermal tides is provided. The main goal is to derive average GW effects, and this is performed in detail to highlight the assumptions and approximation steps necessary for a consistent description. To begin with, the fundamental conservation laws of an ideal, dry, compressible fluid are reviewed, followed by a step-by-step reduction of the underlying dynamics. First, anelastic theory is presented and discussed. Second, a scale-selective filter is applied over turbulent scales, and then over gravity-wave scales. It is shown that the corresponding gravity-wave forces and heating rates depend on the type of averaging operator, i.e. the Eulerian or the Lagrangian mean. In other words, from a modeler's perspective the formulation of gravity-wave parameterizations depends on the interpretation of mean values.

2.1. From compressible to sound-proof dynamics

2.1.1. Basic compressible flow

The fundamental conservation laws form the starting point of the detailed derivation of average GW effects. An ideal, dry, compressible gas in a rotating frame of reference is considered. Balance equations for the mass, momentum and total energy of a fluid parcel can be derived and found e.g. in *Lange* [2002]:

$$\rho D_t \mathbf{v} + 2\boldsymbol{\Omega} \times \rho \mathbf{v} = \nabla \cdot \mathbf{T} - \rho \nabla \Phi, \quad (2.1)$$

$$D_t \rho + \rho \nabla \cdot \mathbf{v} = 0, \quad (2.2)$$

$$\rho D_t \left(\frac{\mathbf{v}^2}{2} + e + \Phi \right) = \nabla \cdot (\mathbf{T} \cdot \mathbf{v} - \mathbf{q}) + \rho Q, \quad (2.3)$$

where \mathbf{v} is the three-dimensional wind, and $D_t = \partial_t + \mathbf{v} \cdot \nabla$ is the advective derivative. The stress tensor $\mathbf{T} = \mathbf{F} - p\mathbf{1}$ includes contributions of normal pressure p and tension due to molecular friction \mathbf{F} . Φ , e , \mathbf{q} and Q are external force potential, internal energy, heat flux crossing the fluid parcel's boundary and rate of heating, respectively. From the total energy eq. (2.3), changes in the fluid entropy per unit mass s are determined by¹

$$\rho D_t s = \frac{1}{T} (\mathbf{F} \cdot \nabla \mathbf{v} + \rho Q - \nabla \cdot \mathbf{q}). \quad (2.4)$$

The fluid entropy can change by internal heating, i.e. by dissipation of mechanical energy $\epsilon = \mathbf{F} \cdot \nabla \mathbf{v} / \rho$, and by an external input of heat $Q - \nabla \cdot \mathbf{q} / \rho$. The second law of thermodynamics states that the entropy change is greater than the external input of heat for an irreversible process. This constrains the dissipation rate ϵ to be positive definite [*Lange*, 2002].

¹ The notation is adapted from *Lange* [2002]: It is not explicitly distinguished between row and column vectors. The dyadic product between \mathbf{a} and \mathbf{b} is denoted as \mathbf{ab} . The scalar product between vectors is $\mathbf{a} \cdot \mathbf{b}$ and between a dyad and a vector $\mathbf{ab} \cdot \mathbf{c} = \mathbf{a}(\mathbf{b} \cdot \mathbf{c})$. In this way, the double scalar product is defined, where the first scalar product is applied to the inner and the second to the outer vectors of the dyads, i.e. $\mathbf{ab} \cdot \mathbf{cd} = (\mathbf{b} \cdot \mathbf{c})(\mathbf{a} \cdot \mathbf{d})$.

2.1.2. On available energetics of the compressible dynamics

The set of compressible equations (2.1)-(2.3) supports several kinds of waves. But only a small portion of the total potential energy $e + \Phi$ is converted into kinetic energy of these waves and is called “*available energy*” [Lorenz, 1955]. In a resting basic state, acoustic and gravity waves contribute to the available energy of a compressible fluid. In the following, the partitioning of energy between both types of waves is discussed for an inviscid, adiabatic flow. This is the preparation for the construction of the anelastic theory in which acoustic waves are suppressed.

The natural thermodynamic state variables of internal energy e are entropy s and specific volume ρ^{-1} . For an air parcel in isentropic motion, the specific volume changes a lot. However, the pressure p of the air parcel adjusts quickly fast to the surrounding background pressure. This behavior favors the use of specific enthalpy $h(p, s)$ in a discussion of available energetics [Andrews, 1981; Bannon, 2004, 2005]. With $h = e + p/\rho$, the conservation of total energy (2.3) for inviscid, adiabatic motion is

$$\rho D_t \left(\frac{\mathbf{v}^2}{2} + h - p/\rho + \Phi \right) + \nabla \cdot (p\mathbf{v}) = 0. \quad (2.5)$$

For an ideal, dry gas, the specific enthalpy is

$$h(p, s) = c_p T(p, s) = c_p \left(\frac{p}{p_{00}} \right)^{R/c_p} e^{s/c_p}, \quad (2.6)$$

where the standard pressure p_{00} , the heat capacity at constant pressure c_p and the ideal gas constant R appear. In the following, the explicit dependence of h on p and s is not used. Rather the structure of the specific enthalpy suggests the introduction of the Exner pressure

$$\pi(p) = \left(\frac{p}{p_{00}} \right)^{R/c_p} \quad (2.7)$$

and the potential temperature

$$\theta(s) = e^{s/c_p}, \quad (2.8)$$

in which the enthalpy is bi-linear

$$h = c_p \pi \theta. \quad (2.9)$$

Next, using a hydrostatic reference state

$$\nabla p_r = -\rho_r \nabla \Phi, \quad (2.10)$$

the energetics of perturbations, e.g. $p' = p - p_r$, is investigated. As discussed by Lighthill [1978], the rate of work due to the reference pressure p_r

$$\rho^{-1} \nabla \cdot (p_r \mathbf{v}) = \rho^{-1} p_r \nabla \cdot \mathbf{v} + \rho^{-1} \mathbf{v} \cdot \nabla p_r = p_r D_t \rho^{-1} + \rho^{-1} D_t p_r = D_t (p_r / \rho) \quad (2.11)$$

should be excluded from the energy flux $p\mathbf{v}$. Adopting notation of Andrews [1981],

$$\rho D_t \left(\frac{\mathbf{v}^2}{2} + \Pi_{an} + \Pi_{el} \right) + \nabla \cdot (p' \mathbf{v}) = 0 \quad (2.12)$$

is obtained with the anelastic part of available potential energy

$$\Pi_{an} = c_p \pi_r \theta + \Phi = h_{an} + \Phi \quad (2.13)$$

and the elastic (or acoustic) contribution

$$\Pi_{el} = c_p \pi' \theta - p' / \rho. \quad (2.14)$$

In eq. (2.13), the anelastic enthalpy $h_{an} = h(p_r, s)$ is defined. As shown in section 2.1.4, this quantity is important in the construction of anelastic thermodynamics. After linearization, it was demonstrated by *Andrews* [1981], that the available potential energy is a function only of s' , or equivalently of θ' , whereas the acoustic energy only depends on p' . The expressions

$$\Pi_{an} \approx \frac{1}{2N^2} \left(\frac{g \theta'}{\theta_r} \right)^2 \quad \text{and} \quad \Pi_{el} \approx \frac{1}{2c_s^2} \left(\frac{p'}{\rho_r} \right)^2 \quad (2.15)$$

are familiar results from linear wave theory [*Gossard and Hooke*, 1975; *Lighthill*, 1978], where the square of buoyancy frequency $N^2 = g \partial_z \ln \theta_r$, the square of the speed of sound $c_s^2 = \gamma R T_r$ and the ratio of heat capacities $\gamma = c_p / c_v$ were utilized. Hence, in an anelastic approximation the acoustic term Π_{el} should disappear.

2.1.3. Discussion of sound-proof approximations

It is mostly assumed that acoustic waves have a negligible impact on the evolution of flow phenomena connected to weather and climate, and as acoustic waves strongly influence the numerical implementation of the flow equations [*Durran*, 1999], it is of some advantage to filter sound waves from the compressible dynamics in an analytic fashion. For non-hydrostatic motion, the anelastic and the pseudo-incompressible equations are valuable examples of systems of sound-proof equations.

The anelastic system was introduced by *Ogura and Phillips* [1962] for an isentropic reference state in a study of atmospheric convection. In their work, the set of compressible equations was expanded around an isentropic reference atmosphere with uniform potential temperature. Restricting their analysis to low Mach number flows [*Durran*, 2008], they could show that the resulting dynamical equations suppress acoustic modes. *Wilhelmson and Ogura* [1972] extended the work of *Ogura and Phillips* [1962] by allowing for a non-isentropic basic state. As some problems with the conservation of energy of the resulting system were discovered, *Lipps and Hemler* [1982] re-derived the anelastic equations with modifications to fulfill the energetic constraints. As discussed by *Lipps* [1990] (and e.g. *Achatz et al.* [2010]) the anelastic equations are only applicable to weakly stratified basic states in which the vertical variations of basic state entropy are small.

Durran [1989] derived the pseudo-incompressible equations for low Mach number flows. Also for the derivation of the traditional pseudo-incompressible dynamics, thermodynamic reference profiles of a resting basic state are used. In the mass balance, *Durran* [1989] introduced a pseudo-mass, or equivalently a pseudo-density, which is conservatively transported by the flow. This differs from the anelastic dynamics in which the reference mass is assumed to be conserved. Note that in the pseudo-incompressible equations, only variations of the true mass of an air parcel due to pressure perturbations are neglected, whereas in the anelastic equations, also variations due to entropy perturbations are neglected [*Senf and Achatz*, 2012]. Because of this difference, no change

in the pseudo-incompressible momentum balance is needed to maintain total energy conservation for the case of a non-isentropic basic state.

Using a multi-scale technique, *Achatz et al.* [2010] investigated the asymptotic behavior of compressible, pseudo-incompressible, and anelastic flows under conditions representative of GW breaking. They found that the pseudo-incompressible and compressible versions are consistent, but some differences appear for the anelastic system in that limit. *Davies et al.* [2003] performed normal mode analysis to evaluate the accuracy of the sound-proof systems compared to the compressible one. They did not recommend the use of the anelastic system and also partially discouraged the use of the pseudo-incompressible system for the purposes of numerical weather prediction and climate simulations. *Nance and Durran* [1997] reported in a more positive way about the performance of the sound-proof equations. In a study of mountain waves, both sound-proof systems achieved acceptable results with the pseudo-incompressible system performing slightly better. On the other hand, it was discussed by *Durran* [2008] that the performance of the sound-proof systems becomes unsatisfactory if the fluid motion is significantly influenced by Earth's rotation. In fact, this is true for large-scale planetary wave dynamics. *Durran* [2008] recommended the use of a compressible, hydrostatic model for that purpose. The use of an anelastic approximation may thus lead to a less-than-optimal large-scale equation system. However, the averaging of the resulting equations, especially the Lagrangian-mean average, is much simpler than for the compressible dynamics [*Becker and Schmitz*, 2002; *Becker*, 2003b; *Becker*, 2004]. Therefore, despite its shortcomings, the anelastic system is used in the present study.

2.1.4. The anelastic approximation

The anelastic equations, discussed above, will form the basis for the description of GW motion. Its derivation is sketched here. First, the mass balance, eq. (2.2), is altered. As mentioned, the effect of the perturbation density on the actual mass of a fluid parcel is neglected. The reference mass is transported conservatively, so that

$$\nabla \cdot (\rho_r \mathbf{v}) = 0. \quad (2.16)$$

Next, the momentum balance is modified in a way described by *Durran and Arakawa* [2007]. The hydrostatic reference state, eq. (2.10), is included in eq. (2.1) and the stress tensor is expanded:

$$D_t \mathbf{v} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{\nabla p}{\rho} + \frac{\nabla p_r}{\rho_r} + \frac{1}{\rho} \nabla \cdot \mathbf{F}.$$

The first approximation which enters the momentum balance concerns the magnitude of the perturbations of thermodynamic variables. Typically, these are orders of magnitude smaller than the corresponding reference values in geophysical flow situations. For instance, the global mean surface pressure is around 1000 hPa, whereas the difference between low and high pressure events is not larger than 100 hPa. Therefore, linearization of the momentum equation with respect to ρ and p and replacing $\rho \rightarrow \rho_r$ in the viscous force leads to

$$D_t \mathbf{v} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\frac{\nabla p'}{\rho_r} - \frac{g\rho'}{\rho_r} \mathbf{e}_z + \frac{1}{\rho_r} \nabla \cdot \mathbf{F}, \quad (2.17)$$

where a vertical gravitational force $-\nabla\Phi = -g\mathbf{e}_z$ has been assumed. Again, ρ' and p' denote deviations from the reference state ρ_r and p_r , respectively. With the help of the

linearized equation of state of an ideal gas

$$\frac{\rho'}{\rho_r} = \frac{p'}{c_s^2 \rho_r} - \frac{\theta'}{\theta_r}, \quad (2.18)$$

terms involving p' and ρ' in eq. (2.17) are reformulated

$$-\frac{\nabla p'}{\rho_r} - \frac{g\rho'}{\rho_r} \mathbf{e}_z = -\nabla \left(\frac{p'}{\rho_r} \right) + \left(\frac{p'}{\rho_r} \frac{\partial_z \theta_r}{\theta_r} + \frac{g\theta'}{\theta_r} \right) \mathbf{e}_z.$$

It was argued by *Lippis* [1990] that the term which couples the pressure perturbation with the static stability is negligible at this order of the anelastic expansion. Clearly, this is only valid for weakly stratified flows [*Achatz et al.*, 2010]. Hence, the anelastic momentum equation becomes

$$D_t \mathbf{v} + 2\boldsymbol{\Omega} \times \mathbf{v} = -\nabla \phi + b \mathbf{e}_z + \frac{1}{\rho_r} \nabla \cdot \mathbf{F}, \quad (2.19)$$

where the normalized pressure $\phi = p'/\rho_r$ and the buoyancy $b = g\theta'/\theta_r$ have been introduced.

For the anelastic thermodynamics, the entropy evolution eq. (2.4) is modified.² The anelastic entropy is defined via [*Bannon*, 1996]

$$D_t s_{an} = \frac{c_p}{\theta_r} D_t \theta = \frac{Q_{an}}{T_r}, \quad (2.20)$$

where

$$Q_{an} = \frac{1}{\rho_r} \mathbf{F} \cdot \nabla \mathbf{v} + Q - \frac{1}{\rho_r} \nabla \cdot \mathbf{q} \quad (2.21)$$

is the anelastic rate of heating. Several equivalent forms of the thermodynamics can be stated. From the partitioning of the available energy in section 2.1.2 only the anelastic part (2.13) is retained to define the anelastic enthalpy [*Durran and Arakawa*, 2007]

$$h_{an} = c_p \pi_r \theta. \quad (2.22)$$

Using the evolution of anelastic entropy (2.20), the hydrostatic assumption for the reference values, i.e. $c_p \theta_r \partial_z \pi_r = -g$ and $T_r = \pi_r \theta_r$, an equation for the anelastic enthalpy results:

$$D_t h_{an} = -(g + b)w + Q_{an}. \quad (2.23)$$

Finally, the evolution equation of the buoyancy is

$$D_t b + wN^2 \left(1 + \frac{b}{g} \right) = \frac{gQ_{an}}{c_p T_r}. \quad (2.24)$$

For the discussion of the anelastic energetics, the enthalpy form of the thermodynamic equations is the most useful. The kinetic energy evolution equation is derived by $\mathbf{v} \cdot (2.19)$, i.e.

$$\rho_r D_t \left(\frac{\mathbf{v}^2}{2} \right) = -\nabla \cdot (\rho_r \mathbf{v} \phi - \mathbf{v} \cdot \mathbf{F}) + \rho_r w b - \nabla \mathbf{v} \cdot \mathbf{F}. \quad (2.25)$$

² Very recently, an elegant discussion of the thermodynamics of the pseudo-incompressible approximation appeared in *Klein and Pauluis* [2012].

Energy is exchanged via stresses at the boundaries of a fluid parcel and with anelastic enthalpy due to buoyant production and mechanical dissipation. Meanwhile, h_{an} is influenced by the external force potential. The total anelastic energy obeys

$$\rho_r D_t \left(\frac{\mathbf{v}^2}{2} + \Pi_{an} \right) = -\nabla \cdot (\rho_r \mathbf{v} \phi - \mathbf{v} \cdot \mathbf{F} - \mathbf{q}) + \rho_r Q, \quad (2.26)$$

where $\Pi_{an} = h_{an} + \Phi$. Obviously, the acoustic energy Π_{el} , involved in the available energetics of a compressible fluid in (2.12), is absent here.

In the following, the set of anelastic equations is used to derive the effective forces and heating rates due to subgrid-scale processes. Later in section 4.2 and 4.3.2, the anelastic equations are linearized and GW solutions within a tidal background flow are investigated.

2.2. On scale-selective filtering of subgrid-scale processes

2.2.1. Introduction

The careful reduction of the dynamics of a complex system is one of the most ambitious tasks in scientific research, but it is the necessary basis for understanding and modeling of many phenomena. In fluid dynamics, it has proven useful to distinguish different flow phenomena by their temporal and spatial scales. Unfortunately, the grouping of different flow structures is not as unique as one might wish. Usually no clear spectral gaps separate regions with different temporal and spatial properties. Nevertheless, one of the most common approaches is to select just a few flow features and assume that they exist in superposition in the flow under consideration. For the anelastic system, the flow may be divided into turbulence, gravity waves and large-scale motion, whereas acoustic waves were removed in advance. Filter operators are defined which average over the corresponding subgrid-scales. For instance a box average over several kilometers for turbulence and several hundred kilometers for gravity waves might be one choice. Another possibility is the ensemble filter in which one would average over an ensemble of stochastically-distributed realizations of the subgrid-scale process, e.g. an ensemble of gravity waves with randomly shifted initial phases. Regardless of the definition of the averaging operator, it has to fulfill certain Reynolds assumptions about the filtering [see e.g. *Lange, 2002*]. Without loss of generality, the ensemble average is considered in the following. It is assumed that all field variables continuously depend on the ensemble parameter α and are denoted by \mathbf{v}_α , etc. Two different formulations of the averaging operators are introduced:

The Eulerian average is a local average. The mean field is estimated via a weighted integration over all realizations of the ensemble at a fixed position \mathbf{x} on the large-scale grid and a fixed time t . With weight W_α , the Eulerian average of velocity \mathbf{v} is

$$\langle \mathbf{v} \rangle(\mathbf{x}, t) = \int d\alpha W_\alpha \mathbf{v}_\alpha(\mathbf{x}, t). \quad (2.27)$$

The Lagrangian average is a non-local average. Here, contrary to the Eulerian mean, a field variable is evaluated at a displaced position $\mathbf{x} + \boldsymbol{\xi}_\alpha$ and the integration is done over an ensemble of different fluid trajectories. The displacement $\boldsymbol{\xi}_\alpha$ is induced by subgrid-scale motion. Hence, the Lagrangian mean is

$$\langle \mathbf{v} \rangle^L(\mathbf{x}, t) = \langle \mathbf{v}^\xi \rangle(\mathbf{x}, t) = \int d\alpha W_\alpha \mathbf{v}_\alpha(\mathbf{x} + \boldsymbol{\xi}_\alpha(\mathbf{x}, t), t). \quad (2.28)$$

For small displacements, calculating the Lagrangian average only requires information about the (unresolved) local neighborhood of \mathbf{x} . The position \mathbf{x} on the large-scale grid is interpreted as the location of the ensemble-mean fluid trajectory. Therefore, the Lagrangian average $\langle \mathbf{v} \rangle^L$ is an average over the set of all individual fluid trajectories.

It is not quite obvious which type of average is more appropriate for description of the dynamics of the resolved flow. It clearly depends on the desired interpretation of the mean fields and perhaps on the straightforwardness of the formulation of the corresponding subgrid-scale parameterization.

2.2.2. On turbulent structures

Applying a scale-selective filter over turbulent structures yields an averaged flow which includes all the effects of gravity waves and the large-scale dynamics and furthermore some average turbulence impact. In the following, the Eulerian average is chosen for the filtering of subgrid-scale turbulence. Note, however that there has been intensive research in the field of Lagrangian averaging of turbulent fluctuations [see *Holm*, 1999, 2002, and references therein], in which the so-called Camassa-Holm equations or Navier-Stokes alpha models arise. Even though the application of the Lagrangian average to filter turbulence is very interesting, it is clearly outside the scope of the present work.

In the following, each flow quantity is divided into an Eulerian mean and a turbulent part, e.g. $\langle \mathbf{v} \rangle_t$ and \mathbf{v}'' , where the subscript t denotes the average over random turbulent fluctuations. The averaging operator is applied to the anelastic continuity equation (2.16),

$$\nabla \cdot (\rho_r \langle \mathbf{v} \rangle_t) = 0, \quad (2.29)$$

to the momentum eq. (2.19)

$$\left(\partial_t + \langle \mathbf{v} \rangle_t \cdot \nabla \right) \langle \mathbf{v} \rangle_t + 2\mathbf{\Omega} \times \langle \mathbf{v} \rangle_t = -\nabla \langle \phi \rangle_t + \langle b \rangle_t \mathbf{e}_z + \frac{1}{\rho_r} \nabla \cdot \left(\langle \mathbf{F} \rangle_t + \mathbf{R}_t \right), \quad (2.30)$$

and to the enthalpy eq. (2.23)

$$\begin{aligned} \rho_r \left(\partial_t + \langle \mathbf{v} \rangle_t \cdot \nabla \right) \langle h_{an} \rangle_t &= -\rho_r g \langle w \rangle_t - \rho_r \langle b \rangle_t \langle w \rangle_t - \rho_r \langle b'' w'' \rangle_t \\ &\quad - \nabla \cdot \left(\langle \mathbf{q} \rangle_t + \mathbf{J}_t \right) + \langle \mathbf{F} \rangle_t \cdot \nabla \langle \mathbf{v} \rangle_t + \langle \mathbf{F}'' \cdot \nabla \mathbf{v}'' \rangle_t + \rho_r \langle Q \rangle_t. \end{aligned} \quad (2.31)$$

The Reynolds stress tensor

$$\mathbf{R}_t = -\rho_r \langle \mathbf{v}'' \mathbf{v}'' \rangle_t \quad (2.32)$$

and the turbulent sensible heat-flux

$$\mathbf{J}_t = \rho_r \langle \mathbf{v}'' h''_{an} \rangle_t \quad (2.33)$$

arise from the non-linear advection terms and give the average transport of turbulent momentum and sensible heat by turbulent fluctuations. The vertical turbulent buoyancy flux $\rho_r \langle b'' w'' \rangle_t$ and the average frictional heating due to dissipation of turbulent kinetic energy $\langle \mathbf{F}'' \cdot \nabla \mathbf{v}'' \rangle_t$ influence the dynamics of the mean anelastic enthalpy. The average

flow equations are completed by an evolution equation for the mean turbulent kinetic energy $E_{t,kin} = \rho_r \langle |\mathbf{v}''|^2/2 \rangle_t$,

$$\begin{aligned} \partial_t E_{t,kin} + \nabla \cdot \left(\langle \mathbf{v} \rangle_t E_{t,kin} + \frac{\rho_r}{2} \langle \mathbf{v}'' |\mathbf{v}''|^2 \rangle_t + \rho_r \langle \phi'' \mathbf{v}'' \rangle_t - \langle \mathbf{F}'' \cdot \mathbf{v}'' \rangle_t \right) = \\ \rho_r \langle b'' w'' \rangle_t - \langle \nabla \mathbf{v}'' \cdot \mathbf{F}'' \rangle_t + \mathbf{R}_t \cdot \nabla \langle \mathbf{v} \rangle_t \equiv \dot{E}_{t,kin}, \end{aligned} \quad (2.34)$$

where the shorthand $\dot{E}_{t,kin}$ is defined following *Becker* [2003b]. The vertical turbulent buoyancy flux $\rho_r \langle b'' w'' \rangle_t$ and the average frictional heating due to dissipation of turbulent kinetic energy $\langle \mathbf{F}'' \cdot \nabla \mathbf{v}'' \rangle_t$ are eliminated from the average anelastic enthalpy eq. (2.31) to yield

$$\begin{aligned} \rho_r \left(\partial_t + \langle \mathbf{v} \rangle_t \cdot \nabla \right) \langle h_{an} \rangle_t = -\rho_r g \langle w \rangle_t - \rho_r \langle b \rangle_t \langle w \rangle_t - \nabla \cdot \left(\langle \mathbf{q} \rangle_t + \mathbf{J}_t \right) \\ + \left(\langle \mathbf{F} \rangle_t + \mathbf{R}_t \right) \cdot \nabla \langle \mathbf{v} \rangle_t + \rho_r \langle Q \rangle_t - \dot{E}_{t,kin}. \end{aligned} \quad (2.35)$$

When the vertical turbulent buoyancy flux $\rho_r \langle b'' w'' \rangle_t$, shear production $\mathbf{R}_t \cdot \nabla \langle \mathbf{v} \rangle_t$ and dissipation of turbulent kinetic energy $\langle \nabla \mathbf{v}'' \cdot \mathbf{F}'' \rangle_t$ are in balance then $\dot{E}_{t,kin} = 0$. This is assumed in the following and called “*quasi-stationary turbulence*”. However, when a time-dependent turbulence model is used in a study of the interaction between gravity waves and thermal tides this point *must* be revised. The major change from the unfiltered to the filtered anelastic dynamics is the inclusion of the Reynolds stresses and the subgrid-scale enthalpy flux which are added to their molecular counterparts. For convenience, the Eulerian-mean anelastic equations of quasi-stationary turbulence are summarized here:

$$\nabla \cdot \left(\rho_r \langle \mathbf{v} \rangle_t \right) = 0, \quad (2.36)$$

$$\left(\partial_t + \langle \mathbf{v} \rangle_t \cdot \nabla \right) \langle \mathbf{v} \rangle_t + 2\boldsymbol{\Omega} \times \langle \mathbf{v} \rangle_t = -\nabla \langle \phi \rangle_t + \langle b \rangle_t \mathbf{e}_z + \frac{1}{\rho_r} \nabla \cdot \left(\langle \mathbf{F} \rangle_t + \mathbf{R}_t \right), \quad (2.37)$$

$$\begin{aligned} \rho_r \left(\partial_t + \langle \mathbf{v} \rangle_t \cdot \nabla \right) \langle h_{an} \rangle_t = -\rho_r \left(g + \langle b \rangle_t \right) \langle w \rangle_t - \nabla \cdot \left(\langle \mathbf{q} \rangle_t + \mathbf{J}_t \right) \\ + \left(\langle \mathbf{F} \rangle_t + \mathbf{R}_t \right) \cdot \nabla \langle \mathbf{v} \rangle_t + \rho_r \langle Q \rangle_t. \end{aligned} \quad (2.38)$$

Remarks on the turbulent available potential energy

The total kinetic energy of the flow is a non-linear, or more precisely a quadratic, function of the flow velocity \mathbf{v} . This bi-linearity leads to the appearance of two kinetic energy terms in the averaged dynamics: the kinetic energy of the average flow $\rho_r \langle \mathbf{v} \rangle_t^2/2$ and the averaged turbulent kinetic energy $E_{t,kin}$. As discussed above the sum of the anelastic enthalpy and the force potential, i.e. $\Pi_{an} = h_{an} + \Phi$, plays the role of available potential energy and is a function of the entropy perturbation $s - s_r$, or equivalently the buoyancy b , for a given reference state s_r [see e.g. *Andrews*, 1981]. Averaging of $\Pi_{an}(b)$ leads to

$$\langle \Pi_{an}(b) \rangle_t = \Pi_{an}(\langle b \rangle_t) + \left(\langle \Pi_{an}(b) \rangle_t - \Pi_{an}(\langle b \rangle_t) \right) \quad (2.39)$$

where the first term is the available potential energy of the averaged flow. The term in parentheses is the averaged turbulent available potential energy. Note that, the explicit

dependence of Π_{an} on b is only known for special cases such as constant temperature reference states and linear perturbations [Andrews, 1981]. From eq. (2.15)

$$\langle \Pi_{an}(b) \rangle_t \approx \frac{\langle b \rangle_t^2}{2N^2} + \frac{\langle b'^2 \rangle_t}{2N^2} \quad (2.40)$$

is obtained. A similar partitioning was utilized by Plumb [1983] in his discussion on the atmospheric energy cycle. An exact partitioning can also be done in the Boussinesq case [Achatz, 2007].

2.2.3. On filtering of gravity waves

Motivation

In the following, it is assumed that the dynamics of gravity waves is not sufficiently resolved within a general circulation model (GCM). The model variables describe the average large-scale flow where GW motion is filtered out, but the choice of the average is to some extent arbitrary and strongly depends on the interpretation of the resolved variables. In section 2.2.1, two different filter operators acting on a random ensemble of gravity waves were introduced: the Eulerian average and the Lagrangian average.

If the set of anelastic equations (2.36)-(2.38) is locally averaged in an Eulerian sense, then mean equations result which are equivalent to the set of turbulence-averaged equations. GW forces result from the divergence of the GW momentum flux tensor and GW direct heating rates depend on the dissipation of GW kinetic energy, the vertical GW buoyancy flux and the convergence of GW sensible heat fluxes [Becker and Schmitz, 2002; Becker, 2003a, b; Becker, 2004]. The classical GCM can be formulated in abstract notation as

$$\mathcal{M}[\mathbf{Y}] = \mathbf{Z}, \quad (2.41)$$

where \mathcal{M} is the non-linear model operator acting on the resolved, Eulerian-averaged variables summarized in the vector \mathbf{Y} , and \mathbf{Z} is the parameterization of the impact of subgrid-scale processes which is needed when the underlying set of equations is averaged in the classical Eulerian sense.

However, it has been argued, e.g. by Andrews *et al.* [1987], that in contrast to the Eulerian circulation, the residual circulation is more appropriate for describing the transport of air. The two differ especially, when the circulation is driven by waves affected by the Earth's rotation. The discussion comes up most frequently in the context of globally-averaged circulations driven by Rossby waves, but easily translates to locally averaged flows driven by inertia-gravity waves. As the residual description is equivalent to the Lagrangian-mean one, it might be favorable to establish a GCM based on a Lagrangian-average. Therefore, if the underlying dynamics is averaged in a Lagrangian sense the variable vector of the Lagrangian-mean flow \mathbf{Y}^L is different from its Eulerian counterpart \mathbf{Y} . The difference, which might be called the Stokes correction $\mathbf{Y}^S = \mathbf{Y}^L - \mathbf{Y}$, arises due to different formulations of the subgrid-scale parameterization. Starting from the Eulerian-mean model (2.41), a Lagrangian-mean model can be constructed with the form

$$\mathcal{M}[\mathbf{Y}^L] = \mathbf{Z}^L, \quad (2.42)$$

where the same kernel \mathcal{M} is used to integrate the Lagrangian-averaged variables \mathbf{Y}^L , but a different set of subgrid-scale parameterizations

$$\mathbf{Z}^L = \mathbf{Z} + \mathcal{M}[\mathbf{Y}^L] - \mathcal{M}[\mathbf{Y}] \quad (2.43)$$

is used. If the difference between \mathbf{Y} and \mathbf{Y}^L in the initial state is ignored (as this might be hard to specify), the main difference between the Eulerian-mean and Lagrangian-mean flow arises due to the specification of the subgrid processes \mathbf{Z} and \mathbf{Z}^L . That is, for a given GCM, the type of the parameterizations determines whether the variables should be interpreted as Eulerian-mean or Lagrangian-mean. In any case, either can be converted into the other by adding or subtracting the Stokes correction \mathbf{Y}^S at any time step.

The equation (2.43) gives a hint as to how the subgrid-scale parameterizations differ: For small-amplitude waves, the Stokes corrections \mathbf{Y}^S are also small compared to \mathbf{Y}^L . Hence, $\mathcal{M}[\mathbf{Y}]$ can be expanded around \mathbf{Y}^L and the subgrid-scale parameterizations in the Lagrangian-mean model are

$$\mathbf{Z}^L \approx \mathbf{Z} + \mathcal{L}[\mathbf{Y}^S; \mathbf{Y}^L], \quad (2.44)$$

where $\mathcal{L}[\mathbf{Y}^S; \mathbf{Y}^L]$ is the corresponding linear, tangent model which advances \mathbf{Y}^S given the basic state \mathbf{Y}^L . Even if this connection between the Eulerian-mean and Lagrangian-mean model is quite straightforward, a different (hopefully more didactic) route is chosen in deriving the parameterizations of the Lagrangian-mean model in the proceeding paragraphs. In the following, first the conventional Eulerian model is formulated, and then the Lagrangian-mean model is derived by investigating the displaced dynamics as done in *Andrews and McIntyre [1978]*.

Eulerian average

An Eulerian average is applied to the set (2.36)-(2.38). The former mean-flow variables, e.g. $\langle \mathbf{v} \rangle_t$, are decomposed into a new Eulerian mean $\langle \mathbf{v} \rangle_w$ describing the large-scale part of the flow, and a GW perturbation \mathbf{v}' . An ensemble of randomly shifted gravity waves is assumed and the Eulerian average is performed with respect to the set of random realizations, but a spatial filter with a horizontal extent of a few hundred kilometers can also be appropriate for GW filtering. Hence, the large-scale continuity equation is

$$\nabla \cdot (\rho_r \langle \mathbf{v} \rangle_w) = 0. \quad (2.45)$$

In the average momentum equation

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w \cdot \nabla \right) \langle \mathbf{v} \rangle_w + 2\boldsymbol{\Omega} \times \langle \mathbf{v} \rangle_w = & -\nabla \langle \phi \rangle_w + \langle b \rangle_w \mathbf{e}_z \\ & + \frac{1}{\rho_r} \nabla \cdot \left(\langle \mathbf{F} \rangle_w + \langle \mathbf{R}_t \rangle_w + \mathbf{R}_w \right) \end{aligned} \quad (2.46)$$

the divergence of the GW momentum flux tensor

$$\mathbf{R}_w = -\rho_r \langle \mathbf{v}' \mathbf{v}' \rangle_w \quad (2.47)$$

appears as an additional force. Again, this arises due to the quadratic non-linearity in the momentum advection. The Eulerian-averaged thermodynamic equation,

$$\begin{aligned} \rho_r \left(\partial_t + \langle \mathbf{v} \rangle_w \cdot \nabla \right) \langle h_{an} \rangle_w = & -\rho_r g \langle w \rangle_w - \rho_r \langle b \rangle_w \langle w \rangle_w \\ & - \nabla \cdot \left(\langle \mathbf{q} \rangle_w + \langle \mathbf{J}_t \rangle_w + \mathbf{J}_w \right) + \left(\langle \mathbf{F} \rangle_w + \langle \mathbf{R}_t \rangle_w \right) \cdot \nabla \langle \mathbf{v} \rangle_w + \rho_r \left(\langle Q \rangle_w + E \right), \end{aligned} \quad (2.48)$$

has two different GW heating sources. The first results from the convergence of GW sensible heat-flux

$$\mathbf{J}_w = \rho_r \langle \mathbf{v}' h'_{an} \rangle_w \quad (2.49)$$

which originates from the non-linear advection of specific anelastic enthalpy, whereas the second is the rate of GW energy deposition

$$E = - \langle b' w' \rangle_w + \frac{1}{\rho_r} \langle (\mathbf{F}' + \mathbf{R}'_t) \cdot \nabla \mathbf{v}' \rangle_w \quad (2.50)$$

and emerges from the non-linear nature of mechanical dissipation and vertical entropy transport. According to *Becker* [2004], both induce an additional heating of the Eulerian-mean flow. As the structures of the heat sources Q and heat fluxes \mathbf{q} and \mathbf{J}_t were not specified, possible interactions between radiation and gravity waves as well as chemical reactions and gravity waves are hidden in the mean terms, $\langle Q \rangle_w$, $\langle \mathbf{q} \rangle_w$ and $\langle \mathbf{J}_t \rangle_w$. As discussed by *Akmaev* [2007], these GW impacts also contribute to the GW energy deposition onto the mean flow. The GW impacts on the Eulerian-mean flow are summarized in the box below:

$$\mathbf{f}_{gw} = -\frac{1}{\rho_r} \nabla \cdot (\rho_r \langle \mathbf{v}' \mathbf{v}' \rangle_w) \quad : \text{GW force} \quad (2.51)$$

$$Q_{gw} = -\frac{1}{\rho_r} \nabla \cdot (\rho_r \langle \mathbf{v}' h'_{an} \rangle_w) + E \quad : \text{GW heating rate} \quad (2.52)$$

$$E = - \langle b' w' \rangle_w + \epsilon_{gw} \quad : \text{GW energy deposition} \quad (2.53)$$

$$\epsilon_{gw} = \frac{1}{\rho_r} \langle (\mathbf{F}' + \mathbf{R}'_t) \cdot \nabla \mathbf{v}' \rangle_w \quad : \text{GW dissipation rate} \quad (2.54)$$

A very short introduction to the Generalized Lagrangian-mean theory

The generalized Lagrangian-mean (GLM) theory invented by *Andrews and McIntyre* [1978] explains how to construct a Lagrangian-mean model including the representation of subgrid-scale effects. Its main features are sketched here briefly. The interested reader is referred to the original article or to a didactic introduction given in *Bühler* [2009].

As mentioned above, a set of fluid trajectories results from the different realizations of the subgrid-scale process. In GLM theory, the coordinate \mathbf{x} on the large-scale grid is interpreted as the location of the ensemble-mean fluid trajectory.³ The trajectory of an individual realization is displaced to the position $\mathbf{x}^\xi = \mathbf{x} + \boldsymbol{\xi}$ where $\boldsymbol{\xi}(\mathbf{x}, t)$ denotes the displacement of the fluid trajectory.⁴ These relations are sketched in fig. 2.1. Each variable can be evaluated at the mean position, e.g. $\mathbf{v} = \mathbf{v}(\mathbf{x}, t)$, and at the actual fluid position, denoted by $\mathbf{v}^\xi = \mathbf{v}(\mathbf{x}^\xi, t)$. The lift from the mean to the actual position defines a map $\mathbf{x} \rightarrow \mathbf{x} + \boldsymbol{\xi}(\mathbf{x}, t)$, the “*lifting map*”, and is denoted by the superscript ξ . Note that also expressions including differential operators can be lifted, such as $\nabla \phi \rightarrow (\nabla \phi)^\xi$, but the lifting does generally not commute with the operators, i.e. $(\nabla \phi)^\xi \neq \nabla(\phi^\xi)$. The Lagrangian average was defined in section 2.2.1. With the present notion of the lifting map it can be written $\langle \mathbf{v} \rangle_w^L = \langle \mathbf{v}^\xi \rangle_w$, where the fluid variable is evaluated at the

³ This is comparable to the semi-Lagrangian approach as each time step a new mean trajectory is chosen which ends at the large-scale grid point.

⁴ In the following, the notation of the ensemble parameter α is suppressed

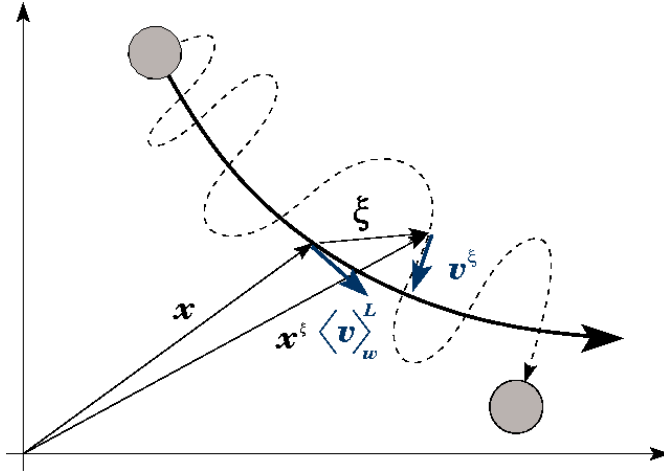


Figure 2.1.: A sketch of a mean trajectory (thick solid line) and one realization of an actual fluid trajectory (dashed line). The mean position \mathbf{x} , the actual position \mathbf{x}^ξ and the displacement $\boldsymbol{\xi}$ are plotted, and both, the actual fluid velocity \mathbf{v}^ξ and the Lagrangian-mean velocity $\langle \mathbf{v} \rangle_w^L$, are shown (after *Bühler and McIntyre* [1998]).

actual position \mathbf{x}^ξ of each individual trajectory. Then the Lagrangian disturbance is $\mathbf{v}^\ell = \mathbf{v}^\xi - \langle \mathbf{v} \rangle_w^L$.

In the following, transformations of infinitesimal line, surface and volume elements under the lifting map are considered [*Bühler*, 2009]. These transformation rules are important in the derivation of Lagrangian-mean dynamics. A line element is distorted via

$$d\mathbf{x}^\xi = d\mathbf{x} \cdot \nabla \mathbf{x}^\xi = d\mathbf{x} \cdot (\mathbf{1} + \nabla \boldsymbol{\xi}) , \quad (2.55)$$

where the chain rule was applied and $\mathbf{1}$ denotes the identity matrix. A surface element is rotated and deformed via the transformation matrix \mathbf{K} . The lifted surface element obeys

$$d\mathbf{A}^\xi = \mathbf{K} \cdot \mathbf{A} , \quad (2.56)$$

where \mathbf{K} has to be determined. Lastly, the change in size of a volume element is given by the Jacobian $J = |\nabla \mathbf{x}^\xi|$ of the transformation, i.e.

$$dV^\xi = J dV . \quad (2.57)$$

The last relation can lead to an explicit expression for \mathbf{K} . A volume element is determined by a surface element and a line element,

$$dV^\xi = d\mathbf{x}^\xi \cdot d\mathbf{A}^\xi = d\mathbf{x} \cdot \nabla \mathbf{x}^\xi \cdot \mathbf{K} \cdot d\mathbf{A} = J dV . \quad (2.58)$$

As the last equality should hold for any meaningful combination of $d\mathbf{x}$ and $d\mathbf{A}$, the transformation \mathbf{K} is obtained by inversion, i.e.

$$\mathbf{K} = J \left(\nabla \mathbf{x}^\xi \right)^{-1} = |\nabla \mathbf{x}^\xi| \left(\nabla \mathbf{x}^\xi \right)^{-1} . \quad (2.59)$$

For displacements $\boldsymbol{\xi}$ small compared to the local wavelength of a perturbation, a linear approximation is

$$\mathbf{K} \approx (1 + \nabla \cdot \boldsymbol{\xi}) (\mathbf{1} - \nabla \boldsymbol{\xi}) \approx (1 + \nabla \cdot \boldsymbol{\xi}) \mathbf{1} - \nabla \boldsymbol{\xi} , \quad (2.60)$$

where the Jacobian is approximately the trace of $\nabla \mathbf{x}^\xi$, i.e. $J \approx 1 + \nabla \cdot \boldsymbol{\xi}$ for small $\boldsymbol{\xi}$.

Lifted anelastic dynamics and Lagrangian average

The transformation rules for dV^ξ and $d\mathbf{A}^\xi$ are now used to derive some integral relationships for conservation laws. For instance, the reference mass within a small volume \mathcal{V}^ξ centered at position \mathbf{x}^ξ is given by

$$\int_{\mathcal{V}^\xi} dV \rho_r = \int_{\mathcal{V}} dV^\xi \rho_r^\xi = \int_{\mathcal{V}} dV J \rho_r^\xi. \quad (2.61)$$

A sequence of two variable changes $dV \rightarrow dV^\xi \rightarrow dV$ was performed. The volume \mathcal{V} centered at position \mathbf{x} moves with the velocity of the mean trajectory, thus with $\langle \mathbf{v} \rangle_w^L$. Therefore, the lifted anelastic continuity equation at \mathbf{x}^ξ is

$$\frac{d}{dt} \int_{\mathcal{V}} dV (J \rho_r^\xi) = 0 \quad \Leftrightarrow \quad \partial_t (J \rho_r^\xi) + \nabla \cdot (\langle \mathbf{v} \rangle_w^L J \rho_r^\xi) = 0. \quad (2.62)$$

The quantity $\tilde{\rho}_r = J \rho_r^\xi$ might be called pseudo-reference density. $\tilde{\rho}_r$ is a mean quantity because it obeys the same dynamics as $\langle \tilde{\rho}_r \rangle_w$ and no initial shift between the two is assumed. From $\tilde{\rho}_r = \langle \tilde{\rho}_r \rangle_w$ a simple linear relation for the continuity equation arises [Bühler, 2009; Senf and Achatz, 2012]

$$\tilde{\rho}_r = J \rho_r^\xi \approx (1 + \nabla \cdot \boldsymbol{\xi})(\rho_r + \boldsymbol{\xi} \cdot \nabla \rho_r) \approx \rho_r \Rightarrow \quad \nabla \cdot (\rho_r \boldsymbol{\xi}) \approx \nabla \cdot (\tilde{\rho}_r \boldsymbol{\xi}) \approx 0 \quad (2.63)$$

Next, the major simplification appears in the advection term, which is the overall advantage of the GLM theory. Using conservation of reference mass, the anelastic momentum within \mathcal{V}^ξ changes according to

$$\frac{d}{dt} \int_{\mathcal{V}^\xi} dV \rho_r \mathbf{v} = \frac{d}{dt} \int_{\mathcal{V}} dV \tilde{\rho}_r \mathbf{v}^\xi = \int_{\mathcal{V}} dV \tilde{\rho}_r (\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla) \mathbf{v}^\xi. \quad (2.64)$$

The advection of any quantity at \mathbf{x}^ξ happens with Lagrangian-mean velocity, thus $(D_t \mathbf{v})^\xi = (\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla) \mathbf{v}^\xi$. Therefore, when the non-linear advection is averaged in a Lagrangian sense, no second moments of subgrid-scale perturbations result. This is the main difference to the Eulerian-mean technique.

For the lifting of flux divergences, the theorem of Gauss is applied at the fluid position \mathbf{x}^ξ , e.g.

$$\begin{aligned} \int_{\mathcal{V}^\xi} dV \nabla \cdot \mathbf{R}_t &= \iint_{\partial \mathcal{V}^\xi} d\mathbf{A} \cdot \mathbf{R}_t = \iint_{\partial \mathcal{V}} d\mathbf{A}^\xi \cdot \mathbf{R}_t^\xi = \iint_{\partial \mathcal{V}} (\mathbf{R}_t^\xi)^T \cdot d\mathbf{A}^\xi \\ &= \iint_{\partial \mathcal{V}} (\mathbf{R}_t^\xi)^T \cdot \mathbf{K} \cdot d\mathbf{A} = \iint_{\partial \mathcal{V}} d\mathbf{A} \cdot \mathbf{K}^T \cdot \mathbf{R}_t^\xi = \int_{\mathcal{V}} dV \nabla \cdot (\mathbf{K}^T \cdot \mathbf{R}_t). \end{aligned} \quad (2.65)$$

Hence, each flux tensor is rotated and deformed by \mathbf{K}^T to compensate for changes of the surface through which the flux goes. Finally, a recipe for lifting the anelastic set is obtained:

- (i) lift of a quantity b via $\rho_r b \rightarrow \tilde{\rho}_r b^\xi$
- (ii) Lagrangian mean advection $\rho_r (\partial_t + \mathbf{v} \cdot \nabla) \mathbf{v} \rightarrow \tilde{\rho}_r (\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla) \mathbf{v}^\xi$
- (iii) transformation of lifted flux tensor in the divergence $\nabla \cdot \mathbf{R}_t \rightarrow \nabla \cdot (\mathbf{K}^T \cdot \mathbf{R}_t^\xi)$

Applying these rules to the anelastic momentum eq. (2.37) and anelastic enthalpy eq. (2.38), their lifted counterparts are

$$\left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla\right) \mathbf{v}^\xi + 2\mathbf{\Omega}^\xi \times \mathbf{v}^\xi = -\left(\nabla\phi\right)^\xi + b^\xi \mathbf{e}_z^\xi + \frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\mathbf{K}^T \cdot (\mathbf{F}^\xi + \mathbf{R}_t^\xi)\right), \quad (2.66)$$

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla\right) h_{an}^\xi &= -\left(g + b^\xi\right) w^\xi - \frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\mathbf{K}^T \cdot (\mathbf{q}^\xi + \mathbf{J}_t^\xi)\right) \\ &+ \epsilon^\xi + Q^\xi, \end{aligned} \quad (2.67)$$

where for convenience the notion of the turbulent filter operator $\langle \cdot \rangle_t$ was dropped and \mathbf{v}^ξ is interpreted as turbulence-averaged velocity vector lifted to the fluid position \mathbf{x}^ξ . Note that the term $(\nabla\phi)^\xi$ will be the major source of GW forces. Next, the set of lifted equations is averaged which then forms the basis of the Lagrangian mean dynamics

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla\right) \langle \mathbf{v} \rangle_w^L + \left\langle 2\mathbf{\Omega}^\xi \times \mathbf{v}^\xi \right\rangle_w &= -\left\langle (\nabla\phi)^\xi \right\rangle_w + \left\langle b^\xi \mathbf{e}_z^\xi \right\rangle_w \\ &+ \frac{1}{\tilde{\rho}_r} \nabla \cdot \left\langle \mathbf{K}^T \cdot (\mathbf{F}^\xi + \mathbf{R}_t^\xi) \right\rangle_w, \end{aligned} \quad (2.68)$$

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla\right) \langle h_{an} \rangle_w^L &= -\left(g + \langle b \rangle_w^L\right) \langle w \rangle_w^L - \left\langle b^\ell w^\ell \right\rangle_w \\ &- \frac{1}{\tilde{\rho}_r} \nabla \cdot \left\langle \mathbf{K}^T \cdot (\mathbf{q}^\xi + \mathbf{J}_t^\xi) \right\rangle_w + \langle \epsilon \rangle_w^L + \langle Q \rangle_w^L. \end{aligned} \quad (2.69)$$

In some parts, the Lagrangian-mean system resembles the Eulerian-mean system. It is the appearance of the several lifted terms, e.g. $\langle (\nabla\phi)^\xi \rangle_w$, that makes further simplification necessary. Therefore, two major approximations are imposed:

- (i) “*the scale-separation assumption*”: smallness of ξ compared to the background scales, and
- (ii) “*the small-amplitude assumption*”: smallness of ξ compared to the corresponding GW length,

After considerable manipulations (discussed in appendix A.1), the final set of Lagrangian-mean equations is

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla\right) \langle \mathbf{v} \rangle_w^L + 2\mathbf{\Omega} \times \langle \mathbf{v} \rangle_w^L &= -\nabla \langle \phi \rangle_w + \langle b \rangle_w^L \mathbf{e}_z + \frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\langle \mathbf{F} \rangle_w + \langle \mathbf{R}_t \rangle_w\right) \\ &+ \frac{1}{\tilde{\rho}_r} \nabla \cdot \left(-\tilde{\rho}_r \langle \xi \nabla \phi' \rangle_w + \langle \xi \nabla \cdot (\mathbf{F}' + \mathbf{R}_t') \rangle_w\right), \end{aligned} \quad (2.70)$$

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla\right) \langle h_{an} \rangle_w^L &= -\left(g + \langle b \rangle_w^L\right) \langle w \rangle_w^L - \langle b' w' \rangle_w \\ &- \frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\langle \mathbf{q} \rangle_w + \langle \mathbf{J} \rangle_w + \langle \xi \nabla \cdot (\mathbf{q}' + \mathbf{J}_t') \rangle_w\right) + \epsilon_m + \epsilon_{gw} + \langle Q \rangle_w^L, \end{aligned} \quad (2.71)$$

where ϵ_m and ϵ_{gw} denote the mean dissipation rates of the resolved flow and of the subgrid-scale gravity waves, respectively. The term $\tilde{\rho}_r \langle \xi \nabla \phi' \rangle_w$ can be interpreted as the average work of the GW displacements against the GW pressure perturbations [McIntyre, 1980]. By integration by parts, $-\tilde{\rho}_r \langle \xi \nabla \phi' \rangle_w \approx \tilde{\rho}_r \langle \phi' (\nabla \xi)^T \rangle_w$ is the work done by the perturbation pressure $\tilde{\rho}_r \phi'$ against the corrugated fluid surface. In mountain wave studies, the term $\langle p' \partial_x \zeta \rangle_w$ is known as the zonal form drag exerted on the mountain

ridge by the flow where ζ is the displacement of the flow by the topography [Nappo, 2002]. Bretherton [1969] pointed out that Jones [1967] had shown that $\langle p' \partial_x \zeta \rangle_w$ can be “interpreted [...] in terms of the vertical flux of angular momentum”. Furthermore, with reference to Jones [1967] and Bretherton [1969], Smith [1979] considered it as the “correct form of wave drag in a rotating fluid”. Similarly, the term $\langle \boldsymbol{\xi} \nabla \cdot (\mathbf{F}' + \mathbf{R}'_t) \rangle_w$ is the average work of the GW displacements against the molecular and turbulent stresses due to GW motion. The corresponding Lagrangian-mean GW force is

$$\mathbf{f}_{gw}^L = -\frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\tilde{\rho}_r \langle \boldsymbol{\xi} \nabla \phi' \rangle_w - \langle \boldsymbol{\xi} \nabla \cdot (\mathbf{F}' + \mathbf{R}'_t) \rangle_w \right) \quad : \text{GW force} \quad (2.72)$$

The second part, which couples the GW and the turbulence parameterizations, will be neglected in the remainder of the thesis. This is justified only if the wave Reynolds number is sufficiently large. For the thermal impact of gravity waves, the convergence of sensible heat-flux is absent. GW heating results from the vertical GW buoyancy flux, the GW dissipation rate and additionally from the interaction between the GW parameterization and the molecular and turbulent diffusive fluxes. Hence, the Lagrangian-mean heating due to gravity waves is

$$Q_{gw}^L = -\frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\langle \boldsymbol{\xi} \nabla \cdot (\mathbf{q}' + \mathbf{J}'_t) \rangle_w \right) + E \quad : \text{GW heating rate} \quad (2.73)$$

$$E = -\langle b' w' \rangle_w + \epsilon_{gw} \quad : \text{GW energy deposition} \quad (2.74)$$

$$\epsilon_{gw} = \frac{1}{\rho_r} \langle (\mathbf{F}' + \mathbf{R}'_t) \cdot \nabla \mathbf{v}' \rangle_w \quad : \text{GW dissipation rate} \quad (2.75)$$

Note that in the above formulation, the energy deposition E is equal in the Eulerian and Lagrangian average for the imposed approximations. It is up to the modeler whether the Eulerian-mean forces and heating rates, \mathbf{f}_{gw} and Q_{gw} , or their Lagrangian-mean counterparts, \mathbf{f}_{gw}^L and Q_{gw}^L , are preferred. Comments on that can be found in Bretherton [1969] who stated that “the force exerted across a surface always consisting of the same material particles differs in this instance from the force exerted across surface $z = \text{constant}$. [...] Thus there is an essential difference between equations formulated in terms of Eulerian means (at a fixed point or in horizontal planes) or Lagrangian means (following fluid particles), and care is needed before applying the results to ensure that the formulation is internally consistent.” The discussion of that topic will be renewed in section 4.4 with the focus on generalized Eliassen-Palm theorems and in section 5.5 concerning the analysis of GW forces obtained from global ray-tracing experiments.

2.3. Summary

- The dynamical and thermal impacts of subgrid-scale gravity waves and turbulence were consistently derived.
- For the fundamental compressible equations, the partitioning between acoustic and available potential energies was presented.
- The anelastic system which inhibits propagating acoustic waves was reviewed and several implications were discussed.
- After averaging over turbulent scales, the filtered anelastic system remains structurally equal to its unfiltered counterpart, only turbulent heat and momentum fluxes are added to the molecular fluxes (quasi-stationary turbulence).
- The dynamics of the mean flow is obtained by a second average over GW scales.
- Two types of averages, the Eulerian-mean and the Lagrangian-mean, were introduced and discussed.
- In the Eulerian-mean flow, GW forces result from the divergence of the momentum-flux tensor and heating rates are due to the convergence of sensible heat fluxes, vertical entropy fluxes and dissipation of GW mechanical energy.
- In contrast, in the Lagrangian-mean flow, GW forces are induced by the divergence of GW stresses which arise from the average work of GW displacements against GW pressure perturbations and frictional stresses.
- Furthermore, for GW impacts on the Lagrangian-mean thermodynamics, the convergence of sensible fluxes is absent and an additional interaction between gravity waves and the diffusivity-parameterization appears.

3. Overview of diurnal tides

Thermal tides are introduced in this chapter. Special focus is on diurnal variations in the middle atmosphere excited by the daily cycle in solar heating. Both the diurnal heating rates and tides are obtained from the global climate model HAMMONIA which is briefly presented.

3.1. HAMMONIA model

The HAMMONIA model is a spectral general circulation model with coupled chemistry. It is an upward extension of the ECHAM5 model up to the lower thermosphere and built and operated at the Max Planck Institute of Meteorology at Hamburg [*Schmidt et al.*, 2006]. The flow dynamics are based on a set of compressible, hydrostatic equations equivalent to *Becker* [2003b], with a generalized vertical coordinate η including the effect of orography. It was shown by several studies that simulation results from HAMMONIA compare quite well with recent observations [e.g. *Achatz et al.*, 2008; *Yuan et al.*, 2008] and that it is one of the leading state-of-the-art climate models of the middle atmosphere. This makes HAMMONIA simulations highly valuable for this study of gravity wave-tide interactions. However, it is questionable whether all sub-grid parameterizations in HAMMONIA are consistent with the general conservation law structure of the underlying dynamics. Several issues concerning the frictional heating due to molecular and turbulent motion and direct heating due to GW breakdown in current state-of-the-art general circulation models are discussed by *Becker* [2001]; *Becker and Schmitz* [2002]; *Becker* [2004]; *Burkhardt and Becker* [2006].

Global horizontal wind, temperature and geopotential data have been provided from a twenty year time slice experiment from 1980 to 1999 in typical solar maximum conditions. The simulation output has a spectral truncation at T31¹ and 67 vertical levels. The corresponding equivalent horizontal grid spacing is about 700 km and the distance between vertical levels is about one kilometer in the lower atmosphere and increases to three kilometers in the mesosphere. The data were re-gridded to a 48×48 horizontal grid. Monthly averaged values at eight different times a day with an increment of 3 hours have been taken to calculate a monthly-mean diurnal cycle. In the current study, only mean January data, representative of Northern hemisphere winter conditions, are used. With the help of a Fourier analysis in time, the monthly-mean diurnal cycle has been analyzed for the temporal average and the average diurnal tide. Semi-diurnal and shorter-period tidal variations have been excluded from the present analysis.

Beside several other physical parameterizations, HAMMONIA applies the Hines parameterization (see e.g. *McLandress* [1998]) to estimate the so-called non-orographic GW drag. Hence, the tidal data used in this study already felt the impact of gravity waves due to the parametrized forces and heating rates. As will be shown, it is questionable whether these GW effects on thermal tides are quantitatively as well as qualitatively realistic due to the drastic assumptions made in the parameterization.

¹ Spectral truncation: The model equations are expanded in spherical harmonics $Y_n^{m'}$ with the total wavenumber n' and the zonal wavenumber m' . The triangular truncation T31 means that the maximal m' is set to the maximal $n' = 31$.

3.2. Diurnal heating rates

The day-side of the earth is heated by solar radiation. The emissions by atmospheric greenhouse gases cool the atmosphere. Due to the rotation of the earth, the local heating rates are approximately periodic with a period of a day. The resulting daily cycle is dominated by the first harmonic, but higher harmonics are also excited by inherent asymmetries in the heating signal. For instance, latent heating due to deep convection is highly intermittent and more likely to appear in the afternoon.

HAMMONIA incorporates several processes which lead to a periodic heating of the atmosphere. Contributions are grouped into:

- large-scale cumulus convection and condensation heating
- heating due to absorption of solar insolation which are sorted into three bands: (i) 5 - 105 nm, (ii) 105 - 250 nm and (iii) ≥ 250 nm
- cooling by infrared emissions of atmospheric greenhouse gases, especially water vapor and carbon dioxide
- frictional heating due to molecular and turbulent diffusion, convergence of molecular and turbulent heat fluxes, energy deposition of gravity waves, convergence of GW entropy fluxes (as argued before, these effects may not be consistently described in HAMMONIA)

The sum of all different contributions is shown in fig. 3.1 at two times, 06:00 UT and 12:00 UT, in the middle troposphere and stratopause region. In the troposphere,

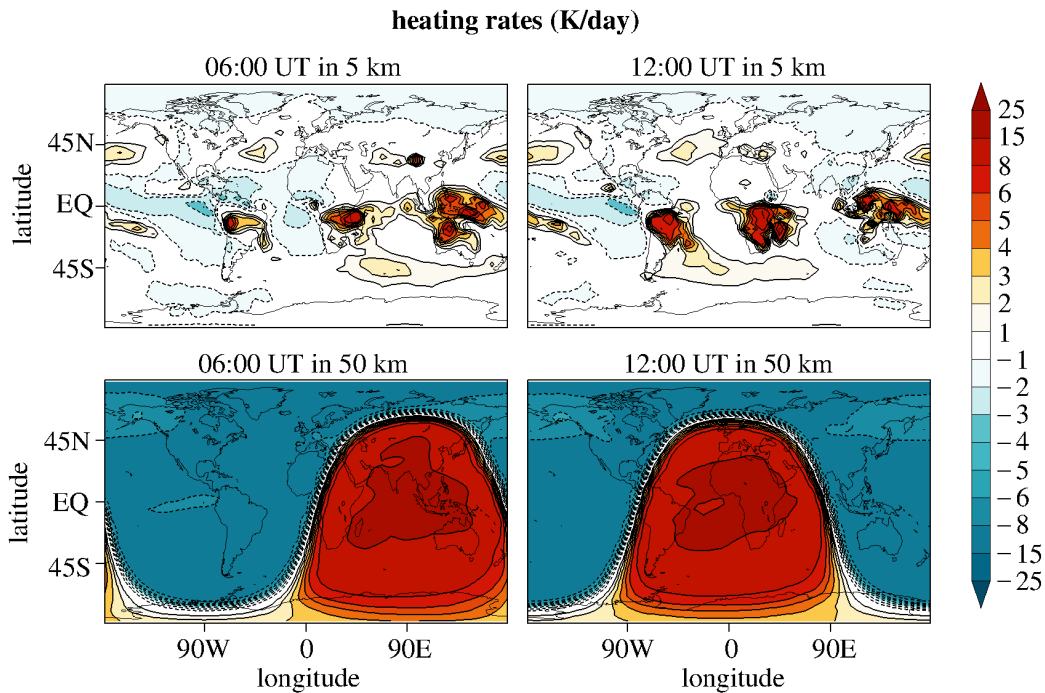


Figure 3.1.: The sum of all heating rates in HAMMONIA at two times, 06:00 UT (left column) and 12:00 UT (right column). In the mid-troposphere (upper row) the heating rates are dominated by latent heating and direct radiative effects of water vapor, whereas in the stratopause region (lower row) ozone has the strongest impact on heating rates.

deep convection over Indonesia, and the Congo and Amazon basins lead to a local heating of the atmosphere of up to 10 K per day. Large-scale condensation processes are also important in the troposphere, especially in the mid-latitude Atlantic. In the stratopause region (about 50 km), the heating rates are about two times larger than in the troposphere and mainly induced by the absorption of ultraviolet solar radiation by ozone and infrared emission by carbon dioxide. The ozone heating only occurs in the day-time due to incident sun light. The hemispheric asymmetry of ozone heating is due to the inclination of solar insolation angle (Northern winter). Infrared emissions of carbon dioxide are much less affected by the daily cycle.

The zonally averaged amplitude of the diurnal heating rates extracted from the HAMMONIA simulations is shown in fig. 3.2. There are two main regions where diurnal heat-

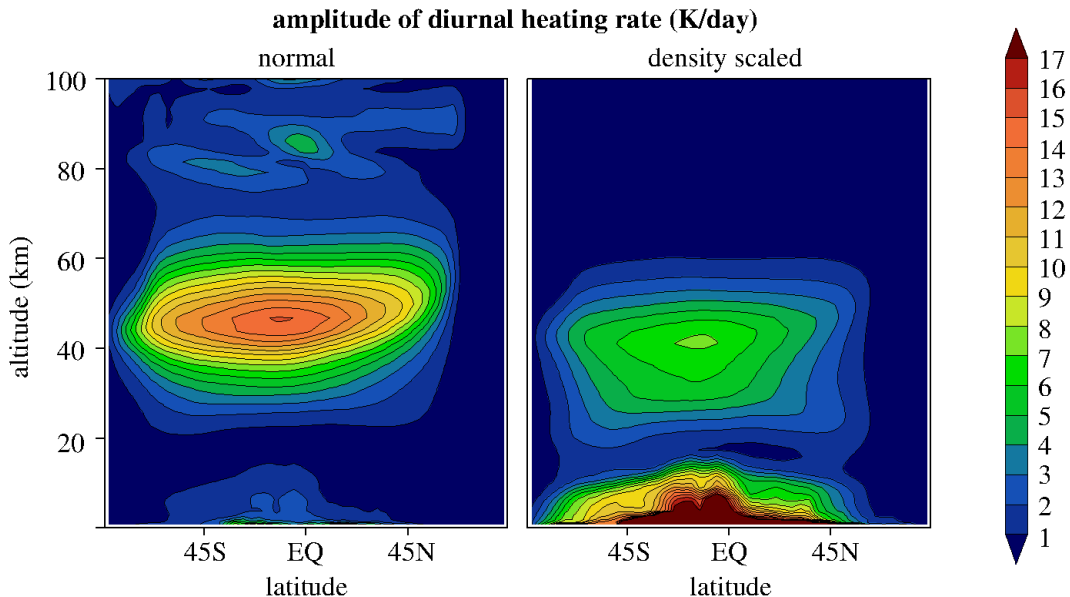


Figure 3.2.: Total amplitudes of diurnal heating rates (left) and density-scaled heating rates (right). In the scaling factor $\sqrt{\rho_r/\rho_{00}}$, a reference density of $\rho_{00} = 0.01 \text{ kg m}^{-3}$ was chosen.

ing is large: (i) in the troposphere by the impact of convection, condensation and direct radiative effects of water vapor, and (ii) in the stratopause region by periodic heating due to ozone. The diurnal heating amplitudes have maxima of about 5 and 15 K per day for the troposphere and the stratopause region respectively. Thus, a significant amount of the heating shown in fig. 3.1 is connected to higher harmonics.

The left part of fig. 3.2 seems to suggest that the diurnal tide is mainly excited close to the stratopause. However, this underestimates the impact of the troposphere. As discussed by *Lindzen* [1966, 1967], the relevant forcing in the tidal equations is a density-scaled heating rate given by $\sqrt{\rho_r/\rho_{00}} Q$, where ρ_{00} is an arbitrary reference density. The impact of troposphere is thus more than a factor 8 larger than in the regions above. The concept of a rescaled heating rate is explained by the fact that throughout the middle atmosphere, the density decrease strongly enhances tidal amplitudes as less dense air is transported by tidal oscillations. Since the variations of kinetic energy of the thermal tides due to shear and buoyant production as well as dissipation processes are mostly of the same order, the wind amplitudes increase like the square root of inverse density, i.e. $u_T \propto \rho_r^{-1/2}$, where u_T denotes the diurnal wind perturbation. The tidal temperature perturbations grow similarly. Hence, if the density factor is eliminated from the thermodynamics, the heating rate changes to $Q \rightarrow \sqrt{\rho_r/\rho_{00}} Q$. Therefore, the

troposphere has an enormous impact on the formation of thermal tides.

3.3. Diurnal tides

The periodic heating of the troposphere and stratopause region induces a large-scale temperature perturbation, causing buoyant motion which feeds back into the horizontal kinetic energy of the thermal tides and excites horizontal tidal winds. The tidal perturbation propagates away from the source region in both the upward and downward directions. The downward branch is reflected at the surface and propagates upward again. The two branches may interfere. The mean background wind and temperature affect the propagation of the tides by forming wave guides and changing the tides' modal structure [Ortland, 2005a, b]. Furthermore, resonant interactions between large-scale planetary waves and tides may form secondary tidal waves [Achatz *et al.*, 2008].

The diurnal part of the tidal variations was extracted from the HAMMONIA data via Fourier analysis. For instance, the zonal tidal wind is represented by

$$u_T = u_R \cos(\Omega t) + u_I \sin(\Omega t), \quad (3.1)$$

where the Fourier coefficients u_R and u_I are functions of longitude λ , latitude φ and altitude z . $\Omega = 2\pi(24\text{ h})^{-1}$ is the diurnal frequency. Again for the sake of simplicity, semi-diurnal and shorter-period tidal variations have been excluded from the present analysis, with corresponding extensions left to future work. The migrating parts of diurnal variations have zonal wavenumber one and follow the apparent motion of the sun. Thus, the migrating part of the diurnal zonal wind variation is given by [Achatz *et al.*, 2008]

$$u_{T,\text{mig}} = u_{R,\text{mig}} \cos(\Omega t + \lambda) + u_{I,\text{mig}} \sin(\Omega t + \lambda). \quad (3.2)$$

As investigated by several authors [Achatz *et al.*, 2008, and references therein], the heating in equatorial convection zones and the interaction of the migrating tide with planetary waves excites diurnal variations which are not sun-synchronous. They are called non-migrating tides and defined as the residual diurnal variation after the subtraction of the migrating parts, i.e.

$$u_{T,\text{non}} = u_T - u_{T,\text{mig}}. \quad (3.3)$$

They are composed of several components with different zonal wavenumbers.

The real and imaginary parts of the diurnal zonal wind, diurnal meridional wind and diurnal temperature are shown in fig. 3.3 at 110 km altitude. The local wind maxima easily exceed 50 m/s. The zonal wind is more symmetric about the equator whereas the meridional wind tends to be more anti-symmetric. In addition to the migrating zonal wind part, non-migrating parts with zonal wavenumbers two and three are important. The meridional wind, however, is more controlled by its migrating parts. The diurnal temperature perturbation has local maxima of up to 30 K.

A vertical cross-section of the real part of zonal wind u_R is shown in fig. 3.4 for different latitudes. u_R has been scaled by the square root of density $\sqrt{\rho_r/\rho_{00}}$. There are two altitude regions where dissipation of tidal-wave energy seems to happen: in the lower stratosphere and in the upper mesosphere. In the southern hemisphere lower stratosphere (at about 20 km), the phase lines of the convectively-excited tidal components become shallower which can indicate that there exists a critical mean-wind level.²

² For critical levels of gravity waves see also the discussion in sections 4.3.1 and 5.3.5.

diurnal perturbations in 110 km

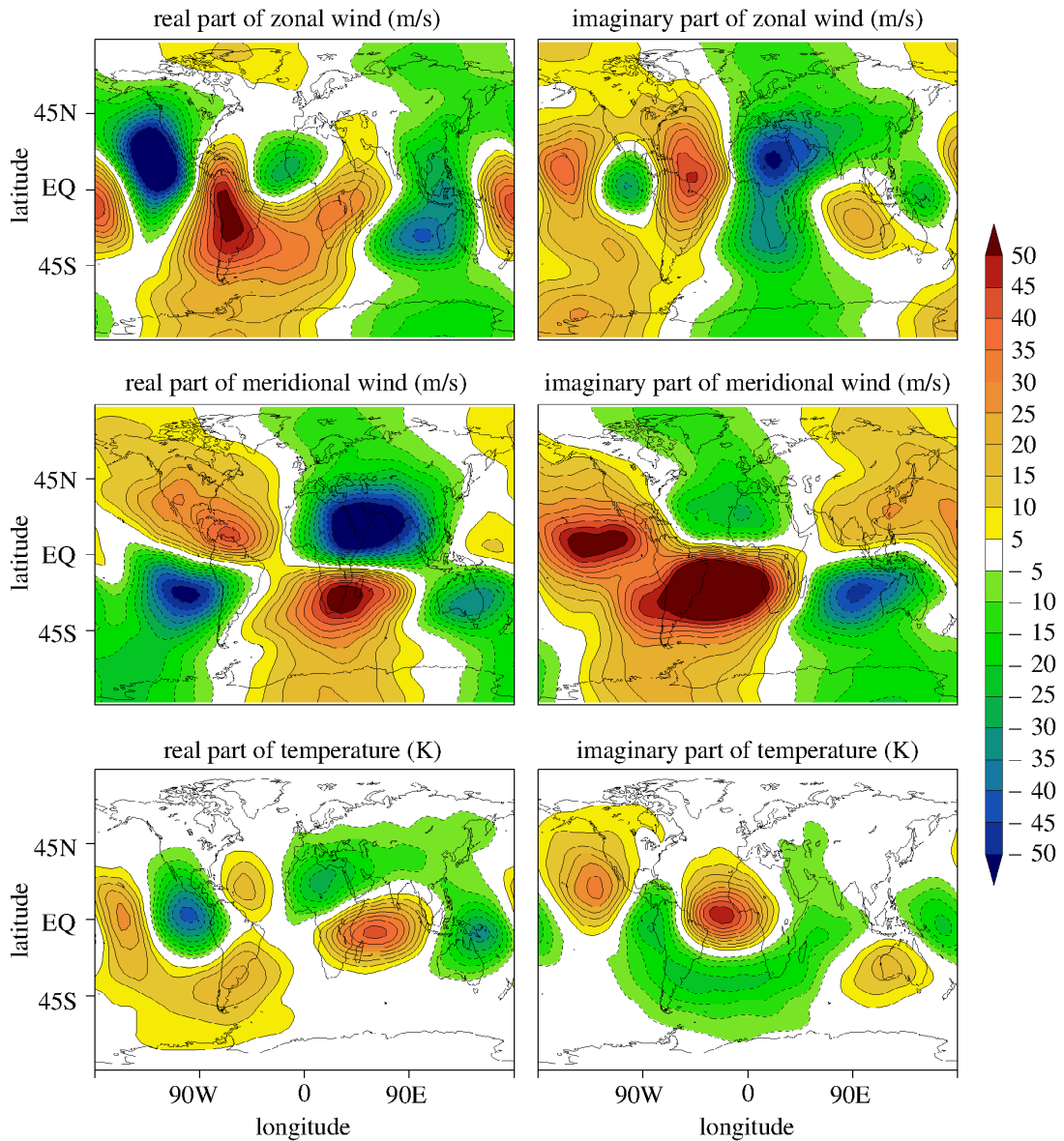


Figure 3.3.: Real part (left) and imaginary part (right) of diurnal zonal wind (upper row), diurnal meridional wind (middle row) and diurnal temperature (lower row) at 110 km altitude from HAMMONIA simulations. A contour interval of 5 m/s and 5 K was chosen for wind and temperature, respectively.

The phase lines are negatively tilted which suggests that the phase progression is westward. As there is an east-wind jet in the southern (summer) hemisphere, the difference between the tidal phase velocity and the mean wind velocity is reduced. Another hint to the critical level filtering is that the height of the extinction of the westward moving non-migrating tides rises from 30S to 0 and eventually vanishes at 15N and further northward. In the northern hemisphere, the west-wind jet supports westward tidal-wave motion. Another possible cause for the damping of the westward-moving tide at 20 km and 30S could be the destructive interference with the tidal component excited in the stratopause region. Furthermore, there exists a tendency for the non-migrating parts of the diurnal tide to be preferentially eastward-moving in the southern (summer) hemi-

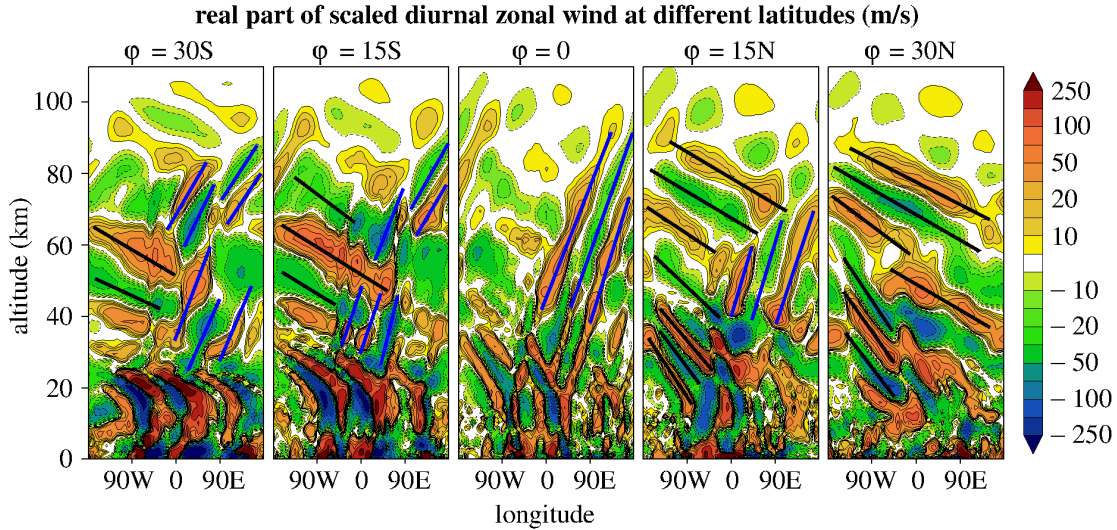


Figure 3.4.: A vertical cross-section of the real part of scaled diurnal zonal wind at different latitudes 30S, 15S, 0, 15N and 30N (from left to right) from HAMMONIA simulations. The wind was scaled with a factor $\sqrt{\rho_r/\rho_{00}}$ where a reference density of $\rho_{00} = 10^{-5} \text{ kg m}^{-3}$ was chosen. Phase lines of the westward-moving parts were indicated by black lines, whereas phases of eastward-moving parts were accentuated by blue lines.

sphere and westward-moving in the northern (winter) hemisphere. These are typical properties of (in this case large-scale) gravity waves. Another interesting feature can be seen at $\varphi = 0$ and altitude 20 km: a so-called St. Andrew's cross, where several beams are radiated away from a localized source [Nappo, 2002], but due to the typical wind conditions only the eastward-moving branch can escape. The westward-moving migrating part has larger vertical wavelength and is superposed on the non-migrating part in both hemispheres. Dissipation of tidal-wave energy in the upper mesosphere is due to molecular viscosity and thermal diffusivity. Energy originating in the lower atmosphere and transported into and dissipated in the upper atmosphere via thermal tides is one of the major coupling mechanisms between the different atmospheric layers.

The total amplitude of diurnal variations is defined in analogy to a zonal and temporal root-mean square, i.e. as

$$U = \sqrt{[u_R^2 + u_T^2]}, \quad (3.4)$$

which corresponds to the amplitude definition of *Ortland and Alexander* [2006]. The bracket $[\cdot]$ denotes the zonal average. The total amplitudes of the diurnal zonal wind U , diurnal meridional wind V and diurnal temperature disturbance are shown in fig. 3.5. The wind amplitudes have a double maximum structure with peaks at about 20N and 20S and between 100 km and 110 km. The meridional wind amplitude V is larger than the zonal wind amplitude U . A hemispheric asymmetry between northern hemisphere winter and southern hemisphere summer is also visible. The total temperature amplitude peaks in the tropics between 110 km and 120 km. The amplitudes of the migrating parts are shown in fig. 3.5. The migrating tides peak at 33 m/s and 48 m/s for the zonal and meridional winds, respectively, which is only about 60% and 85% of their total diurnal variances. The migrating temperature variation of 21 K accounts for only 35% of the diurnal variance. The amplitudes of non-migrating tides shown in fig. 3.5 have a single maximum in the equatorial region and are comparable to or, in the case of the diurnal zonal wind variation, even larger than the amplitudes of the migrating tide.

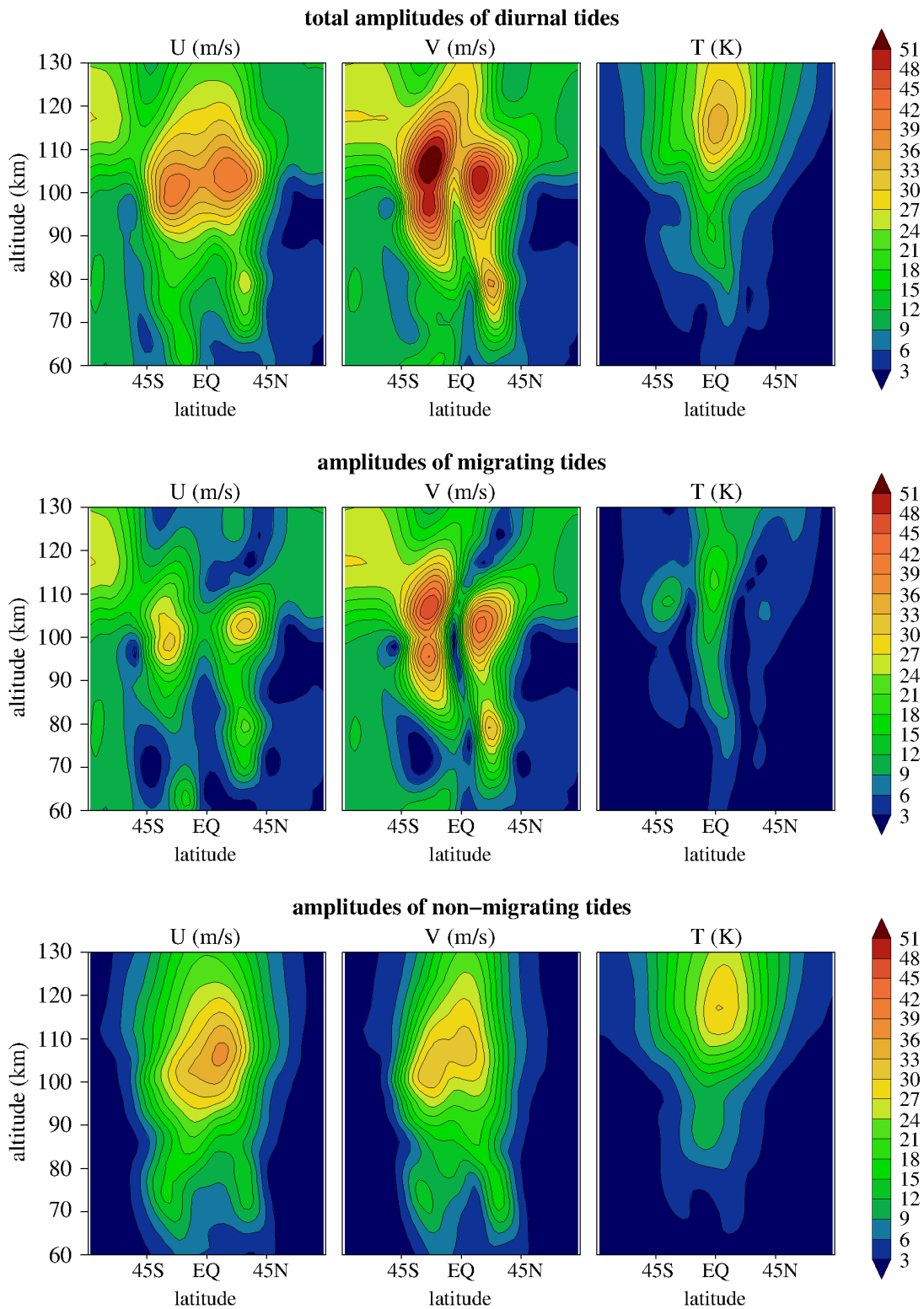


Figure 3.5.: Zonally-averaged total amplitudes of diurnal tides (upper row), and amplitudes of migrating (middle row) and non-migrating parts (lower row) of zonal wind (left column), meridional wind (middle column) and temperature (right column) obtained from HAMMONIA simulation data.

3.4. Discussion

It has been shown that tidal waves dominate the daily variability in the mesosphere/lower thermosphere with wind amplitudes up to 50 m/s and temperature perturbations of up to 30 K. Smaller scale gravity waves are strongly affected by changes in the background conditions imposed by thermal tides. This impact seems to be far from small and justifies the present interest in and detailed investigation of the gravity wave-tidal wave interaction. Another important point is that the periodic wind and temperature changes due to tides are relatively fast and can directly modify GW propagation and certain GW properties, as will be explained in chapter 4 and 5. This temporal effect is one of the major focuses of the present work and was not investigated properly in the past [as discussed in *Senf and Achatz, 2011; Achatz et al., 2012*]. In the ray-tracing simulations discussed in chapter 5, migrating as well as non-migrating parts of diurnal variations are considered as background for GW propagation.

3.5. Summary

- Simulation data from the complex chemistry-climate model HAMMONIA have been analyzed with respect to diurnal variations.
- Diurnal heating by solar insolation is strong in the troposphere due to radiative interactions of water vapor, latent heating of large-scale condensation and cumulus convection, especially in the tropics.
- Ozone heating maximizes in the upper stratosphere.
- Migrating tides follow the apparent motion of the sun and have large amplitudes in the subtropical mesopause region.
- Non-migrating tides also strongly impact the diurnal cycle in the tropical mesopause and lower thermosphere.

4. Dynamics of gravity waves

The dynamics of gravity waves (GWs) can be divided into three stages: (i) excitation, (ii) propagation and (iii) dissipation. The various, mainly tropospheric, gravity-wave sources are discussed first. Then the temporal evolution of small-amplitude waves in a resting isothermal basic state is introduced. The extension to linear gravity waves in a general moving basic state with special focus on middle-atmosphere thermal tides is addressed using multiple-scale analysis and WKB theory. A set of ray-tracing equation is derived which describes the vertical and horizontal evolution of nearly monochromatic gravity-wave trains. This method allows one to follow the wave field from its (prescribed) source to the region, where it is dissipated, and to monitor the possible changes in wave properties. In the dissipation region, here the mesosphere/lower thermosphere, gravity waves force a mean flow via deposition of pseudo-momentum and energy. The Eliassen-Palm (EP) theorems are used to explain the impact on the mean flow. As the influence of the temporal and horizontal dependence due to thermal tides on the gravity-wave propagation is taken into account, the EP-theorems are generalized.

4.1. Discussion of gravity-wave sources

Before the dynamics of gravity waves is derived explicitly, their sources, mainly located in the troposphere, will be discussed. There is a variety of different mechanisms through which gravity waves are excited, all connected to vertical lifting of air, localized body forces and heating of air. A comprehensive overview of GW sources was given by *Fritts and Alexander* [2003] and is summarized here.

In mountainous areas, a flow across topography induces oscillations in a stably-stratified environment. Propagating gravity waves result in the lee of the mountains. The horizontal scales of the mountain waves are dictated by the horizontal structure of the mountains, but their vertical wavelength is determined by the background stability and the speed of the surface wind. Because the wave source, i.e. the mountain range, does not move with respect to the geostationary frame, topographic waves have zero frequency and phase velocity.

Moist convection is also able to produce gravity waves. *Fritts and Alexander* [2003] proposed three excitation mechanisms (i) thermal forcing, (ii) a mechanical oscillator, and (iii) the obstacle or “transient mountain” mechanism. In the thermal forcing mechanism, latent heat release inside a convective cloud causes a temperature anomaly that moves away from the heating source. The extreme updraft within a convective cloud can intermittently push the cloud top against the tropopause. The moving cloud behaves like a mechanical oscillator. Finally, the cloud can be viewed as a moving obstacle which, due to convective heating, modifies the shape of the isentropes in the lowest stable layers [*Fritts and Alexander*, 2003]. Air parcels transported by the background wind rise to negotiate the cloud obstacle and GW motion is excited. The momentum flux by convectively forced gravity waves is distributed over a large range of non-zero phase velocities [*Song and Chun*, 2005].

Dynamical instabilities are another mechanism responsible for GW excitation. The collapse of a shear layer via Kelvin-Helmholtz instability can lead to the so-called enve-

lope radiation of GW packets in which a coherent sequence of Kelvin-Helmholtz billows develops. Similarly, in the adjustment process of a nearly geostrophically balanced flow, gravity waves are emitted. This happens for instance in exit regions of the tropospheric jet streams. Another lifting process occurs in the vicinity of fronts. Close to a cold front, warmer air is lifted rapidly, convective clouds are formed, and the air column is lifted. Frontal, convective and dynamical excitation of gravity waves tend to operate together and are hard to distinguish.

A variety of different GW sources exists and the properties of the excited gravity waves are extremely variable. A superposition of all types of gravity waves forms a virtually continuous GW spectrum. The interaction between gravity waves and the background flow and non-linear wave-wave interactions further shape the spectrum of gravity waves, making a quantitative description of the GW properties close to a source highly uncertain [Alexander *et al.*, 2010]. With this problem in mind, a simple toy ensemble of vertically moving gravity waves described in section 5.2.1 was chosen as the lower boundary conditions for the global ray-tracing simulations described in chapter 5. The simplicity of the employed GW ensemble made it possible to correctly attribute the tidal-wave impacts on the propagation of the gravity waves. As pointed out in chapter 6, however, major advancements are expected from the use of more realistic choices of GW sources.

4.2. Propagation of linear gravity waves in a resting, isothermal atmosphere

It is assumed that gravity waves, which are able to reach the middle atmosphere, have small amplitudes at their source altitude. If a GW packet propagates vertically over several scale-heights, its amplitude increases by orders of magnitudes. This is because less mass is transported by the GW oscillation as the mean air density decreases exponentially. Hence in its early stage, GW motion might be described within a linear frame. Furthermore, the non-linear advection terms remain quite small for nearly monochromatic gravity waves until the onset of instability.

This motivates the study of the dynamics of small-amplitude gravity waves. In the following, the anelastic equations (2.36) - (2.38) are linearized around an isothermal, hydrostatic background at rest. Linear dynamics is analyzed in Cartesian coordinates on a local tangent plane at latitude φ_0 , e.g. mid-latitudes with $\varphi_0 = 45^\circ\text{N}$. This is a typical textbook example [Pedlosky, 2003] in which several interesting properties of gravity waves can be investigated. The important extension to GW packets in moving basic states is presented in later sections. Ignoring the effects of friction and diabatic heating, the linear anelastic equations are ¹

$$\partial_t u' - f v' + \partial_x \phi' = 0, \quad (4.1)$$

$$\partial_t v' + f u' + \partial_y \phi' = 0, \quad (4.2)$$

$$\partial_t w' + \partial_z \phi' = b', \quad (4.3)$$

$$\partial_t b' + w' N^2 = 0, \quad (4.4)$$

$$\partial_x u' + \partial_y v' + \partial_z w' - w'/H_p = 0. \quad (4.5)$$

¹The parts of the Coriolis term containing the vertical velocity are ignored. As seen in section 4.3.2 inertia-gravity waves are nearly hydrostatic and the corresponding w' -terms are small. For non-hydrostatic gravity waves, the whole Coriolis term $2\mathbf{\Omega} \times \mathbf{v}'$ is negligible.

The density scale-height $H_\rho = -(\partial_z \ln \rho_r)^{-1}$ is constant and equal to RT_r/g with T_r the constant reference temperature. The square of the buoyancy frequency $N^2 = g^2/(c_p T_r)$ and the Coriolis parameter $f = 2\Omega \sin(\varphi_0)$ are also constants. The background density decreases exponentially, $\rho_r(z) = \rho_{00} \exp(-(z-z_{00})/H_\rho)$ where ρ_{00} is the reference density at a reference level z_{00} .

The solution of the set (4.1)-(4.5) is sketched in the following. To eliminate the impact of the background density ρ_r , all GW variables are scaled with the inverse square-root of ρ_r . Then the resulting equation system admits wave solutions of the form

$$\{\mathbf{v}', \phi', b'\} = \Re \left\{ \left\{ \hat{\mathbf{v}}, \hat{\phi}, \hat{b} \right\} \sqrt{\frac{\rho_{00}}{\rho_r}} e^{i\Theta} \right\}. \quad (4.6)$$

$\{\hat{\mathbf{v}}, \hat{\phi}, \hat{b}\}$ are the complex wave amplitudes of wind \mathbf{v}' , scaled pressure ϕ' and buoyancy b' . The $\rho_r^{-1/2}$ -term is responsible for the enormous increase in magnitude with height of vertically propagating gravity waves. The GW phase Θ obeys

$$\Theta = kx + ly + mz - \omega t, \quad (4.7)$$

which defines the zonal wavenumber k , meridional wavenumber l , vertical wavenumber m and ground-based frequency ω , but is only valid for this simple example. For slowly varying basic states, the definition of the wavenumber vector $\mathbf{k} = k\mathbf{e}_x + l\mathbf{e}_y + m\mathbf{e}_z$ and frequency is generalized in section 4.3.3.

The Fourier ansatz (4.6) is inserted into the set of linear partial differential equations (4.1)-(4.5) and an algebraic, homogeneous linear system results. Non-trivial solutions are only possible if the determinant of the corresponding system matrix vanishes. A relation results which connects the wave frequency ω to the wavenumbers k , l and m , the reference buoyancy frequency N and the Coriolis parameter f . This is the dispersion relation for gravity waves [Fritts and Alexander, 2003]

$$\omega^2 = \frac{N^2 k_h^2 + f^2 (m^2 + (2H_\rho)^{-2})}{k_h^2 + m^2 + (2H_\rho)^{-2}}, \quad (4.8)$$

where $k_h = \sqrt{k^2 + l^2}$ denotes the horizontal wavenumber. The term $(2H_\rho)^{-2}$ in eq. (4.8) can be viewed as a small shift of the vertical wavenumber m , becoming equal to m for a vertical wavelength of $4\pi H_\rho$. For typical scale heights of 7 km, this occurs for wavelengths larger than 100 km. For gravity waves with vertical wavelengths smaller than 40 km, the frequency shift by the $(2H_\rho)^{-2}$ -term is less than 5%. Hence, it might not be critical to neglect this term in the following discussion so that the relevant dispersion relation reduces to

$$\omega^2 = \frac{N^2 k_h^2 + f^2 m^2}{k_h^2 + m^2}. \quad (4.9)$$

There is a large range of possible wave periods for which propagating gravity waves can exist. The high-frequency limit of eq. (4.9) is obtained for large aspect ratios $k_h^2/m^2 \gg 1$, i.e. for $m^2 \rightarrow 0$, and the low frequency limit for small aspect ratios $k_h^2/m^2 \ll 1$, i.e. $m^2 \rightarrow \infty$:

$$\omega^2 \rightarrow N^2 \quad \text{for } m/k_h \gg 1 \quad : \text{ high-frequency limit}, \quad (4.10)$$

$$\omega^2 \rightarrow f^2 \quad \text{for } m/k_h \ll 1 \quad : \text{ low-frequency limit}. \quad (4.11)$$

The oscillations of fluid parcels are purely vertical in the high-frequency limit. For typical mean temperatures $T_r = 240$ K, the minimum wave period is about 5 minutes. When

a GW field approaches the limit of highest frequencies it is reflected (known as turning level). On the other hand, so-called inertial oscillations constitute the low-frequency limit, and are purely horizontal and unaffected by gravity. At $\varphi_0 = 45^\circ\text{N}$, the maximum wave period is about 17 hours. This limit is approached, for instance, near a so-called critical layer which can be responsible for an extreme increase in m^2 . In between both limits, i.e. for periods between a few minutes to about a day, free wave propagation is possible.

The dispersion relation (4.9) can be reformulated in terms of the vertical wavenumber m , i.e.

$$m^2 = k_h^2 \frac{N^2 - \omega^2}{\omega^2 - f^2} = \frac{k_h^2 N^2}{\omega^2} \frac{1 - \omega^2/N^2}{1 - f^2/\omega^2}. \quad (4.12)$$

There are three different approximations of (4.12) frequently in use [see e.g. *Fritts and Alexander, 2003*]:

- the non-hydrostatic GW regime with $\omega \gg f$, where

$$m^2 = \frac{k_h^2 N^2}{\omega^2} \left(1 - \frac{\omega^2}{N^2}\right) \Leftrightarrow \omega^2 = \frac{N^2 k_h^2}{k_h^2 + m^2} \quad (4.13)$$

- the medium-frequency GW regime with $\omega \gg f$ and $\omega \ll N$, where

$$m^2 = \frac{k_h^2 N^2}{\omega^2} \Leftrightarrow \omega^2 = \frac{N^2 k_h^2}{m^2} \quad (4.14)$$

- and the inertia-gravity wave regime with $\omega \ll N$, where

$$m^2 = \frac{k_h^2 N^2}{\omega^2 - f^2} \Leftrightarrow \omega^2 = \frac{N^2 k_h^2}{m^2} + f^2. \quad (4.15)$$

The non-hydrostatic GW regime and the inertia-gravity wave regime are branches of the full dispersion relation which show the asymptotic behavior of either the high- or low-frequency limits. The medium-frequency regime, in between the two, is the most simple, valid for wave periods on the order of hours. It is usually used in GW parameterizations [*McLandress, 1998*]. As shown in section 4.3.2, the different scaling behavior of the non-hydrostatic and the inertia-GW regimes, i.e. $\omega \sim N$ for the first and $\omega \sim f$ for the latter, makes separate multiple-scale analyses necessary when a slowly varying background state is permitted.

The motion of a GW field is characterized by two sets of velocities. The phase velocities $c_h = \omega/k_h$ and $c_z = \omega/m$ describe the motion of GW phase lines in the horizontal and vertical directions, respectively. The group velocities

$$c_{gh} = \frac{\partial \omega}{\partial k_h} = \frac{k_h}{|\mathbf{k}|^2} \frac{N^2 - \omega^2}{\omega} \quad (4.16)$$

and

$$c_{gz} = \frac{\partial \omega}{\partial m} = -\frac{m}{|\mathbf{k}|^2} \frac{f^2 - \omega^2}{\omega} \quad (4.17)$$

are even more important and describe the horizontal and vertical propagation of the envelope of a GW packet [*Pedlosky, 2003*]. The wave energy is transported by the group

wavenumber	> 0	< 0
zonal wavenumber k	eastward	westward
meridional wavenumber l	northward	southward
vertical wavenumber m	downward	upward

Table 4.1.: Sign convention for GW wavenumbers. The given directions indicate the GW group flow.

velocity. c_{gh} and c_h have the same sign, but c_{gz} and c_z have opposite signs. Hence for upward moving GW energy, the GW phases move downward. The wavenumber vector and the group velocity form a right angle. GW motion is dispersive. Assuming $\omega > 0$, GW group propagation is eastward for $k > 0$, northward for $l > 0$ and upward for $m < 0$. The sign convention is summarized in tab. 4.1.

The structure of a simple Gaussian GW packet is plotted in fig. 4.1. GW winds are shifted by 90 relative to the buoyancy or potential temperature field. Upward winds are associated with upward displacements of air parcels with smaller background entropy than the air above and thus lead to negative temperature perturbations.

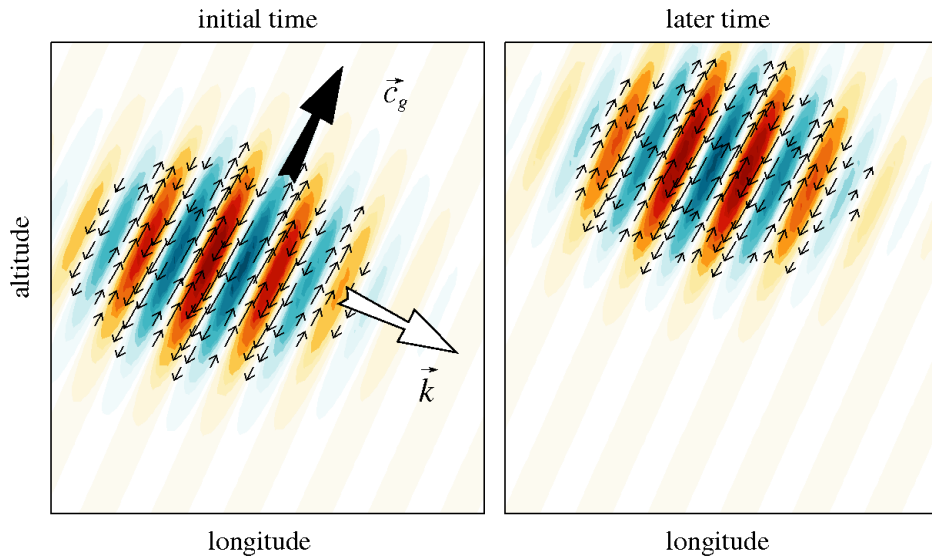


Figure 4.1.: Phase structure of a Gaussian GW packet. The buoyancy field is plotted in colors with positive areas in red and negative areas in blue, and the wind is shown with arrows in arbitrary units. The group velocity c_g (large filled arrow) and the wavenumber vector k (large open arrow) are also shown.

4.3. Propagation of linear gravity waves in a general moving basic state

4.3.1. Possible impacts of background wind and stability

So far, the impact of the background wind which advects the GW field has been neglected. The speed by which GW phase crests approach an observer depends on whether the observer is fixed to the ground or moves with the mean wind.² Hence in the two reference frames different frequencies are measured. The frequency obtained in the geostationary frame is called the observed or ground-based frequency ω and the frequency measured in the moving frame is called the intrinsic frequency $\hat{\omega}$. The transport by the mean wind induces a shift in frequency called a Doppler shift $\mathbf{k} \cdot \mathbf{u}$. This is the frequency due to wave crest advection. The Doppler shift is also familiar from everyday life. An observer waiting on the street will hear an approaching ambulance siren with a higher pitched sound than if the ambulance is stopped. This instantly reverses to a lower pitched sound after the passage of the ambulance. Also, no changes in the static stability of the background atmosphere were treated in the previous section which may alter the restoring force.

If the intrinsic frequency is denoted by $\hat{\omega}$ then the equation

$$\omega = \hat{\omega} + \mathbf{k} \cdot \mathbf{u} \quad (4.18)$$

holds. The dispersion relation (4.9) is a relation between wave quantities in a moving frame. When constant background motion is included then

$$\hat{\omega}^2 = (\omega - \mathbf{k} \cdot \mathbf{u})^2 = \frac{N^2 k_h^2 + f^2 m^2}{k_h^2 + m^2} \quad (4.19)$$

relates the (constant) mean wind \mathbf{u} and (constant) stability frequency N to the intrinsic GW frequency $\hat{\omega}$.³ With the horizontal phase speed $c_h = \omega/k_h$ and the horizontal background wind $u_h = \mathbf{k} \cdot \mathbf{u}/k_h$ projected onto the wave direction \mathbf{k}_h/k_h , the medium-frequency approximation of m in eq. (4.14) simplifies to

$$m = \frac{N}{u_h - c_h}, \quad (4.20)$$

where the root with $c_z = \omega/m < 0$ was chosen. For the medium-frequency approximation, the two limits $m \rightarrow 0$ for $N \rightarrow 0$ and $|m| \rightarrow \infty$ for $u_h \rightarrow c_h$ are possible. The first refers to wave reflection whereas the second is called critical level filtering at the critical level where u_h approaches c_h . However, as noted before, the medium-frequency dispersion relation (4.20) is not appropriate for describing either one of these limiting cases.

The impact of wind and stability variations on the GW dispersion may be quantified using the HAMMONIA data introduced in section 3.1. As in section 4.2, a constant temperature $T_r = 240$ K and zero wind $\mathbf{u}_r = 0$ m/s are assumed. The resulting reference buoyancy frequency is $N_r = \sqrt{g^2/(c_p T_r)} = 72$ cycles per hour. The variations of the buoyancy frequency ΔN and zonal background wind Δu with respect to these reference values are shown in fig. 4.2. The range between the minimum and maximum time-mean

² The mean wind is either the Eulerian-mean or the Lagrangian-mean wind depending on the way it is obtained (as discussed in section 2.2.1). In the simulations in chapter 5, \mathbf{u} is taken from the climate model HAMMONIA and is more closely connected to the Eulerian-mean. For simplicity of notation the filter operator is dropped in the following discussions.

³ Concerning the sign convention: From eq. (4.19) it becomes clear that $\hat{\omega} > 0$ is only possible for $\omega > \mathbf{k} \cdot \mathbf{u}$. In the case $\omega > \mathbf{k} \cdot \mathbf{u}$ and $\hat{\omega} < 0$, all directions given in tab. 4.1 are reversed.

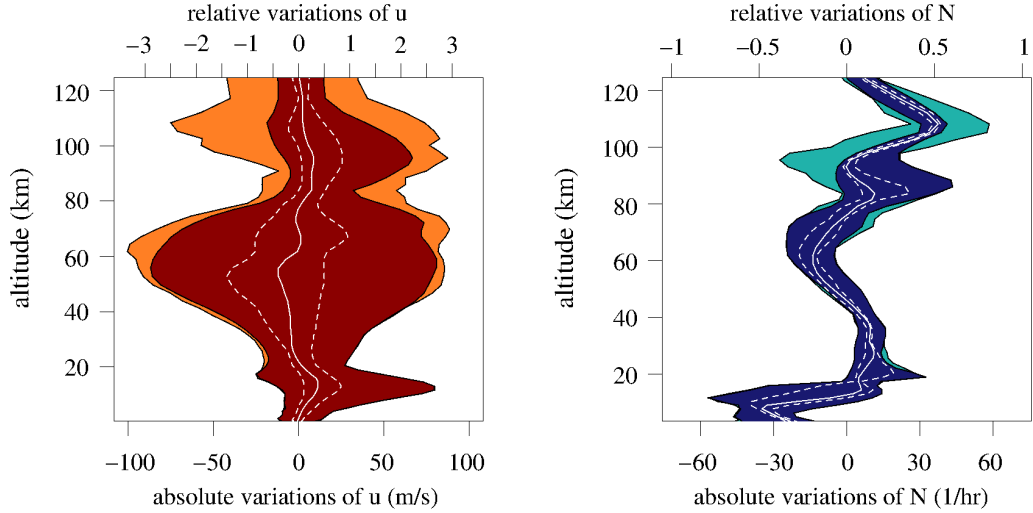


Figure 4.2.: Variations of zonal wind Δu (left) and buoyancy frequency ΔN (right) as functions of altitude. The range between minimum and maximum difference between the actual and the reference value ($u_r = 0$, $N_r = \sqrt{g^2/(c_p T_r)}$) is indicated by colors, dark red and dark blue for temporally averaged HAMMONIA January and orange and light blue with diurnal variations also included. The solid white line indicates the median and the dashed white lines the 25th and 75th percentiles of the darker range. The scaling of the relative variations was made with $c_0 = 30$ m/s and $N_r = 72$ cycles per hour.

wind achieved globally at each altitude level is plotted with dark red shading. In the tropopause region, the west-wind jets reach up to 80 m/s. In the stratopause region, winter westerly and summer easterly winds are maximized. A third maximum is visible at about 95 km and is connected to the wind reversal in the summer mesopause region. Altogether, the wind variations are quite large. The variations in buoyancy frequency, plotted with dark blue shading, are much narrower at each altitude and vertical changes dominate the overall distribution. Hence, a mean vertical profile of buoyancy frequency might capture most of the buoyancy structure. This confirms the assumptions made in the anelastic expansion derived in section 2.1.4. The use of a mean reference profile for the buoyancy frequency seems to be sufficient for this study.

Using lighter colors, the impact of the diurnal tides on the range of wind and stability variations is also shown in fig. 4.2. In the upper mesosphere / lower thermosphere, the diurnal wind variations become significant and extend the range between the extreme values up to 50 m/s. The diurnal variations in the buoyancy frequency are also shown and are mainly visible between 80 km and 120 km. As discussed next, they might be negligible compared to the diurnal wind variations.

Consider the medium-frequency dispersion relation (4.20). If the horizontal phase velocity $c_h = c_0$ is kept constant, then changes in u and N are uniquely attributable to changes in m , i.e.

$$\Delta m = \frac{\partial m}{\partial N} \Delta N + \frac{\partial m}{\partial u_h} \Delta u_h \quad (4.21)$$

which leads via eq. (4.20) to

$$\frac{\Delta m}{m} = \frac{\Delta N}{N} + \frac{\Delta u_h}{c_h - u_h}. \quad (4.22)$$

Therefore, the absolute stability variation ΔN is scaled by the full value N , but the absolute wind variation Δu_h is only scaled by the difference $c_h - u_h$ which can be significantly smaller than u_h alone. With respect to the reference values, the relative variations are approximately $\Delta N/N_r$ and $\Delta u/c_0$ with $c_0 = 30$ m/s and $N_r = 72$ cycles per hour, also shown in fig. 4.2. The relative variations in the zonal wind can reach up to 300 %, whereas relative variations in the buoyancy frequency are no larger than 30 % when an appropriate altitude profile is chosen. One should keep in mind that this oversimplified picture changes as time-dependence and horizontal inhomogeneities of the background are taken into account. Nevertheless, it demonstrates that GW fields are mainly influenced by the background wind due to Doppler shifting $\mathbf{k} \cdot \mathbf{u}$.

4.3.2. Multiple-scale asymptotics

Motivation

In the previous sections, linear GW solutions within a simplified setup, a resting, isothermal basic state, have been derived. Even in the last discussion of possible mean wind effects, a constant background wind was assumed and only the sensitivity of the dispersion relation to variations in the constant background quantities was investigated. This is quite unrealistic because winds and temperatures vary as function of space and time. However, no simple linear GW solution exists in the general case and the full set of equations has to be solved. Even where this might be feasible, the gain in information about the GW motion and its interpretation is quite restricted for the full solution.

Fortunately, variations of the basic state often happen on temporal and spatial scales different from those of the GW motion. This two-scale behavior, i.e. large scales for the background and small scales for gravity waves, will be used in the following to derive an asymptotic set of equations which is simpler to solve and interpret than the full equation set and its solution. Furthermore for GW parameterizations in climate models where fast GW variations are beyond the resolution threshold, an asymptotic description of GW impacts is needed.

In the following, multiple-scale asymptotics are presented with a focus on small-scale gravity waves. Similar derivations for rotating flows were provided in a number of studies. *Grimshaw* [1975a] used the square of the so-called Froude number as a small parameter in which he expanded the set of compressible equations (also assuming a small Mach number). In his derivation he additionally assumed that the wave Rossby number is order one which is equivalent to a buoyancy frequency N on the order of the Earth's angular frequency Ω . However, this assumption is not realistic for atmospheric conditions. *Shaw and Shepherd* [2009] provided a very general asymptotic framework for subgrid-scale motion, but they did not consider the scale separation between the vertical scale of the waves and the basic state resulting in a set of equations equivalent to the medium-frequency GW regime discussed below. They were not able to consistently include low- or high-frequency limits within their derivation. These issues will be addressed in the following.

Typical scales of the background state

The horizontal scales of the background range from several thousand kilometers up to the circumference of the earth. Within this range, a representative horizontal length scale \hat{L} is chosen. As given in tab. 4.2, $\hat{L} \approx 10\,000$ km is a typical value for the zonal wavelength of diurnal tides and for (quasi)-stationary planetary waves. However, no distinction between zonal and meridional scales is made here, even though the meridional

structure of the mean wind jet and the diurnal tides might be narrower. The vertical scale of the basic state \hat{H} is associated with the density-scale height of the atmosphere. Hence, $\hat{H} \approx 10$ km (as listed in tab. 4.2) is an appropriate choice. The aspect ratio $\hat{a}_B = \hat{H}/\hat{L}$ of the background flow is about 10^{-3} . This is closely related to the ratio of f to N in mid-latitudes where the Rossby deformation radius $R_D = N\hat{H}/f$ is of the order of \hat{L} [Pedlosky, 1982]. The background time scale \hat{T} ranges from one day for diurnal tides up to several weeks for (quasi-)stationary planetary waves. $\hat{T} \approx 1$ day is a reasonable choice for the investigation of tidal-wave effects.

quantity	symbol	magnitude
background time scale	\hat{T}	$O(1 \text{ day})$
background horizontal scale	\hat{L}	$O(10\,000 \text{ km})$
background vertical scale	\hat{H}	$O(10 \text{ km})$
aspect ratio	$\hat{a}_B = \hat{H}/\hat{L}$	$O(10^{-3})$

Table 4.2.: List of typical background scales.

The average zonal wind has typical magnitudes in the order of several 10 m/s. For diurnal tides, these values are only reached in the upper mesosphere / lower thermosphere. The horizontal phase velocity of the diurnal tides scales like $\Omega \hat{L}$ and is about one order of magnitude larger than \hat{U} . Nevertheless, the zonal wind scale \hat{U} is connected to the tidal phase velocity via $\hat{U} = \text{Sr}_m^{-1} \Omega \hat{L}$ where the mean Strouhal number $\text{Sr}_m = \hat{L} \Omega / \hat{U}$ measures the ratio between advective and tidal time-scale [Klein, 2000] and is typically large, $\text{Sr}_m = O(10)$. As discussed in chapter 3, the meridional tidal wind has approximately the same magnitude as the zonal wind. Thus, choosing one horizontal wind-scale \hat{U} seems to be sufficient for diurnal tides in the MLT region. The ratio between Earth's rotation time-scale Ω^{-1} and the background time-scale \hat{T} in the Coriolis force defines the mean Rossby number $\text{Ro}_m = (\Omega \hat{T})^{-1}$. For diurnal tides Ro_m is approximately one. The time-mean flow including planetary waves is close to geostrophic equilibrium [Holton, 2004] in which the horizontal pressure gradient is mostly balanced by the Coriolis force. Hence the corresponding Rossby number is much smaller than one. The magnitude of the density-scaled mean pressure ϕ is denoted by \hat{P} . For tides, pressure changes are governed by changes in the inertia and Coriolis forces, thus, $\hat{P} = \text{Sr} \hat{U}^2$ is appropriate. For the geostrophic components of the background flow, the balance between pressure gradient and Coriolis force implies the relation $\hat{P} = \Omega \hat{L} \hat{U}$ which also reduces to $\hat{P} = \text{Sr} \hat{U}^2$. The hydrostatic balance of the basic state enforces a relation between the mean buoyancy scale \hat{B} and the mean pressure scale \hat{P} , i.e. $\hat{B} = \hat{P}/\hat{H}$. The mean vertical velocity is in the order of several cm/s. The ratio Ω/\hat{N} is in the order of 10^{-3} where \hat{N} is a typical scale of the buoyancy frequency of about 100 cycles per hour, thus $\Omega^2/\hat{N}^2 \sim \hat{a}_B^2$. The advective change of the mean buoyancy field scales like $\hat{B}/\hat{T} = \hat{W}\hat{N}^2$. Diabatic effects are assumed to be weaker. The scale of the vertical velocity \hat{W} is then

$$\hat{W} = \frac{\hat{B}}{\hat{N}^2 \hat{T}} = \frac{\hat{P}}{\hat{H} \hat{N}^2 \hat{T}} = \frac{\Omega \hat{L} \hat{U}}{\hat{N}^2 \hat{H} \hat{T}} = \frac{\Omega^2}{\hat{N}^2} \frac{\hat{U}}{\hat{a}_B \Omega \hat{T}} = \hat{a}_B \text{Ro}_m \hat{U}. \quad (4.23)$$

The typical background scales are summarized in tab. 4.3.

quantity	relation
background horizontal wind	$\hat{\mathbf{U}} = \text{Sr}_m^{-1} \hat{\mathbf{L}} \Omega$
background density-scaled pressure	$\hat{\mathbf{P}} = \text{Sr}_m \hat{\mathbf{U}}^2$
background buoyancy	$\hat{\mathbf{B}} = \text{Sr}_m \hat{\mathbf{U}}^2 / \hat{\mathbf{H}}$
background vertical wind	$\hat{\mathbf{W}} = \hat{\mathbf{a}}_B \text{Ro}_m \hat{\mathbf{U}}$
mean Strouhal number	$\text{Sr}_m = \hat{\mathbf{L}} \Omega / \hat{\mathbf{U}}$
mean Rossby number	$\text{Ro}_m = (\Omega \hat{\mathbf{T}})^{-1}$

Table 4.3.: Relations between the different background scales, variables and mean flow numbers.

Typical wave scales

There is a large variety of possible horizontal GW scales. First, they depend on the geometry of the source and, second, on the interaction of the excited GW packet with the local flow. Hence the choice of the typical GW scale is somewhat arbitrary and is not meant to exclusively restrict the analysis to one very specific problem. Nevertheless, two possible choices for the horizontal GW scale $\hat{\mathbf{l}}$ are introduced: $\hat{\mathbf{l}} = O(100 \text{ km})$ for medium-frequency waves and $\hat{\mathbf{l}} = O(1 \text{ km})$ for non-hydrostatic waves.⁴ A vertical GW scale of $\hat{\mathbf{h}} = O(1 \text{ km})$ is specified, even though the local vertical GW scale might change a lot due to the impact of the mean wind shears. The aspect ratio $\hat{\mathbf{a}} = \hat{\mathbf{h}}/\hat{\mathbf{l}}$ becomes $\approx 10^{-2}$ and ≈ 1 for the medium-frequency and non-hydrostatic GW regimes, respectively. For fixed GW length-scales, the wave time-scale can be determined using the dispersion relation (4.9) and its approximations (4.13) and (4.14). For medium-frequency gravity waves, the typical time-scale is $\hat{\mathbf{t}} = (\hat{\mathbf{a}}\hat{\mathbf{N}})^{-1}$ and in the order of hours, and for non-hydrostatic gravity waves, $\hat{\mathbf{t}} = \hat{\mathbf{N}}^{-1}$ is in the order of minutes. The dominant scales for the medium-frequency regime are listed in tab. 4.4.

quantity	symbol	magnitude
wave time scale	$\hat{\mathbf{t}}$	$O(1 \text{ hr})$
wave horizontal scale	$\hat{\mathbf{l}}$	$O(100 \text{ km})$
wave vertical scale	$\hat{\mathbf{h}}$	$O(1 \text{ km})$
aspect ratio	$\hat{\mathbf{a}} = \hat{\mathbf{h}}/\hat{\mathbf{l}}$	$O(10^{-2})$
two-scale ratio	$\varepsilon = \hat{\mathbf{h}}/\hat{\mathbf{H}}$	$O(10^{-1})$

Table 4.4.: List of typical GW scales for the medium-frequency regime.

The typical wave and background scales were chosen in such a way that the combined flow system exhibits two-scale behavior. The ratio between the vertical wave scale $\hat{\mathbf{h}}$ and the vertical background scale $\hat{\mathbf{H}}$ is a small number in the order of 10^{-1} and referred

⁴ Clearly, whether a GW field is called medium-frequency or non-hydrostatic depends on the size of the horizontal *and* the vertical scale.

to as the two-scale ratio

$$\boxed{\varepsilon \equiv \varepsilon_z = \frac{\hat{h}}{\hat{H}} \quad : \text{small two-scale ratio.}} \quad (4.24)$$

Just as different wave aspect-ratios were assumed for the different GW regimes, the other scale-ratios, i.e. for horizontal and temporal scales, depend on the regime considered and are not necessarily equal to ε . For instance, the ratio between the horizontal scale of the wave \hat{l} and the background \hat{L} denoted by ε_x behaves like

$$\varepsilon_x = \frac{\hat{l}}{\hat{L}} = \frac{\hat{a}_B}{\hat{a}} \varepsilon. \quad (4.25)$$

For the medium-frequency and for the non-hydrostatic regime, ε_x takes values of $\approx 10^{-2}$ and $\approx 10^{-4}$, respectively. The ratio of time scales is

$$\varepsilon_t = \frac{\hat{t}}{\hat{T}} = \frac{\text{Ro}_m}{\text{Ro}_w}, \quad (4.26)$$

where $\text{Ro}_w = (\Omega\hat{t})^{-1}$ is the wave Rossby number. Ro_w gives a typical ratio between inertial and Coriolis forces and is typically large. For medium-frequency gravity waves, $\text{Ro}_w = O(10)$, whereas for non-hydrostatic gravity waves, $\text{Ro}_w = O(10^3)$.

In section 4.2, it was discussed that GW perturbations grow exponentially with the inverse square root of background density. With the ansatz (4.6), this effect was removed from the GW perturbations. The remaining scales of horizontal wind \hat{u} , vertical wind \hat{w} , density-scaled pressure \hat{p} and buoyancy \hat{b} are considered. In the following, the focus is on small-amplitude gravity waves which obey

$$\hat{u} = \varepsilon \hat{U}. \quad (4.27)$$

This connects the scale of the background wind and the GW wind and imposes an additional constraint on the various scale ratios. For $\Omega/\hat{N} = \hat{a}_B$, $\hat{U} = \Omega\hat{L}/\text{Sr}_m$ and $\hat{t} = (\hat{a}\hat{N})^{-1}$, the GW wind \hat{u} can be reformulated as

$$\hat{u} = \varepsilon \frac{\Omega\hat{L}}{\text{Sr}_m} = \varepsilon \frac{\hat{N}\hat{H}}{\text{Sr}_m} = \frac{\hat{N}\hat{h}}{\text{Sr}_m} = \text{Sr}_m^{-1} \frac{\hat{l}}{\hat{t}}. \quad (4.28)$$

As Sr_m^{-1} is a small number $\sim \varepsilon$, the advective time-scale \hat{l}/\hat{u} is about one order of magnitude larger than \hat{t} . The horizontal pressure gradient \hat{p}/\hat{l} is dominated by a temporal change in \hat{u} , thus

$$\hat{p} = \text{Sr}_m^{-1} \left(\frac{\hat{l}}{\hat{t}} \right)^2 = \text{Sr}_m^{-1} (\hat{N}\hat{h})^2. \quad (4.29)$$

For the vertical momentum balance, it is assumed that the local change in w' is in the order of b' or $\partial_z p'$ or smaller. Hence, the pressure gradient \hat{p}/\hat{h} can be used to estimate the buoyancy scale \hat{b} :

$$\hat{b} = \text{Sr}_m^{-1} \hat{N}^2 \hat{h}. \quad (4.30)$$

The last expression is connected to the upper threshold of \hat{b} given by convective instability, where surfaces of equal entropy or equivalently potential temperature overturn. Since, $\partial_z(\theta_r + \theta') = 0$ is the marginal convective stability condition, the upper bound

$$\hat{b}_c = \hat{h}\hat{N}^2 \quad (4.31)$$

results [Achatz *et al.*, 2010]. The investigated wave regime with $\hat{\mathbf{b}} = \text{Sr}_m^{-1} \hat{\mathbf{b}}_c$ is approximately linear and sufficiently far below the convective instability threshold. Finally for the vertical wind-scale,

$$\hat{w} = \frac{\hat{\mathbf{b}}}{\hat{N}^2 \hat{t}} = \text{Sr}_m^{-1} \frac{\hat{\mathbf{h}}}{\hat{t}} \quad (4.32)$$

is obtained because buoyancy perturbations are the result of the lifting of air packets. The last relation also shows that the vertical displacement of an air parcel is about one order of magnitude smaller than the vertical wavelength of small-amplitude waves.

The wave Rossby number can be written as $\text{Ro}_w = \hat{\mathbf{a}}/\hat{\mathbf{a}}_B$ and the associated constraint on the horizontal scale-ratio is $\varepsilon_x = \varepsilon \text{Ro}_w^{-1}$. Additionally, the background buoyancy-scale $\hat{\mathbf{B}}$ is then related to the wave buoyancy-scale $\hat{\mathbf{b}}$, i.e.

$$\frac{\hat{\mathbf{B}}}{\hat{\mathbf{b}}} = \frac{\hat{\mathbf{P}}}{\hat{\mathbf{H}}} \frac{\hat{\mathbf{h}}}{\hat{\mathbf{p}}} = \varepsilon \frac{\Omega \hat{\mathbf{L}} \hat{\mathbf{U}}}{\text{Sr}_m \hat{\mathbf{u}}^2} = \varepsilon^{-1}. \quad (4.33)$$

For consistency, the above requirements should be met within a systematic scaling analysis.

The linear anelastic equations

Based on the anelastic equations (2.36)- (2.38), the dynamics of small-amplitude gravity waves is derived. Quasi-stationary turbulence is assumed and as discussed before, additional turbulent stresses and heat fluxes are added to the molecular contributions. For the description of gravity waves, two different routes were suggested in section 2.2.3. On the one hand, gravity waves can be described as perturbations from the Eulerian-mean. On the other hand, viewing them as disturbances from the Lagrangian-mean might be convenient, too. As in the following small-amplitude waves within a scale-separated flow are considered, both strategies yield the same set of equations. It is therefore sufficient to focus only on one, in this case the Eulerian, description of wave perturbations. GW perturbations are marked by a prime, e.g. \mathbf{u}' , and the notation of the filter operator is dropped for the background flow, e.g. $\langle \mathbf{u} \rangle_w \rightarrow \mathbf{u}$. In the limit of small-amplitude perturbations, the linear GW dynamics is governed by

$$\nabla \cdot (\rho_r \mathbf{v}') = 0, \quad (4.34)$$

$$D_t \mathbf{v}' + 2\boldsymbol{\Omega} \times \mathbf{v}' = -\mathbf{v}' \cdot \nabla \mathbf{u} - \nabla \phi' + b' \mathbf{e}_z + \frac{1}{\rho_r} \nabla \cdot (\mathbf{F}' + \mathbf{R}'_t), \quad (4.35)$$

$$D_t b' + w' N^2 = -\mathbf{v}' \cdot \nabla b - \frac{wb' + w'b}{H_\theta} + \frac{gQ'_{an}}{c_p T_r}, \quad (4.36)$$

where $H_\theta = (\partial_z \ln \theta_r)^{-1}$ is the reference potential temperature scale height.⁵ The background flow mainly affects the wave motion due to the advection by the horizontal wind \mathbf{u} given in the operator $D_t = \partial_t + \mathbf{u} \cdot \nabla_h + w \partial_z$ and changes in the reference profile of buoyancy frequency denoted by N . In the investigations in chapter 5, \mathbf{u} is composed of a climatologically averaged field and temporally varying diurnal tides. In contrast to simple GW studies in vertical shear flows, horizontal and temporal variations in the background medium are also important. The gradients of the background medium change the momentum and heat content of the waves via the terms $\mathbf{v}' \cdot \nabla \mathbf{u}$ and $\mathbf{v}' \cdot \nabla b$.

⁵ The second term on the right-hand-side of eq. (4.36) usually does not appear in the Boussinesq equations, and the term $w'b/H_\theta$ even survives the scale asymptotics performed in section 4.3.2.

The first influences the kinetic energy of the waves and gives rise to the so-called shear production term and the second influences the available potential energy of the waves and is connected to a buoyancy production.⁶

The impact of turbulence is parametrized with a very simple diffusion approach. If in contrast to *Becker* [2001] plane-parallel flow is assumed, the molecular and turbulent stresses might be approximated by

$$\mathbf{F}' + \mathbf{R}'_t = \rho_r (\nu + \nu_t) (\nabla \mathbf{v}' + (\nabla \mathbf{v}')^T), \quad (4.37)$$

where the molecular and turbulent diffusion coefficients, ν and ν_t , have been introduced. The resulting frictional force is

$$\mathbf{f}'_R = (\nu + \nu_t) \{ \nabla^2 \mathbf{v}' + \nabla(\nabla \cdot \mathbf{v}') \} + \frac{\partial_z(\rho_r(\nu + \nu_t))}{\rho_r} (\partial_z \mathbf{v}' + \nabla w'). \quad (4.38)$$

Additionally, the linearized heat source

$$Q'_{an} = \epsilon' + Q' - \frac{1}{\rho_r} \nabla \cdot (\mathbf{q}' + \mathbf{J}'_t) \quad (4.39)$$

and the linearized dissipation rate

$$\epsilon' = \frac{1}{\rho_r} (\mathbf{F}' + \mathbf{R}'_t) \cdot \nabla \mathbf{u} + \frac{1}{\rho_r} (\mathbf{F} + \mathbf{R}_t) \cdot \nabla \mathbf{v}' \quad (4.40)$$

are obtained. Although the linearized dissipation rate also contains gradients of the large-scale flow, it is suggested that they play only a minor role in changing the heat content of a GW field. The molecular and turbulent fluxes of entropy are more important and are approximated by

$$-\frac{g}{c_p T_r \rho_r} \nabla \cdot (\mathbf{q}' + \mathbf{J}'_t) = \frac{1}{\rho_r} \nabla \cdot (\rho_r (\kappa + \kappa_t) \nabla b') = \frac{1}{\rho_r} \nabla \cdot \left(\frac{\rho_r}{\text{Pr}} (\nu + \nu_t) \nabla b' \right), \quad (4.41)$$

with molecular and turbulent diffusivities, κ and κ_t , respectively. In the second step, the Prandtl number Pr was introduced and no distinction between molecular and turbulent Prandtl number was made. Next, the dimensions are removed from the linear anelastic equations.

Non-dimensional equations, gravity-wave regimes and distinguished limits

In the system under consideration two dominant time scales \hat{t} and \hat{T} , two horizontal scales \hat{l} and \hat{L} , and two vertical scales \hat{h} and \hat{H} exist. For the linear GW dynamics, the set of short scales $\{\hat{l}, \hat{h}, \hat{t}\}$ are chosen as dimensions of the coordinates $\{\mathbf{x}, t\}$, so that the non-dimensional coordinates $\{\mathbf{x}^*, t^*\}$ obey

$$\{x, y\} = \hat{l} \{x^*, y^*\} \quad z = \hat{h} z^* \quad \text{and} \quad t = \hat{t} t^*. \quad (4.42)$$

Dependence on the long scales is denoted by $\{\varepsilon_x \mathbf{x}_h^*, \varepsilon z^*, \varepsilon_t t^*\}$. The multiple scaling property of the perturbation variables is given by e.g. $u^*(\mathbf{x}^*, t^*; \varepsilon_x \mathbf{x}_h^*, \varepsilon z^*, \varepsilon_t t^*)$.

In appendix A.2, a set of non-dimensional equations of motion is derived. Besides the small two-scale ratio $\varepsilon = \hat{h}/\hat{H}$,

- the mean Strouhal number $\text{Sr}_m = \hat{L}\Omega/\hat{U}$,

⁶ The buoyancy production in the balance of the available potential energy should not be confused with the vertical wave buoyancy flux $\langle b'w' \rangle_w$.

- the aspect ratios of the wave, $\hat{\mathbf{a}} = \hat{\mathbf{h}}/\hat{\mathbf{l}}$, and the background, $\hat{\mathbf{a}}_B = \hat{\mathbf{H}}/\hat{\mathbf{L}}$,
- the Rossby numbers of the wave, $\text{Ro}_w = (\Omega\hat{\mathbf{t}})^{-1}$, and the background, $\text{Ro}_m = (\Omega\hat{\mathbf{T}})^{-1}$,
- and the wave Reynolds number⁷ $\text{Re}_w = \hat{\mathbf{h}}^2/(\hat{\mathbf{v}}\hat{\mathbf{t}})$

also appear there. Furthermore, the ratios between temporal and horizontal scales, i.e. ε_t and ε_x , are present. The set of parameters can be used to construct asymptotic relations in the limit of certain parameters being small (or large). However, as discussed by *Klein* [2008], a singular limit might be obtained when one single parameter is brought to zero (or infinity) while keeping the others constant. If in the multi-parameter space the limit of several parameters is investigated then the order of the individual limits matters. Under the topic “*distinguished limits*” several parameters are “glued” together and the explicit path in the multi-parameter space is specified [e.g. *Klein*, 2000; *Klein*, 2008; *Achatz et al.*, 2010].

This is performed in the following for the linear anelastic set of equations. All parameters are related to the two-scale ratio ε as shown in tab. 4.5, but with different scaling assumptions for different regimes of GW motion. Therefore, an asymptotic expansion of non-dimensional equations is done with only one single parameter ε .

regime	independent			dependent		
	$\hat{\mathbf{a}}$	Ro_m	Re_w	Ro_w	ε_x	ε_t
inviscid, non-hydrostatic	1	1	ε^{-1}	ε^{-3}	ε^4	ε^3
inviscid, medium-frequency	ε^2	1	ε^{-1}	ε^{-1}	ε^2	ε
inviscid, inertia-gravity wave	ε^3	ε	ε^{-1}	1	ε	ε

Table 4.5.: List of scaling assumptions made for the different GW regimes. The background aspect ratio $\hat{\mathbf{a}}_B = \varepsilon^3$ and mean Strouhal number $\text{Sr}_m = \varepsilon^{-1}$ are set. Then different choices for the aspect ratio $\hat{\mathbf{a}}$, the background Rossby number Ro_m and the wave Reynolds number Re_w as functions of the two-scale ratio ε determine the wave Rossby number Ro_w and the ratios of temporal and horizontal scales, ε_t and ε_x .

The non-hydrostatic regime: Flow situations are investigated in which the clear separation between horizontal and vertical wave scales vanishes. The relevant scalings are listed in tab. 4.5. For aspect ratios in the order of $\hat{\mathbf{a}} \sim 1$, the time scale of the GW perturbation is given by $\hat{\mathbf{t}} \sim \hat{\mathbf{N}}^{-1}$, i.e. in the order of minutes. Hence, the wave Rossby number is large, $\text{Ro}_w \sim \varepsilon^{-3}$, and the separation between temporal and horizontal scales is large. The GW structure is nearly unaffected by Earth’s rotation and horizontal gradients of the background flow. Furthermore, it is assumed that on their way upward, the GW fields are not perturbed by localized heating, hence the heating number He is $O(\varepsilon^{-2})$, and are weakly damped by friction, thus $\text{Re}_w \sim \varepsilon^{-1}$. With all these assumption, the non-hydrostatic GW motion is to $O(\varepsilon^2)$ governed by

$$\nabla_h^* \cdot \mathbf{u}^* + \frac{\partial w^*}{\partial z^*} = \varepsilon \left(\frac{w^*}{H_\rho^*} \right), \quad (4.43)$$

⁷ The wave Reynolds number Re_w is the typical ratio between inertial and viscous forces incorporating molecular and turbulent diffusion and is also usually large.

$$\frac{D\mathbf{v}^*}{DT^*} + \nabla^* \phi^* - b^* \mathbf{e}_z = \varepsilon \left(-w^* \frac{\partial \mathbf{U}^*}{\partial z^*} + (\nu^* + \nu_t^*) \nabla^* \cdot (\nabla^* \mathbf{v}^*) \right), \quad (4.44)$$

$$\frac{Db^*}{DT^*} + w^* (N^*)^2 = \varepsilon \left(-w^* \frac{\partial B^*}{\partial z^*} - \frac{w^* B^*}{H_\theta^*} + \frac{\nu^* + \nu_t^*}{\text{Pr}} \nabla^* \cdot (\nabla^* b^*) \right), \quad (4.45)$$

where all dimensionless quantities have an asterisk. The dimensionless horizontal background wind is denoted by \mathbf{U}^* and the dimensionless background buoyancy is B^* . D/DT^* is the advective derivative with the horizontal wind.

The inviscid, medium-frequency regime: The inviscid, medium-frequency regime is typically described by the wave scales given in tab. 4.4. The wave aspect ratio is small $\hat{a} \approx 10^{-2}$. The wave time-scale is in the order of hours. Thus the time-scale ratio is $\approx 10^{-1}$ and also approximately equal to Ro_w^{-1} . The horizontal scales of the wave and the background are still separated by a factor of $\approx 10^{-2}$. Again, $\text{Re}_w \sim \varepsilon^{-1}$ and $\text{He} \sim \varepsilon^{-2}$ are assumed. These relations are the basis of the scaling of the medium-frequency GW regime in tab. 4.5. Then the non-dimensional GW equations from appendix A.2 are to $O(\varepsilon^2)$ governed by

$$\nabla_h^* \cdot \mathbf{u}^* + \frac{\partial w^*}{\partial z^*} = \varepsilon \left(\frac{w^*}{H_\rho^*} \right), \quad (4.46)$$

$$\begin{aligned} \frac{D\mathbf{u}^*}{DT^*} + \nabla_h^* \phi^* = \varepsilon \left(-W^* \frac{\partial \mathbf{u}^*}{\partial z^*} - w^* \frac{\partial \mathbf{U}^*}{\partial z^*} \right. \\ \left. - f^* \mathbf{e}_z \times \mathbf{u}^* + (\nu^* + \nu_t^*) \frac{\partial^2 \mathbf{u}^*}{\partial z^* \partial z^*} \right), \end{aligned} \quad (4.47)$$

$$\frac{\partial \phi^*}{\partial z^*} - b^* = 0, \quad (4.48)$$

$$\frac{Db^*}{DT^*} + w^* (N^*)^2 = \varepsilon \left(-W^* \frac{\partial b^*}{\partial z^*} - w^* \frac{\partial B^*}{\partial z^*} - \frac{w^* B^*}{H_\theta^*} + \frac{\nu^* + \nu_t^*}{\text{Pr}} \frac{\partial^2 b^*}{\partial z^* \partial z^*} \right), \quad (4.49)$$

where $f^* = 2 \sin \varphi$ is the non-dimensional Coriolis parameter.

The inertia-gravity wave regime: In the limit of large wave time-scales, the wave structure becomes sensitive to the rotation of earth. In the inertia-gravity wave regime, the scale separation between the wave and background time-scales is questionable. Thus, if $\hat{t} \sim \Omega^{-1}$, then the background Rossby number should be small, i.e. $\text{Ro}_m = \varepsilon$, for scale-separated motion. These requirements are met for geostrophically balanced flows, but certainly not for the diurnal tides. Therefore, the scale asymptotics may fail in the limit of very slow inertia-gravity waves.

In the following, the inertia-gravity wave regime is constructed with the scaling assumptions from tab. 4.5 ignoring possible problems resulting from tidal background flows. To $O(\varepsilon^2)$

$$\nabla_h^* \cdot \mathbf{u}^* + \frac{\partial w^*}{\partial z^*} = \varepsilon \left(\frac{w^*}{H_\rho^*} \right), \quad (4.50)$$

$$\frac{D\mathbf{u}^*}{DT^*} + f^* \mathbf{e}_z \times \mathbf{u}^* + \nabla_h^* \phi^* = \varepsilon \left(-W^* \frac{\partial \mathbf{u}^*}{\partial z^*} - \mathbf{v}^* \cdot \nabla^* \mathbf{U}^* + (\nu^* + \nu_t^*) \frac{\partial^2 \mathbf{u}^*}{\partial z^* \partial z^*} \right), \quad (4.51)$$

$$\frac{\partial \phi^*}{\partial z^*} - b^* = 0, \quad (4.52)$$

$$\frac{Db^*}{DT^*} + w^*(N^*)^2 = \varepsilon \left(-W^* \frac{\partial b^*}{\partial z^*} - \mathbf{v}^* \cdot \nabla^* B^* - \frac{w^* B^*}{H_\theta^*} + \frac{\nu^* + \nu_t^*}{\text{Pr}} \frac{\partial^2 b^*}{\partial z^* \partial z^*} \right). \quad (4.53)$$

Hybrid dynamics

One could think to use each of the three regimes, the non-hydrostatic, the medium-frequency and the inertia-gravity wave, separately to describe the dynamics of their corresponding gravity waves. Beside the redundant work this would entail, an intelligent strategy would have to be invented to jump smoothly between the regimes if suddenly the conditions for the validity of one are violated. This seems to be unnecessarily complicated.

In the following, a hybridization of the different GW regimes is proposed. The main focus remains on the medium-frequency dynamics, but both limiting cases, the non-hydrostatic and the inertia-gravity wave limit, enter the flow description. The hybrid system is chosen to be

$$\nabla_h^* \cdot \mathbf{u}^* + \frac{\partial w^*}{\partial z^*} = \varepsilon \left(\frac{w^*}{H_\rho^*} \right), \quad (4.54)$$

$$\begin{aligned} \frac{D\mathbf{v}^*}{DT^*} + f^* \mathbf{e}_z \times \mathbf{u}^* + \nabla^* \phi^* - b^* \mathbf{e}_z = \varepsilon \left(-W^* \frac{\partial \mathbf{u}^*}{\partial z^*} - \mathbf{v}^* \cdot \nabla^* \mathbf{U}^* \right. \\ \left. + (\nu^* + \nu_t^*) \nabla^* \cdot (\nabla^* \mathbf{v}^*) \right), \quad (4.55) \end{aligned}$$

$$\frac{Db^*}{DT^*} + w^*(N^*)^2 = \varepsilon \left(-W^* \frac{\partial b^*}{\partial z^*} - \mathbf{v}^* \cdot \nabla^* B^* - \frac{w^* B^*}{H_\theta^*} + \frac{\nu^* + \nu_t^*}{\text{Pr}} \nabla^* \cdot (\nabla^* b^*) \right). \quad (4.56)$$

Unlike the medium-frequency dynamics, it also contains non-hydrostatic motion, the direct impact of Earth's rotation and the three-dimensional gradients of the background flow. Note, however, that GW motions in the high and low-frequency limit are inherently different, and therefore, besides the somewhat arbitrary hybridization, no systematic way exists to obtain this asymptotic set of equations for typical atmospheric conditions. If a study of GW motion e.g. *Marks and Eckermann* [1995] is able to describe non-hydrostatic and inertial effects within one single formulation, the discussed hybridization strategy must be implicitly hidden in the asymptotic derivation of the resulting system which has not been clarified by the corresponding authors.

4.3.3. Wentzel-Kramers-Brillouin theory for gravity waves

Motivation

Wentzel-Kramers-Brillouin (WKB) theory was first invented in the context of quantum mechanics. Semi-classical, approximate solutions of the Schrödinger equation were found which are assumed to be sufficiently described by locally monochromatic wave packets [*Messiah*, 1961] Since they are slow compared to variations of the wave phase, amplitude changes are ignored. Hence, WKB theory uses the inherent two-scale behavior of wave packets and easily translates to the atmospheric context.

As discussed in section 4.3.2, GW motion can happen on temporal and spatial scales different from those of the basic state. This two-scale behavior will be used in the

following to construct solutions of the asymptotic hybrid system (4.54)-(4.56). Locally monochromatic GW packets are investigated in which the short scales are associated with variations in the wave phase whereas the large-scale inhomogeneities of the basic state induce slow changes in other wave-packet properties such as the wave amplitudes or wavenumber vectors. Hence, the wave packet is a generic object which has the scaling properties of a realistic GW field.

The wave-packet ansatz

The hybrid set (4.54)-(4.56) is asymptotically solved with the wave-packet ansatz

$$\mathbf{v}^* = \Re \left\{ \hat{\mathbf{v}}^*(\mathbf{X}^*, T^*) e^{i\Theta(\mathbf{X}^*, T^*)/\varepsilon} \right\}, \quad (4.57)$$

$$\phi^* = \Re \left\{ \hat{\phi}^*(\mathbf{X}^*, T^*) e^{i\Theta(\mathbf{X}^*, T^*)/\varepsilon} \right\}, \quad (4.58)$$

$$b^* = \Re \left\{ \hat{b}^*(\mathbf{X}^*, T^*) e^{i\Theta(\mathbf{X}^*, T^*)/\varepsilon} \right\}. \quad (4.59)$$

The dimensionless wave amplitudes $\hat{\mathbf{v}}^*$, $\hat{\phi}^*$ and \hat{b}^* only depend on the slow variables $\mathbf{X}^* = \varepsilon \mathbf{x}^*$ and $T^* = \varepsilon t^*$ for which $\varepsilon_x = \varepsilon$ and $\varepsilon_t = \varepsilon$ was assumed within the hybridization. The GW phase Θ depends on the fast variables while its derivatives only show slow variations. This is emphasized by the additional ε^{-1} factor. The partial temporal and spatial derivatives of the wave packet, e.g. of velocity

$$\frac{\partial \mathbf{v}^*}{\partial t^*} = \Re \left\{ \left(i \frac{\partial \Theta}{\partial T^*} \hat{\mathbf{v}}^* + \varepsilon \frac{\partial \hat{\mathbf{v}}^*}{\partial T^*} \right) e^{i\Theta/\varepsilon} \right\}, \quad (4.60)$$

$$\nabla_{\mathbf{x}}^* \mathbf{v}^* = \Re \left\{ \left(i (\nabla_{\mathbf{x}}^* \Theta) \hat{\mathbf{v}}^* + \varepsilon \nabla_{\mathbf{x}}^* \hat{\mathbf{v}}^* \right) e^{i\Theta/\varepsilon} \right\} \quad (4.61)$$

are dominated to $O(\varepsilon^0)$ by the variations of the phase. Amplitude variations however appear in the next order ε^1 . $\nabla_{\mathbf{x}}^*$ is the nabla operator in the slow coordinates \mathbf{X}^* . Extending the definition of the GW phase in a resting basic state eq. (4.7), the dimensionless GW wavenumber vector \mathbf{k}^* and the dimensionless GW observed frequency ω^* are defined as [Hayes, 1970]

$$\mathbf{k}^* = \nabla_{\mathbf{x}}^* \Theta, \quad (4.62)$$

$$\omega^* = -\frac{\partial \Theta}{\partial T^*}. \quad (4.63)$$

The dimensionless intrinsic frequency measured in a frame moving with the background is given by

$$\hat{\omega}^* = -\frac{D\Theta}{DT^*} = \omega^* - \mathbf{k}^* \cdot \mathbf{U}^*. \quad (4.64)$$

As discussed in section 4.3.1, the main difference between $\hat{\omega}^*$ and ω^* is the Doppler shift by the horizontal background wind \mathbf{U}^* . In contrast to section 4.3.1, \mathbf{U}^* can vary slowly over the length and time-scales of a GW packet. Finally, the amplitude of the wave packet is expanded in orders of ε to obtain e.g. for the wind amplitude

$$\hat{\mathbf{v}}^* = \mathbf{v}^* + \varepsilon \mathbf{v}_1^* + \dots \quad (4.65)$$

Lowest order $O(\varepsilon^0)$: Dispersion and polarization relations

The wave-packet ansatz is inserted into the hybrid system (4.54)-(4.56) and the corresponding definitions of wavenumber vectors and frequencies are used. To the lowest order in ε , a homogeneous system of linear equation results:

$$\begin{pmatrix} -i\hat{\omega}^* & -f^* & 0 & ik^* & 0 \\ f^* & -i\hat{\omega}^* & 0 & il^* & 0 \\ 0 & 0 & -i\hat{\omega}^* & im^* & -N^* \\ ik^* & il^* & im^* & 0 & 0 \\ 0 & 0 & N^* & 0 & -i\hat{\omega}^* \end{pmatrix} \cdot \begin{pmatrix} u_0^* \\ v_0^* \\ w_0^* \\ \phi_0^* \\ b_0^*/N^* \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (4.66)$$

where k^* , l^* and m^* denote the dimensionless zonal, meridional and vertical wavenumber, respectively. First of all, the matrix above, in the following denoted as \mathbf{M} , is anti-Hermitian, i.e. the adjoint matrix \mathbf{M}^+ is equal to $-\mathbf{M}$. Second, the linear system only has trivial solutions unless \mathbf{M} is singular, i.e. unless the determinant of \mathbf{M} is zero. This connects the intrinsic frequency $\hat{\omega}^*$ to all other variables, i.e. to k^* , l^* , m^* , f^* and N^* . The resulting relation is called “*dispersion relation*” and is now obtained after a series of manipulations. The amplitudes of v_0^* and ϕ_0^* are rewritten as function of b_0^* . From thermodynamics,

$$w_0^* = \frac{i\hat{\omega}^*}{(N^*)^2} b_0^* \quad (4.67)$$

is obtained. The vertical momentum balance

$$i\hat{\omega}^* w_0^* + b_0^* = im^* \phi_0^* \quad (4.68)$$

is rewritten as

$$\phi_0^* = -\frac{i}{m^*} \frac{(N^*)^2 - (\hat{\omega}^*)^2}{(N^*)^2} b_0^*, \quad (4.69)$$

and the horizontal balance

$$\begin{pmatrix} -i\hat{\omega}^* & -f^* \\ f^* & -i\hat{\omega}^* \end{pmatrix} \cdot \begin{pmatrix} u_0^* \\ v_0^* \end{pmatrix} = -i \begin{pmatrix} k^* \\ l^* \end{pmatrix} \phi_0^*, \quad (4.70)$$

changes to

$$\begin{pmatrix} u_0^* \\ v_0^* \end{pmatrix} = \frac{(N^*)^2 - (\hat{\omega}^*)^2}{(\hat{\omega}^*)^2 - (f^*)^2} \begin{pmatrix} -i\hat{\omega}^* k^* + l^* f^* \\ -i\hat{\omega}^* l^* - k^* f^* \end{pmatrix} \frac{b_0^*}{m^* (N^*)^2}. \quad (4.71)$$

The relations between v_0^* , ϕ_0^* and b_0^* are called “*polarization relations*” and determine to lowest order the phase relations between the wind, pressure and buoyancy perturbations in a GW field. They can also be used to estimate the subgrid-scale stresses and heat fluxes due to gravity waves. The continuity equation is used to obtain the dispersion relation

$$(\hat{\omega}^*)^2 = (\omega^* - \mathbf{k}^* \cdot \mathbf{U}^*)^2 = \frac{(N^* k_h^*)^2 + (f^* m^*)^2}{(k_h^*)^2 + (m^*)^2} \quad (4.72)$$

The last relation is identical to eq. (4.19) from the linear GW solution in a resting basic state with the major exception that eq. (4.72) is an asymptotic result for small ε in which slow variations of background properties N^* , f^* and \mathbf{U}^* are allowed. Therefore, eq. (4.72) gives the natural extension to GW motion in general moving basic states.

First order $O(\varepsilon^1)$: Amplitude equations

In the lowest order of the asymptotic expansion, relations between the GW wind, pressure and buoyancy perturbations were obtained, but still one amplitude measure, say b_0^* , remained unknown. Now, in the next order, $O(\varepsilon^1)$, an equation for the evolution of the GW amplitude results. Inserting the wave-packet ansatz into the hybrid system (4.54)-(4.56) and collecting terms of order ε , the first order system

$$\begin{pmatrix} -i\hat{\omega}^* & -f^* & 0 & ik^* & 0 \\ f^* & -i\hat{\omega}^* & 0 & il^* & 0 \\ 0 & 0 & -i\hat{\omega}^* & im^* & -N^* \\ ik^* & il^* & im^* & 0 & 0 \\ 0 & 0 & N^* & 0 & -i\hat{\omega}^* \end{pmatrix} \cdot \begin{pmatrix} u_1^* \\ v_1^* \\ w_1^* \\ \phi_1^* \\ b_1^*/N^* \end{pmatrix} + \begin{pmatrix} D_{T^*}u_0^* + \partial_{X^*}\phi_0^* \\ D_{T^*}v_0^* + \partial_{Y^*}\phi_0^* \\ D_{T^*}w_0^* + \partial_{Z^*}\phi_0^* \\ \nabla_{X^*}^* \cdot \mathbf{v}_0^* \\ D_{T^*}(b_0^*/N^*) \end{pmatrix} \quad (4.73)$$

$$= \begin{pmatrix} -W^*(im)u_0^* - \mathbf{v}_0^* \cdot \nabla_{X^*}^* U^* - (\nu^* + \nu_t^*)|\mathbf{k}^*|^2 u_0^* \\ -W^*(im)v_0^* - \mathbf{v}_0^* \cdot \nabla_{X^*}^* V^* - (\nu^* + \nu_t^*)|\mathbf{k}^*|^2 v_0^* \\ -(\nu^* + \nu_t^*)|\mathbf{k}^*|^2 w_0^* \\ w_0^*/H_\rho^* \\ -W^*(im)b_0^*/N^* - \mathbf{v}_0^* \cdot \nabla_{X^*}^* B^*/N^* - w_0^* B^*/(N^* H_\theta^*) - (\nu^* + \nu_t^*)|\mathbf{k}^*|^2 b_0^*/(\text{Pr } N^*) \end{pmatrix}$$

arises. Eq. (4.73) looks quite complicated on the first sight, and one might wonder that the next order amplitude corrections v_1^* , ϕ_1^* and b_1^* appear. They are coupled to the matrix \mathbf{M} . However, as \mathbf{M} has a right eigenvector projecting onto its null space, it is also possible to find a left eigenvector doing the same, and thus eliminate this term by left multiplying (4.73) with the left eigenvector. As \mathbf{M} is anti-Hermitian, the left eigenvector projecting onto its null space is equal to the complex conjugate of the right eigenvector. Therefore,

$$\begin{pmatrix} (u_0^*)^* \\ (v_0^*)^* \\ (w_0^*)^* \\ (\phi_0^*)^* \\ (b_0^*)^*/N^* \end{pmatrix}^T \cdot \begin{pmatrix} D_{T^*}u_0^* + \partial_{X^*}\phi_0^* \\ D_{T^*}v_0^* + \partial_{Y^*}\phi_0^* \\ D_{T^*}w_0^* + \partial_{Z^*}\phi_0^* \\ \nabla_{X^*}^* \cdot \mathbf{v}_0^* \\ D_{T^*}(b_0^*/N^*) \end{pmatrix}$$

$$= \begin{pmatrix} (u_0^*)^* \\ (v_0^*)^* \\ (w_0^*)^* \\ (\phi_0^*)^* \\ (b_0^*)^*/N^* \end{pmatrix}^T \cdot \begin{pmatrix} -W^*(im)u_0^* - \mathbf{v}_0^* \cdot \nabla_{X^*}^* U^* \\ -W^*(im)v_0^* - \mathbf{v}_0^* \cdot \nabla_{X^*}^* V^* \\ 0 \\ w_0^*/H_\rho^* \\ -W^*(im)b_0^* - \mathbf{v}_0^* \cdot \nabla_{X^*}^* B^*/N^* - w_0^* B^*/(N^* H_\theta^*) \end{pmatrix}$$

$$- (\nu^* + \nu_t^*)|\mathbf{k}^*|^2 \begin{pmatrix} (u_0^*)^* \\ (v_0^*)^* \\ (w_0^*)^* \\ (\phi_0^*)^* \\ (b_0^*)^*/N^* \end{pmatrix}^T \cdot \begin{pmatrix} u_0^* \\ v_0^* \\ w_0^* \\ 0 \\ b_0^*/(\text{Pr } N^*) \end{pmatrix}, \quad (4.74)$$

where $()^*$ denotes the complex conjugate and $()^T$ is the transpose.⁸ This is a known strategy to obtain an evolution equation for wave energy, which here obeys

$$\begin{aligned} & \frac{D}{DT^*} \left(\frac{|v_0^*|^2}{2} + \frac{|b_0^*|^2}{2(N^*)^2} \right) + \nabla_x^* \cdot \left(\Re \{ v_0^* (\phi_0^*)^* \} \right) - \frac{\Re \{ w_0^* (\phi_0^*)^* \}}{H_\rho^*} \\ &= -\Re \{ u_0^* (v_0^*)^* \} \cdot \nabla_x^* U^* - \Re \{ b_0^* (v_0^*)^* \} \cdot \frac{\nabla_x^* B^*}{(N^*)^2} - \Re \{ b_0^* (w_0^*)^* \} \frac{B^*}{(N^*)^2 H_\rho^*} \\ & \quad + (\nu^* + \nu_t^*) |\mathbf{k}^*|^2 \left(\frac{|v_0^*|^2}{2} + \text{Pr}^{-1} \frac{|b_0^*|^2}{2(N^*)^2} \right), \quad (4.75) \end{aligned}$$

when the real part of eq. (4.74) has been taken. Using the polarization relations from the previous section, the evolution of the unknown amplitude measure is determined.

4.3.4. Gravity-wave propagation and the ray-tracing method

The dispersion relation (4.72) is a diagnostic relation between the GW frequency, GW wavenumbers and basic state properties. But as GW frequency and GW wavenumbers are calculated from one single GW phase function, it is also possible to construct prognostic relations for them. This enables one to explicitly follow the GW motion in time and monitor changes of relevant GW properties. Called “*ray tracing*”, it is a simplified view of the propagation of GW packets and eases the interpretation of changes in the GW fields, but it has only a restricted range of validity due to the assumption of scale-separated wave packets. The ray-tracing method will be developed and explained below, keeping in mind that it will be applied to GW propagation through middle-atmosphere thermal tides in chapter 5.

In the subsequent sections, dimensions are included in all variables, so the asterisks are removed, e.g. $\mathbf{k}^* \rightarrow \mathbf{k}$, the background wind changes like $U^* \rightarrow \mathbf{u}$ and lowest order GW amplitudes change like $v_0^* \rightarrow \hat{v}$. All results easily translate to dimensional quantities, but some important relations will be restated.

Ray tracing in a plane-parallel atmosphere

Ray tracing is based on the assumption that a locally monochromatic GW field obeys a dispersion relation, in our case

$$\boxed{\hat{\omega}^2 = (\omega - \mathbf{k} \cdot \mathbf{u})^2 = \frac{N^2 k_h^2 + f^2 m^2}{k_h^2 + m^2} \quad : \text{GW dispersion relation.}} \quad (4.76)$$

The horizontal background wind \mathbf{u} is a function of time and space and the (reference) buoyancy frequency N and the Coriolis parameter f only depend on altitude and latitude, respectively. The GW field is divided into small parts in which local values of GW frequency ω , GW wavenumber vector \mathbf{k} and an appropriate amplitude measure can be defined. Each part of the GW field is called a wave parcel and moves with its group velocity \mathbf{c}_g . The geometric position \mathbf{x} of the wave parcel is determined by its initial position and the solution of

$$d_t \mathbf{x} = \mathbf{c}_g \quad (4.77)$$

⁸ It is not distinguished between column and row vectors in most of this thesis.

where $d_t = \partial_t + \mathbf{c}_g \cdot \nabla$ is the derivative along the group ray $\mathbf{x}(t)$. The group velocity has been already introduced in section 4.2. It measures the speed with which the envelope of the wave packet propagates and has the components

$$c_{g\lambda} = \frac{\partial\omega}{\partial k} = u + \frac{k}{|\mathbf{k}|^2} \frac{N^2 - \hat{\omega}^2}{\hat{\omega}}, \quad (4.78)$$

$$c_{g\varphi} = \frac{\partial\omega}{\partial l} = v + \frac{l}{|\mathbf{k}|^2} \frac{N^2 - \hat{\omega}^2}{\hat{\omega}}, \quad (4.79)$$

$$c_{gz} = \frac{\partial\omega}{\partial m} = -\frac{m}{|\mathbf{k}|^2} \frac{\hat{\omega}^2 - f^2}{\hat{\omega}}. \quad (4.80)$$

As discussed in section 4.2, the sign of vertical group velocity c_{gz} is opposite to the sign of the ratio $m/\hat{\omega}$. Hence, if the positive root of (4.76) with $\hat{\omega} > 0$ is chosen, then $m < 0$ for upward moving gravity waves. The sign of the wavenumbers gives the direction of the GW phase progression, $k > 0$ ($k < 0$) for eastward (westward), $l > 0$ ($l < 0$) for northward (southward), and $m > 0$ ($m < 0$) for upward (downward) phase motion. Again, as in the simple example in section 4.2, phase motion of gravity waves is downward for upward-moving wave energy.

In the following, the dispersion relation (4.76) will be written in the abstract notation $\omega = \omega(\mathbf{k}, \mathbf{\Lambda})$, where the local properties of the background medium are summarized in the vector $\mathbf{\Lambda}$ [Bretherton and Garrett, 1968]. Additionally, as discussed before and given e.g. by Hayes [1970], the GW observed frequency ω and the GW wavenumber vector \mathbf{k} are defined by the local variations of the GW phase Θ ,

$\omega = -\partial_t \Theta$: GW observed frequency	(4.81)
$\mathbf{k} = \nabla \Theta$: GW wavenumber vector	(4.82)

The dispersion relation (4.76) gives a Hamilton-Jacobi equation for the GW phase Θ

$$-\partial_t \Theta = \omega(\nabla \Theta, \mathbf{\Lambda}) \quad (4.83)$$

and can be used to derive evolution equations for ω and \mathbf{k} . The local rate of change in observed frequency is

$$\frac{\partial\omega}{\partial t} = \frac{\partial\omega}{\partial k_i} \frac{\partial k_i}{\partial t} + \frac{\partial\omega}{\partial \Lambda_n} \frac{\partial \Lambda_n}{\partial t} \quad (4.84)$$

where $k_i = \mathbf{k} \cdot \mathbf{e}_i$ and i and n index coordinate directions and the number of background quantities, respectively.⁹ The conservation of wave crests [Pedlosky, 2003]

$$\partial_t \mathbf{k} = -\nabla \omega \quad (4.85)$$

follows from Schwarz's theorem, i.e. $\partial_t \nabla \Theta = \nabla \partial_t \Theta$. Therefore, using $c_{gi} = \partial\omega/\partial k_i$ and $d_t = \partial_t + \mathbf{c}_g \cdot \nabla$, a ray equation for the observed frequency is found:

$$d_t \omega = \frac{\partial\omega}{\partial \Lambda_n} \frac{\partial \Lambda_n}{\partial t}. \quad (4.86)$$

The temporal variability of the background medium induces changes of the GW observed frequency along the ray. For the GW wavenumber vector \mathbf{k} , the same procedure applies.

⁹Summation over repeated indices, i.e. $a_i b_i = \sum_{i=1}^I a_i b_i$, is implied where the number I depends on the context.

The gradient in eq. (4.85) is expressed in Cartesian coordinates, i.e. $\nabla = \mathbf{e}_i \partial_{x_i}$ where x_i is one of x , y and z . The corresponding tendencies of the wavenumbers k_j are

$$\frac{\partial k_j}{\partial t} = -\frac{\partial \omega}{\partial k_i} \frac{\partial k_i}{\partial x_j} - \frac{\partial \omega}{\partial \Lambda_n} \frac{\partial \Lambda_n}{\partial x_j}. \quad (4.87)$$

Fortunately, the tensor $\partial_{x_j} k_i$ is symmetric, due to $\partial_{x_j} \partial_{x_i} \Theta = \partial_{x_i} \partial_{x_j} \Theta$, and thus a ray equation for the wavenumbers can be established:

$$d_t k_j = -\frac{\partial \omega}{\partial \Lambda_n} \frac{\partial \Lambda_n}{\partial x_j}. \quad (4.88)$$

The spatial inhomogeneities lead to changes of the GW wavenumbers along the ray. From the dispersion relation (4.76), the derivatives $\partial \omega / \partial \Lambda_n$ are

$$\frac{\partial \omega}{\partial u} = k, \quad \frac{\partial \omega}{\partial v} = l, \quad \frac{\partial \omega}{\partial f} = \frac{f m^2}{\hat{\omega} |\mathbf{k}|^2} \quad \text{and} \quad \frac{\partial \omega}{\partial N} = \frac{N k_h^2}{\hat{\omega} |\mathbf{k}|^2}. \quad (4.89)$$

The ray-tracing equations for a plane-parallel atmosphere are then

$$d_t \omega = \mathbf{k} \cdot \partial_t \mathbf{u}, \quad (4.90)$$

$$d_t k = -\mathbf{k} \cdot \partial_x \mathbf{u}, \quad (4.91)$$

$$d_t l = -\mathbf{k} \cdot \partial_y \mathbf{u} - \frac{f m^2}{\hat{\omega} |\mathbf{k}|^2} \partial_y f, \quad (4.92)$$

$$d_t m = -\mathbf{k} \cdot \partial_z \mathbf{u} - \frac{N k_h^2}{\hat{\omega} |\mathbf{k}|^2} \partial_z N. \quad (4.93)$$

The time-dependence of the background wind, in our case the effect of the diurnal tide, induces a modulation of GW observed frequency ω along the ray. The horizontal gradients in the background wind, the beta-effect and vertical changes in stability N lead to changes in the GW wavenumbers.

For a complete description of the wave field, the wave amplitude is needed. As shown in section 4.3.3, the wave energy equation arose in the first order of the WKB expansion. Including dimensions, eq. (4.75) translates to

$$\begin{aligned} & \partial_t E_w + \nabla \cdot \left(\mathbf{u} E_w + \frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{v}} \hat{\phi}^* \right\} \right) \\ &= -\frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{u}} \hat{\mathbf{v}}^* \right\} \cdot \nabla \mathbf{u} - \frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{b}} \hat{\mathbf{v}}^* \right\} \cdot \frac{\nabla b}{N^2} - \frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{b}} \hat{\omega}^* \right\} \frac{b}{N^2 H_\theta} - \tau^{-1} E_w, \end{aligned} \quad (4.94)$$

where the continuity equations (2.36) and (4.34) were used and

$$E_w = \rho_r \left(\frac{\langle |\mathbf{v}'|^2 \rangle_w}{2} + \frac{\langle b'^2 \rangle_w}{2 N^2} \right) = \rho_r \left(\frac{|\hat{\mathbf{v}}|^2}{4} + \frac{|\hat{\mathbf{b}}|^2}{4 N^2} \right) \quad (4.95)$$

is the wave energy. The damping rate

$$\tau^{-1} = 2(\nu + \nu_t) |\mathbf{k}|^2 (\mu + \text{Pr}^{-1} (1 - \mu)) \quad (4.96)$$

depends on the ratio $\mu = E_{kin}/E_w$ of GW kinetic energy $E_{kin} = \rho_r \langle |\mathbf{v}'|^2 \rangle_w / 2$ and total wave energy E_w . μ changes due to the impact of rotation on wave motion, i.e. for $\hat{\omega} \rightarrow N$: $\mu \rightarrow 1/2$ and for $\hat{\omega} \rightarrow f$: $\mu \rightarrow 1$. τ depends on the Prandtl number. If $\text{Pr} = 1$ then simply $\tau^{-1} = 2(\nu + \nu_t) |\mathbf{k}|^2$. *Dunkerton* [1989] argued that $\text{Pr} \gtrsim 3$ is more

realistic for breaking gravity waves as turbulent damping will be strongly localized in GW phases where mixing of entropy is low.

As shown in appendix A.4, the non-dissipative wave energy sources are related to the change of $\hat{\omega}$ along the ray so that eq. (4.94) reduces to the wave action law (see also *Bretherton and Garrett* [1968])

$$d_t A = -A \nabla \cdot \mathbf{c}_g - \tau^{-1} A, \quad (4.97)$$

where $A = E_w/\hat{\omega}$ denotes the wave action density and τ^{-1} is due mainly to wave breaking processes. The change in the volume of a ray bundle [*Walterscheid*, 2000] is determined by the divergence of the group flow. Wave action conservation was also studied in a much more general context [*Andrews and McIntyre*, 1978; *Grimshaw*, 1984].

The damping rate τ^{-1} is estimated via a highly simplified turbulence parameterization based on saturation theory [*Lindzen*, 1981]. In this scheme, the GW amplitudes are forced back to the convective instability threshold if they have the tendency to grow above it. τ^{-1} is calculated in a way to ensure that the saturation condition is fulfilled [*Holton*, 1982]. The explicit dependence of τ on the diffusion coefficient and Prandtl number will remain unspecified as in chapter 5 only GW forces are considered. For GW heating rates, it is supposed that a more sophisticated approach in parametrizing the turbulence impact is needed [see e.g. *Marks and Eckermann*, 1995]. Additionally, in the MLT region molecular viscosity and thermal diffusivity become more important and are included into the damping process. Note however that in the middle and upper thermosphere, also the dispersion of GW fields would be strongly affected by molecular motion [*Vadas and Fritts*, 2005].

Ray tracing in a shallow, spherical atmosphere

In the following, special emphasis on metric corrections appearing in a shallow, spherical atmosphere is made. This derivation appeared in *Senf and Achatz* [2011] and may be compared to *Hasha et al.* [2008]. Again following *Hayes* [1970], GW observed frequency ω and GW wavenumber vector \mathbf{k} are connected to local variations of the GW phase Θ by eq. (4.81) and (4.82). But now, the gradient in eq. (4.82) is expressed in spherical coordinates

$$\nabla = \frac{\mathbf{e}_\lambda}{a_E \cos \varphi} \frac{\partial}{\partial \lambda} + \frac{\mathbf{e}_\varphi}{a_E} \frac{\partial}{\partial \varphi} + \mathbf{e}_z \frac{\partial}{\partial z} \quad (4.98)$$

where according to the shallow atmosphere approximation [*Phillips*, 1966] the radial distance r is replaced by the mean radius of earth a_E . Note that \mathbf{k} is defined as local Cartesian quantity, but its projection on the set of spherical unit vectors $\{\mathbf{e}_\lambda, \mathbf{e}_\varphi, \mathbf{e}_z\}$ changes during its evolution.

As the unit vectors \mathbf{e}_i in their respective coordinate directions do not depend on time, the derivation of the ω -equation (4.90) remains unchanged. Only the evolution of \mathbf{k} is affected. Following the same strategy as before, the local tendency of \mathbf{k} in eq. (4.87) is

$$\frac{\partial \mathbf{k}}{\partial t} = -\frac{\partial \omega}{\partial k_i} \nabla k_i - \frac{\partial \omega}{\partial \Lambda_n} \nabla \Lambda_n, \quad (4.99)$$

where now the gradient is not expressed in Cartesian coordinates. The next crucial step is rewriting the term $c_{gi} \nabla k_i$ as an advective derivative supplemented by metric corrections, i.e.

$$c_{gi} \nabla k_i = c_{gi} \nabla (\mathbf{k} \cdot \mathbf{e}_i) = \nabla \mathbf{k} \cdot \mathbf{c}_g + c_{gi} \nabla \mathbf{e}_i \cdot \mathbf{k}. \quad (4.100)$$

Since $\nabla \mathbf{k} = \nabla \nabla \Theta$ is a symmetric tensor of second order, $\nabla \mathbf{k} = (\nabla \mathbf{k})^T$, it follows that $\nabla \mathbf{k} \cdot \mathbf{c}_g = \mathbf{c}_g \cdot \nabla \mathbf{k}$. Now, the projection $\mathbf{k} = k_i \mathbf{e}_i$ is applied once again to arrive at

$$c_{gi} \nabla k_i = (\mathbf{c}_g \cdot \nabla k_i) \mathbf{e}_i + \mathbf{c}_g \cdot \nabla \mathbf{e}_i k_i + c_{gi} \nabla \mathbf{e}_i \cdot \mathbf{k} \quad (4.101)$$

$$= (\mathbf{c}_g \cdot \nabla k_i) \mathbf{e}_i + \mathbf{c}_g \cdot \left(\nabla \mathbf{e}_i - (\nabla \mathbf{e}_i)^T \right) k_i, \quad (4.102)$$

where in the last line $\nabla (\mathbf{e}_i \cdot \mathbf{e}_j) = 0$ was used. Hence, the ray equations for the wavenumbers k_i are

$$(d_t k_i) \mathbf{e}_i = - \frac{\partial \omega}{\partial \Lambda_n} \nabla \Lambda_n - \mathbf{c}_g \cdot \left(\nabla \mathbf{e}_i - (\nabla \mathbf{e}_i)^T \right) k_i, \quad (4.103)$$

valid for general coordinate systems [Hasha *et al.*, 2008].

As before, the shallow atmosphere approximation will be used so that vertical derivatives of all unit vectors and all derivatives of the outward pointing unit vector \mathbf{e}_z are neglected. Thus, only the convergence of the meridians is taken into account and the metric terms are

$$\nabla \mathbf{e}_\lambda = \frac{\tan \varphi}{a_E} \mathbf{e}_\lambda \mathbf{e}_\varphi \quad \text{and} \quad \nabla \mathbf{e}_\varphi = - \frac{\tan \varphi}{a_E} \mathbf{e}_\lambda \mathbf{e}_\lambda. \quad (4.104)$$

These terms lead to an additional geometric refraction of the horizontal wavenumbers to compensate for the changes in the unit vectors due to the curvature of earth. Using the wave action conservation eq. (4.97) with

$$\nabla \cdot \mathbf{c}_g = \frac{\partial_\lambda c_{g\lambda} + \partial_\varphi (\cos \varphi c_{g\varphi})}{a_E \cos \varphi} + \partial_z c_{gz}, \quad (4.105)$$

a ray equation for the vertical flux $F_A = c_{gz} A$ of wave action density A is obtained, i.e. $d_t F_A = d_t c_{gz} A + c_{gz} d_t A$, which can be written as

$$d_t F_A = - (\tau^{-1} - \tau_{non}^{-1}) F_A, \quad (4.106)$$

where all non-dissipative effects have been collected into the rate

$$\begin{aligned} \tau_{non}^{-1} = c_{gz}^{-1} & \left(\partial_t c_{gz} + \frac{c_{g\lambda} \partial_\lambda c_{gz} - c_{gz} \partial_\lambda c_{g\lambda}}{a_E \cos \varphi} \right. \\ & \left. + \frac{(\cos \varphi c_{g\varphi}) \partial_\varphi c_{gz} - c_{gz} \partial_\varphi (\cos \varphi c_{g\varphi})}{a_E \cos \varphi} \right), \end{aligned} \quad (4.107)$$

which can be either positive or negative. τ_{non}^{-1} is derived by expanding and rewriting the terms $d_t c_{gz} A$ and $-A \nabla \cdot \mathbf{c}_g$ via $d_t = \partial_t + c_{g\lambda} / (a_E \cos \varphi) \partial_\lambda + c_{g\varphi} / a_E \partial_\varphi + c_{gz} \partial_z$ and eq. (4.105). Eq. (4.106) extends the relation given by Marks and Eckermann [1995] to time-dependent flows in spherical geometry. Changes in F_A result from dissipation via $-\tau^{-1} F_A$ and from temporal and horizontal variations of group velocity via $\tau_{non}^{-1} F_A$. The latter are connected to a local change of the volume which neighboring GW rays occupy [Broutman *et al.*, 2004]. In the upcoming simulations, the turbulent damping is the major contribution and changes in GW properties, chiefly ω and k_h , modify the GW breakdown, i.e. the damping rate τ^{-1} . Hence, time- and horizontal dependence of the background flow have a mostly indirect impact on the diurnal GW force by changing the turbulence parameterization. This is in contrast to direct non-dissipative forces due to transience and horizontal refraction, i.e. from $\tau_{non}^{-1} F_A$, as discussed in section 5.3.3 and by Dunkerton [1981], Eckermann and Marks [1996], and Bühler [2009].

In summary, the ray-tracing equations in a spherical shallow atmosphere are

$$d_t \omega = \mathbf{k} \cdot \partial_t \mathbf{u}, \quad (4.108)$$

$$d_t k = -\mathbf{k} \cdot \frac{\partial_\lambda \mathbf{u}}{a_E \cos \varphi} + \frac{k \tan \varphi}{a_E} \hat{c}_{g\varphi}, \quad (4.109)$$

$$d_t l = -\mathbf{k} \cdot \frac{\partial_\varphi \mathbf{u}}{a_E} - \frac{f m^2}{\hat{\omega} |\mathbf{k}|^2} \frac{\partial_\varphi f}{a_E} - \frac{k \tan \varphi}{a_E} \hat{c}_{g\lambda}, \quad (4.110)$$

$$d_t m = -\mathbf{k} \cdot \partial_z \mathbf{u} - \frac{N k_h^2}{\hat{\omega} |\mathbf{k}|^2} \partial_z N, \quad (4.111)$$

$$d_t F_A = -(\tau^{-1} - \tau_{non}^{-1}) F_A. \quad (4.112)$$

The metric corrections from \mathbf{u} lead to the change $\mathbf{c}_g \rightarrow \hat{\mathbf{c}}_g$ on the right-hand sides of eq. (4.109) and (4.110), where $\hat{\mathbf{c}}_g$ is the intrinsic group velocity. Again, the time-dependence of the background wind induces the modulation of ω along the ray. In Addition to the horizontal wind gradients, the geometric terms, the last terms of eq. (4.109) and (4.110), are responsible for a turning of the horizontal wavenumber vector \mathbf{k}_h . In a hypothetical isothermal earth at rest, a wave packet is forced to move on a great circle rather than on a latitude circle as in the plane-parallel atmosphere. Several aspects of numerical implementation of global ray tracing are discussed in the appendix A.6.

4.4. Eliassen-Palm theorems

In the study of gravity wave-tide interaction, forces and heating rates resulting from the non-dissipative energy exchanges along the path of GW propagation and from dissipative GW break-down are of major concern. Within simplified basic states, the way gravity waves interact with the background medium was stated in a very elegant form called the “*Eliassen-Palm theorems*”. They enable the estimation of the strength and location of mean-wind changes and are reviewed in the following.

The Eliassen-Palm theorems are stated in the book of *Lindzen* [1990]. “[The first] theorem tells us that the momentum flux is such that if deposited in the mean flow it will bring u towards c ” [*Lindzen*, 1990, p. 146], where c is the zonal GW phase velocity, i.e.

$$\langle p'w' \rangle_w = -\rho_r (u - c) \langle u'w' \rangle_w. \quad (4.113)$$

“[The second] theorem tells us that in the absence of (i) damping, (ii) local thermal forcing, and (iii) critical levels, no momentum flux is deposited or extracted from the basic flow” [*Lindzen*, 1990, p. 148], so that

$$\partial_z (\rho_r \langle u'w' \rangle_w) = 0. \quad (4.114)$$

Both theorems have proven extremely helpful for the interpretation of the interaction between gravity waves and a mean flow. However, several assumptions have been made in their derivation, namely (i) 2D-GW motion, (ii) no rotation, and (iii) only vertical shears in the background flow. In the following, generalized Eliassen-Palm relations are derived when the assumptions are relaxed.

It is shown in appendix A.3 that in the WKB approximation the vertical pressure flux is equal to the intrinsic vertical flux of wave energy E_w , i.e.

$$\langle p'w' \rangle_w = \hat{c}_{gz} E_w. \quad (4.115)$$

Since the considered gravity waves are affected by rotation the momentum flux is replaced by the more meaningful Eliassen-Palm flux (EP flux)

$$F_{\text{EP}} = -\rho_r (\langle u'_\parallel w' \rangle_w - f \langle b' u'_\perp \rangle_w / N^2) \quad (4.116)$$

where u'_\parallel and u'_\perp denote the velocity perturbation along and perpendicular to the horizontal wavenumber vector. As the EP flux $F_{\text{EP}} = -\hat{c}_{gz} k_h A$ gives the (negative of the) vertical flux of horizontal pseudo-momentum $k_h A$, the first Eliassen-Palm theorem generalizes to

$$\langle p' w' \rangle_w = (u_h - c_h) F_{\text{EP}}, \quad (4.117)$$

where c_h is horizontal phase velocity, u_h denotes the horizontal background wind in the wave direction and $A = E_w / \hat{\omega}$ is the wave action density. This may be stated with Lindzen's words: "the EP flux is such that if deposited in the mean flow it will bring u_h towards c_h ".¹⁰ Note that u_h must be interpreted as Lagrangian-mean velocity here, as F_{EP} is closer connected to the pseudo-momentum flux than to the momentum flux itself.

Also for the second EP-theorem, the impact on the Lagrangian-mean flow is of interest. As shown in appendix A.3, $\nabla \cdot (\rho_r \langle \xi \nabla_h \phi' \rangle_w) \approx \nabla \cdot (\hat{c}_g \mathbf{k}_h A)$ is the divergence of the intrinsic flux of the horizontal pseudo-momentum vector [*Grimshaw*, 1975a]. Hence, the dynamics of $\mathbf{k}_h A$ is needed. With the wave action law (4.97) and the \mathbf{k}_h -dynamics (A.83), one obtains

$$\partial_t (\mathbf{k}_h A) + \nabla \cdot (\mathbf{c}_g \mathbf{k}_h A) = -\nabla_h \mathbf{u} \cdot \mathbf{k}_h A - \frac{f \beta m^2}{\hat{\omega} |\mathbf{k}|^2} A \mathbf{e}_\varphi - \tau^{-1} \mathbf{k}_h A. \quad (4.118)$$

The horizontal inhomogeneities in the background conditions are the sources of the horizontal pseudo-momentum. Rewriting the last relation, a generalized variant of the second Eliassen-Palm theorem is found, i.e.

$$\nabla \cdot (\hat{c}_g \mathbf{k}_h A) = -\partial_t (\mathbf{k}_h A) - \nabla_h \cdot (\mathbf{u} \mathbf{k}_h A) - \nabla_h \mathbf{u} \cdot \mathbf{k}_h A - \frac{f \beta m^2}{\hat{\omega} |\mathbf{k}|^2} A \mathbf{e}_\varphi - \tau^{-1} \mathbf{k}_h A. \quad (4.119)$$

Therefore, even if (i) damping, (ii) heating or (iii) critical filtering is excluded, a local mean flow is forced by wave transience and the horizontal inhomogeneities in the background conditions.

¹⁰This actually implies that no pseudo-momentum is locally produced by the interaction with the mean flow.

4.5. Summary

- GW sources were reviewed.
- Linear dynamics of small-amplitude gravity waves were discussed.
- Gravity waves in a resting isothermal basic state were chosen as a didactic introduction.
- Then an extension to gravity waves in a large-scale moving basic states was provided.
- Within a multiple-scale analysis, the dynamics of gravity waves was viewed as a two-scale system with fast variations on the wave scales and slow variations on the background scales.
- Asymptotically, the three different GW regimes were investigated:
 - (i) the non-hydrostatic regime (high-frequency limit),
 - (ii) the medium-frequency regime,
 - (iii) and the inertia-gravity wave regime (low-frequency limit).
- Hybrid GW dynamics was constructed including the essential features of all regimes.
- Wentzel-Kramers-Brillouin theory for locally monochromatic wave trains was utilized.
- Local dispersion, polarization and wave action relations arose for slowly varying background flows.
- The ray-tracing method was introduced in which parts of the GW field are traced along its group velocity.
- The impact of GW fields on the background flow was briefly discussed with the help of generalized Eliassen-Palm theorems.

5. Modulation of gravity waves in thermal tides

Global ray-tracing simulations are presented in this chapter. Parts of gravity-wave fields are traced through a time-changing background flow including tides from HAMMONIA. Effects of temporal and horizontal variations in the background conditions on gravity-wave propagation are investigated. Diurnal gravity-wave forces are calculated and compared to a conventional gravity-wave parameterization.

5.1. Introduction

The periodic change of the diurnal winds modulates GW fields propagating through it. As indicated in section 4.3.1, different Doppler shifts $\mathbf{k} \cdot \mathbf{u}$ occur during different tidal phases. Changes in $\mathbf{k} \cdot \mathbf{u}$ induce changes in certain GW properties and appear as source terms in the set of ray-tracing equations (4.109)-(4.112). The impact of vertical wind shear is of major importance for vertically propagating gravity waves and usually incorporated into conventional GW parameterizations [McLandress, 1998]. However, there are also effects resulting from the time-dependence of the diurnal tides and the horizontal inhomogeneities of the combined background flow. The first is responsible for a periodic modulation of the GW observed frequency while the latter inducing a mostly irreversible refraction of the horizontal wavenumbers. Both effects are introduced and discussed below. Some of the upcoming discussion is published in *Senf and Achatz [2011]* and *Achatz et al. [2012]*.

In the upcoming sections, frequency modulation and horizontal refraction are presented in two different ways: On the one hand, simplified examples are given to illustrate the underlying principles of each phenomenon, and rough estimates are provided regarding how the different effects can impact on the GW-tide interaction. On the other hand, global ray-tracing simulations within realistic tides have been performed for a set of experiments to obtain quantitative results. For this, a new global ray-tracing model was built by the author to fulfill all requirements of a consistent ray description in a tidal background which was not available from existing models.¹ The diurnal GW force is the major concern of this study as well as how it changes due to the impact of frequency modulation and horizontal refraction. The realistic GW simulations are compared to the conventional vertical-column approach which is usually applied in GW parameterizations.

5.2. Global ray-tracing simulations with RAPAGI

5.2.1. Gravity-wave ensemble

A global numerical model was built by the author based on the set of ray-tracing eqns. (4.108)-(4.112). The model, named RAPAGI (Ray PAMeterization of Gravity-wave ImPacts), is a global ray-tracing model, the implementation details of which are given in

¹ For instance, the ray model GROGRAT maintained by Stephen Eckermann, Naval Research Laboratory, Washington is public available, but does not incorporate effects of temporal and horizontal dependence of the basic state in the calculation of wave amplitude.

appendix A.6. The RAPAGI simulations used a small and highly idealized ensemble of gravity waves previously used by *Becker and Schmitz* [2003], listed in tab. 5.1. Gravity

#	α [deg]	L_h [km]	c_h [ms ⁻¹]	F_h [10 ⁻³ J m ⁻³]
1	0	385	6.8	0.32
2	45	410	6.8	0.38
3	90	504	10.2	0.35
4	135	570	6.8	0.38
5	180	596	6.8	0.45
6	225	570	6.8	0.38
7	270	504	10.2	0.35
8	315	410	6.8	0.38
9	0	385	32.8	0.32
10	45	410	20.4	0.38
11	135	570	20.4	0.38
12	180	596	32.8	0.45
13	225	570	20.4	0.38
14	315	410	20.4	0.38

Table 5.1.: 14 members of the GW ensemble used in the simulations. α denotes the azimuth angle towards which the waves propagate, zero towards the east and increasing counter-clockwise. L_h and c_h are the horizontal wavelength and phase velocity in the wave direction and F_h the vertical flux of horizontal pseudo-momentum at the lower boundary $\hat{z}_B = 20$ km.

waves with horizontal wavelengths between about 400 km and 600 km and random initial phases are globally homogeneously and continuously emitted at the lower boundary, $\hat{z}_B = 20$ km. Each of the 14 independent GW components is integrated forward separately. The gravity waves are directed into 8 equally spaced azimuth directions, beginning at east and increasing counter-clockwise at 45 intervals. The GW ensemble is non-isotropic with largest k_h directed to the east, the largest c_h in the zonal directions and the largest momentum flux to the west (see tab. 5.1).

It was shown by *Becker and Schmitz* [2003] that the mean residual circulation of the middle atmosphere is well reproduced in a large-scale GCM when their GW ensemble is used with a Lindzen GW parameterization. Note, however, that as it mostly resulted from tuning the GW parameters, this GW ensemble is just one of many possibilities. Therefore, the simple GW ensemble is viewed as a “toy” configuration in which the effect of temporal and horizontal variation of the background conditions can be investigated.

5.2.2. Experimental setup

A hierarchy of three different experiments, named in order of decreasing complexity “full”, “noREF” and “TS”, was used, as listed in tab. 5.2.

The “full” experiment is a full ray-tracing simulation without any approximations for horizontal and time dependence, i.e. the unmodified eqns. (4.108) - (4.112) are integrated along the ray path given by eqns. (A.98), (A.99), (A.100). Thus, changes in

dependence	“full”	“noREF”	“TS”
time	yes	yes	no
horizontal	yes	no	no

Table 5.2.: Overview of the three different experiments which were performed with the same initial conditions and background medium. “full”, “noREF” and “TS” are the short-cuts of experiments explained in detail in the text.

ω , \mathbf{k}_h and m appear, induced by mean flow changes. Also, the geographical distribution of the GW fields is altered in the process.

“noREF” (no refraction) is a simplified ray-tracing experiment in which neither horizontal refraction nor horizontal propagation are allowed. In the simulation, the right-hand sides of eqns. (A.98), (A.99) and (4.109), (4.110) have been set to zero. Additionally, horizontal derivatives and curvature terms in eq. (4.112) have been ignored. Hence in experiment “noREF”, ray points are only allowed to propagate vertically but have a finite group velocity and feel the transience of the background wind. The horizontal wavenumber vector \mathbf{k}_h is constant along each ray, but ω and m vary to compensate for temporal and vertical changes in the background conditions, respectively.

The third experiment, denoted by “TS” (time slicing), is equivalent to a Lindzen-type vertical column parameterization with temporally fixed background fields at each time step [McLandress, 1998]. Only vertical variations of the background are taken into account. Simulations with a stationary background fixed at each tidal phase, i.e.

$$u_{\text{TS},n} = \bar{u} + u_R \cos(2\pi n/12) + u_I \sin(2\pi n/12) \quad (5.1)$$

for $n = 1, \dots, 12$, were used for a good sampling of the diurnal cycle. At the end, an analysis of the composite of the different phases was performed.

With the three experiments, effects of frequency modulation and refraction of the horizontal wavenumber vector can be extracted. Differences between “TS” and “noREF” are attributed to the first, whereas differences between “noREF” and “full” to the latter. As the simpler simulations “TS” and “noREF” are obtained by successively simplifying the “full” ray-tracing system, a consistent comparison of the results is possible while keeping numerical and implementation parameters the same.

5.3. Gravity-wave frequency and phase-speed modulation

5.3.1. Motivation

The local time-dependence of the diurnal tides has a large impact on GW motion. First, the background medium changes while gravity waves are moving through it and seeing different temporal snapshots of the tides at different altitude levels. Therefore, depending on the travel time of individual gravity waves it might be incorrect to assume a stationary background field, keeping mean and tidal flow fixed, as is done in most GW parameterizations [McLandress, 1998] and in the “TS” experiment. Second, the GW properties are directly affected by the time-dependence of the diurnal tides which seems to be of major importance. From the ray equation (4.108), it is clear that the observed frequency ω is changed by the periodic influence of the diurnal tides. This effect is called “*frequency modulation*” and is explained in detail in the following sections. In the sections 5.3.2-5.3.4, the mechanism of frequency modulation is introduced within

a simple minimalistic example, and the energetics of non-dissipative wave-mean flow interactions and the impact on the saturation region are discussed. In section 5.3.5, the investigation is extended to more realistic flows and the experiments from section 5.2.2 are utilized to quantify frequency modulation of gravity waves in HAMMONIA diurnal tides.

5.3.2. The mechanism of frequency modulation

For an introduction to the frequency modulation, a simple background flow is considered here. It consists only of a zonal wind component $u(z, t) = U \sin(Mz - \Omega t)$ with the large-scale vertical wavenumber $M = -2\pi/L_z$, the vertical wavelength L_z and the amplitude profile $U(z)$ [Broutman, 1984; Broutman and Young, 1986; Eckermann and Marks, 1996]. The thermodynamic background state is isothermal. For diurnal tides with $M = -2\pi (30 \text{ km})^{-1}$ and $\Omega = 2\pi (1 \text{ day})^{-1}$, the phase velocity $C = \Omega/M$ is downward and in the order of 0.3 m/s. Two examples are investigated in the following: two continuously emitted GW trains and a transient GW packet.

First, consider two counter-propagating GW trains, one propagating in the positive zonal direction and one in the negative with zonal wavenumbers $k = \pm k_0$. Both are continuously emitted at the ground with phase velocities $c = \pm c_0$, respectively. A very simple profile of the tidal amplitude is assumed. The wind amplitude U is zero at the ground and slowly increases to a constant value at a certain altitude z_T . Then the ray-tracing eqns. (4.108), (4.109) reduce to

$$d_t \omega = k \partial_t u, \quad (5.2)$$

$$d_t k = 0, \quad (5.3)$$

which can be combined to the ray equation for the zonal phase velocity $c = \omega/k$

$$d_t c = \partial_t u. \quad (5.4)$$

Therefore, a local tendency of background wind is connected to a change of zonal phase velocity c along the ray. But as also shown by Eckermann and Marks [1996] and Walterscheid [2000], phase velocity changes arise only due to frequency changes. Above z_T , u is monochromatic and a solution of the form $c(z - Ct)$ can be found for which eq. (5.4) becomes

$$\left(1 + \frac{c_{gz}}{|C|}\right) \partial_t c = \partial_t u. \quad (5.5)$$

Assuming small U , the ansatz

$$c \approx \pm c_0 + \delta c \sin(Mz - \Omega t) \quad (5.6)$$

gives the phase velocity variation δc to lowest order

$$\delta c = \frac{U}{1 + \frac{c_{gz,0}}{|C|}}, \quad (5.7)$$

where $c_{gz,0}$ is the initial vertical group velocity at the ground. For upward propagating gravity waves, the frequency modulation acts so that c follows u . The exact solution of eq. (5.5) and its linear approximation (5.6) and (5.7) are shown in fig. 5.1 for a monochromatic tidal wave with constant amplitude $U = 40$ m/s. The linear approximations (5.6) and (5.7) perform quite well even in the case of large U . Only the peak

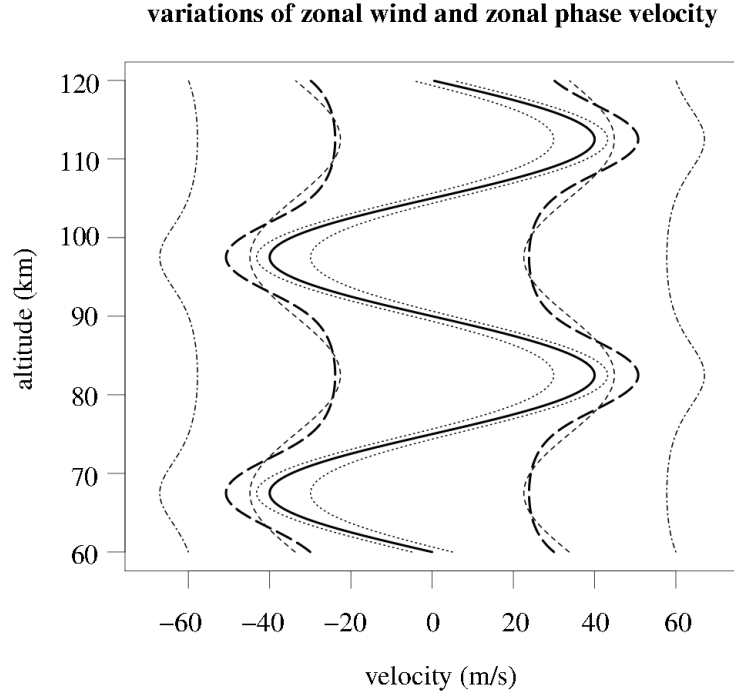


Figure 5.1.: The vertical profiles of zonal wind (thick solid) with $U = 40 \text{ ms}^{-1}$ and the exact solution for c of eq. (5.5) for 3 counter-propagating GW pairs: $c_0 = 5 \text{ ms}^{-1}$ (dotted), $c_0 = 30 \text{ ms}^{-1}$ (thick dashed) and $c_0 = 60 \text{ ms}^{-1}$ (dot-dashed). The linear approximation (5.6) for $c_0 = 30 \text{ ms}^{-1}$ (thin dashed) is also shown. In the calculations, the Coriolis effect was neglected and z_T was far below 60 km.

values of c in fig. 5.1 are under-estimated in regions where vertical GW motion is slowed down.

The second example, the propagation of a transient GW packet is shown in fig. 5.2. A Gaussian wave packet, with its envelope centered at 24 h and with an envelope width of 4 h, is excited at the lower boundary $z = 0$. Initially it has a zonal wavelength of 300 km, a zonal phase velocity of 40 m/s and a zonal momentum flux of 0.4 mJ/m^3 . It is traced through a tidal wind $u(z, t) = U \sin(Mz - \Omega t)$ where the zonal wind amplitude U is Gaussian-shaped with a maximum of 40 m/s at 100 km altitude and a width of 20 km. In the time-altitude plot, the rays are curved with steeper slopes in tidal phases with negative wind and shallower slopes in positive wind phases. The wave packet is vertically accelerated in the former and decelerated in the latter. The GW phases move downward as the packet envelope propagates upward. The GW phase lines are steeper in negative and shallower in positive tidal winds. The tilting of GW phase lines is connected to an increase of vertical wavenumber and a deceleration of the GW packet. Also visible in fig. 5.2 is that the GW period is increased in tidal minima and decreased in tidal maxima. Thus, the transience of the background tide enforces a frequency modulation proportional to the tidal wind itself. The largest GW temperature amplitudes appear in altitudes below the tidal maxima. Further above, GW fields are strongly damped by saturation.

The second panel of fig. 5.2 shows the amplitude of the tidal wind U and the amplitude

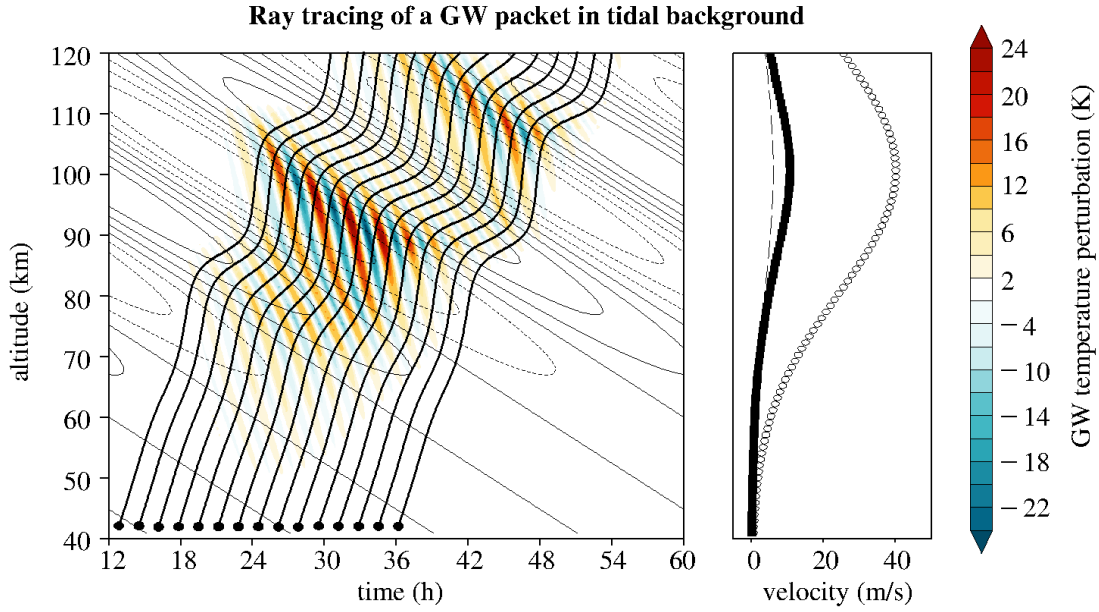


Figure 5.2.: Propagation of a GW packet through a tidal background. Left: The zonal tidal wind and the GW temperature perturbation are shown in contours and colors with intervals of 10 m/s and 2 K, respectively. The time evolution of an initially Gaussian wave packet is obtained by ray tracing (each 20th ray is indicated by thick lines). Right: The amplitude of the tidal wind (open circles), amplitude of c -modulation (filled squares) and the estimate via (5.7) (dashed).

of the GW phase velocity variation δc . At their maxima at 100 km, the variation of GW phase velocity is about 27% of the tidal wind variation. For the given GW packet with $k_h = 2\pi(300\text{km})^{-1}$ and $c_0 = 40$ m/s, the vertical group velocity is 1.7 m/s compared to the local tidal phase velocity of $C \approx -0.3$ m/s. From the earlier (in this case crude) estimate (5.7), a relative variation of 15% is expected. This suggests that eq. (5.7) gives a lower bound for the phase velocity modulation in more realistic tidal winds with possibly substantial positive deviations.

In general, as pointed out e.g. by *Walterscheid* [2000], the slower the GW propagates in the vertical, the more pronounced is the effect of frequency modulation. For eq. (5.7), two limits are obtained: for $c_{gz,0} \ll |C|$, $\delta c \rightarrow U$ and for $c_{gz,0} \gg |C|$, $\delta c \rightarrow 0$. Hence, the effect is more important for slow gravity waves and seems to be negligible for fast gravity waves. Based on this, one might conclude that the effect of frequency modulation is restricted to only very slowly vertical propagating gravity waves. However, it is believed that critical-layer-type interactions with the mean wind are very important to induce breakdown and dissipation of gravity waves in the middle atmosphere and especially in the mesopause region [*Fritts and Alexander, 2003*]. When a spectrum of gravity waves approaches a critical region, a large part of the spectrum is slowed down and becomes aware of the time-dependence of the background conditions. This makes the author believe that the effect of frequency modulation is of general importance in a realistic middle atmosphere including temporal variation of solar-thermal tides.

5.3.3. Simple energetics of frequency modulation

At first glance, it seems to be surprising that the observed frequency ω can change for an observer that remains fixed at the ground. As the frame of reference is not altered in the observation, ω -modulation is not caused by Doppler shifts in contrast to the intrinsic frequency $\hat{\omega}$. It is discussed in the following that the energetic exchange with

the induced mean flow is responsible for the ω -modulation.

As in *Broutman and Grimshaw* [1988], energetics for the simple example above are considered. Only non-dissipative interactions between the waves and the mean flow are taken into account. A dynamical situation is constructed in which a zonal background flow only depending on z and t is sustained. It is assumed that the local inertia of the background is entirely balanced by the Coriolis force. Hence, there is a need to also include a meridional background wind component, but no pressure or entropy perturbations are excited in this (artificial) setup. Throughout the discussion, it is assumed that the GW amplitudes are small and that a convergent expansion of the mean flow in orders of small wave amplitudes exists. Therefore, the zero order mean flow \mathbf{u}_0 is distinguished from the wave-induced flow \mathbf{u}_2 which is second order in wave amplitudes. The two sum up to \mathbf{u} . As defined, \mathbf{u}_0 is governed only by inertial motions, i.e.

$$\partial_t \mathbf{u}_0 + f \mathbf{e}_z \times \mathbf{u}_0 = 0, \quad (5.8)$$

where $f = \Omega$. The last relation defines a suitable meridional flow v_0 . The kinetic energy of this lowest order mean flow does not change, $\partial_t(\rho_r \mathbf{u}_0^2/2) = 0$, and by construction, the corresponding available potential energy is zero.

Next, non-dissipative medium-frequency GW dynamics are assumed. A zonally propagating GW packet extending infinitely in both horizontal directions is excited at a source level where its amplitude and corresponding zonal pseudo-momentum $\mathcal{P} = kA$ might change in time (e.g. as provided in fig. 5.2). On the way up, \mathcal{P} is conserved since (i) no GW dissipation and (ii) no zonal inhomogeneities in the background conditions are taken into account. From eq. (4.119),

$$\partial_t \mathcal{P} = -\partial_z F \quad (5.9)$$

where the pseudo-momentum flux $F = c_{gz} \mathcal{P}$ is equal to the GW momentum flux.² The GW energy $E_w = \rho_r/2 (\langle |\mathbf{u}'|^2 \rangle_w + \langle b'^2 \rangle_w / N^2)$ is not conserved and obeys (compared to eq. (4.94))

$$\partial_t E_w + \nabla \cdot (\mathbf{u}_0 E_w + \rho_r \langle \mathbf{v}' \phi' \rangle_w) = -F \partial_z u_0. \quad (5.10)$$

The GW packet exchanges energy with the mean flow via shear production. The last relation further simplifies as here no horizontal inhomogeneities were taken into account and thus only the vertical energy transport $\rho_r \langle w' \phi' \rangle_w = \hat{c}_{gz} E_w$ matters. Hence, the wave energy equation reduces to

$$\partial_t E_w + \partial_z (\rho_r \langle \mathbf{v}' \phi' \rangle_w) = -F \partial_z u_0. \quad (5.11)$$

Next, it is discussed how this transfer of energy from the waves to the mean flow establishes a wave-induced flow.

The second order mean flow is influenced by GW motion. The momentum balance is

$$\partial_t \mathbf{u}_2 + f \mathbf{e}_z \times \mathbf{u}_2 = -\frac{1}{\rho_r} \partial_z F \mathbf{e}_\lambda, \quad (5.12)$$

where it was assumed that the gravity waves only propagate in the zonal direction. The induced zonal mean flow u_2 behaves like a forced harmonic oscillator,

$$\partial_{tt}(\rho_r u_2) + f^2 \rho_r u_2 = \partial_{tt} \mathcal{P}. \quad (5.13)$$

²Remember that this is only the case if the impact of rotation on GW motion is ignored.

Therefore, u_2 has two parts: the first is a locally-induced mean flow $u_2^{(loc)}$ and equal to the GW pseudo-momentum \mathcal{P} , whereas the second, $u_2^{(sw)}$, describes a train of secondary waves, in this case inertial waves [Broutman and Grimshaw, 1988]. Thus,

$$u_2 = u_2^{(loc)} + u_2^{(sw)} \quad (5.14)$$

where

$$\rho_r u_2^{(loc)} = \mathcal{P} \quad \text{and} \quad \rho_r u_2^{(sw)} = -f \int_{-\infty}^t d\tau \mathcal{P} \sin(f(t - \tau)). \quad (5.15)$$

For example, $u_2^{(loc)}$ and $u_2^{(sw)}$ are plotted in fig. 5.3 for the GW packet traced in fig. 5.2. The locally induced flow is positive and attached to the wave packet as it moves

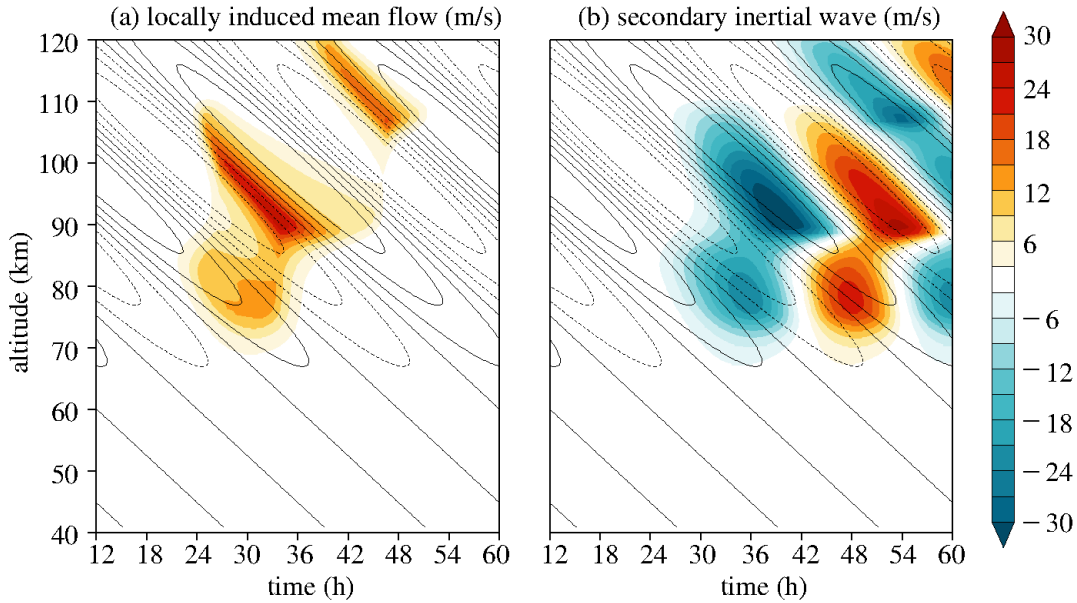


Figure 5.3.: The induced flow of a GW packet propagating through a tidal background. The zonal wind perturbation of the locally induced mean flow (left) and the secondary inertial wave (right) are shown in colors. Again, the first-order zonal tidal wind is plotted in contours with intervals of 10 m/s.

eastward. As dissipative effects were taken into account in the ray-tracing simulation, the wave packet is strongly damped via the saturation scheme below the positive tidal-wave crest at about 100 km and 36 h and then slowly recovers. Also, a secondary inertial wave is induced after the passage of the GW packet and remains. Interestingly, $u_2^{(sw)}$ is anti-correlated to u_0 in this example and would contribute to a reduction of tidal amplitudes. However, it is not obvious that this would generally be the case for induced secondary waves and needs further investigation. In the example of fig. 5.3, the magnitude of the induced flow is comparable to the leading order flow and thus an expansion of the results in wave amplitude is not appropriate. Nevertheless, the rest of this section continues with the description of small-amplitude waves.

The kinetic energy $\rho_r \mathbf{u}_2^2/2$ would be fourth order in wave amplitude and does not contribute to the energetic budget at the given order, but as shown below, the coupling between lowest order and second order flow $\rho_r \mathbf{u}_0 \cdot \mathbf{u}_2$ gives a kinetic energy contribution which is important for the energy exchange between the mean flow and the waves.

Without external heating, the total energy of the wave-mean flow system is conserved and obeys (see eq. (2.26))

$$\partial_t (\rho_r |\mathbf{u} + \mathbf{u}'|^2/2 + \Pi_{an}) + \nabla \cdot \{ \rho_r (\mathbf{u} + \mathbf{v}') (|\mathbf{u} + \mathbf{u}'|^2/2 + \Pi_{an} + \phi) \} = 0, \quad (5.16)$$

where Π_{an} denotes the available potential energy of the coupled wave-mean flow system. The last equation is averaged over GW scales, horizontal energy fluxes are neglected and only terms up to second order in wave amplitudes are retained, i.e.

$$\partial_t (\rho_r |\mathbf{u}_0|^2/2 + \rho_r \mathbf{u}_0 \cdot \mathbf{u}_2 + E_w) + \partial_z (\rho_r \langle u'w' \rangle_w u_0 + \rho_r \langle \phi'w' \rangle_w) = 0. \quad (5.17)$$

The term $\rho_r \langle u'w' \rangle_w u_0$ is the mean advective flux of kinetic energy due to GW perturbations. It is obtained from linearizing $\langle w'|\mathbf{u} + \mathbf{u}'|^2/2 \rangle_w$. The mean advective flux of available potential energy $\langle w'\Pi_{an} \rangle_w$ is negligible as no entropy is transported vertically, i.e. $\langle w'b' \rangle_w = 0$. Using, $\partial_t (\rho_r |\mathbf{u}_0|^2/2) = 0$ and $F = \hat{c}_{gz} \mathcal{P}$, the energy conservation is reformulated as

$$\partial_t (E_w + \rho_r u_0 u_2^{(loc)} + \rho_r \mathbf{u}_0 \cdot \mathbf{u}_2^{(sw)}) + \partial_z (\hat{c}_{gz} (u_0 \mathcal{P} + E_w)) = 0. \quad (5.18)$$

The pseudo-energy, which is the sum of the wave energy and the coupling between lowest order and locally induced mean flow (compare e.g. *Broutman and Grimshaw* [1988]; *Bühler* [2009])

$$\mathcal{E} = E_w + u_0 \mathcal{P}, \quad (5.19)$$

appears in the last equation. Note that $\rho_r u_2^{(loc)} = \mathcal{P}$. The vertical flux of total energy is equal to the vertical flux of \mathcal{E} , i.e.

$$\rho_r \langle u'w' \rangle_w u_0 + \rho_r \langle \phi'w' \rangle_w = \hat{c}_{gz} \mathcal{E}. \quad (5.20)$$

Using $d_t \omega = k \partial_t u_0$, the energy budget

$$\partial_t E_w + \partial_z (\hat{c}_{gz} E_w) = -F \partial_z u_0, \quad (5.21)$$

$$\partial_t (\rho_r u_0 u_2^{(loc)}) + \partial_z (\hat{c}_{gz} u_0 \mathcal{P}) = A d_t \omega + F \partial_z u_0, \quad (5.22)$$

$$\partial_t (\rho_r \mathbf{u}_0 \cdot \mathbf{u}_2^{(sw)}) = -A d_t \omega, \quad (5.23)$$

is obtained. Note the following remarks:

- (i) The wave energy E_w is transported by \hat{c}_{gz} and exchanges with the contribution $\rho_r u_0 u_2^{(loc)}$ from the locally-induced flow via shear production $-F \partial_z u_0$.
- (ii) The energetic contribution $\rho_r u_0 u_2^{(loc)}$, which is connected to the locally-induced mean flow $\rho_r u_2^{(loc)} = \mathcal{P}$, is also transported by \hat{c}_{gz} and remains attached to the GW field. Furthermore, it interacts also with the secondary wave field which is locally excited via the ω -modulation term.
- (iii) The ω -modulation can only appear within a background medium which admits wave motion and furthermore, ω -modulation appears here in the lowest order of wave amplitudes because the background flow was excited in the past by some prior leading-order process, i.e. whatever causes the \mathbf{u}_0 -oscillation.

- (iv) If a horizontally compact GW packet propagates through the oscillating background flow, then the corresponding GW force would also induce a horizontal pressure gradient. As a result, vertically-propagating secondary gravity waves with a horizontal wavelength connected to the horizontal extend of the primary GW packet would emerge.
- (v) In contrast to *Lindzen* [1990], the term Fu_0 is not included in the transport of E_w as it is more appropriate for describing induced mean-flow dynamics.
- (vi) Clearly, the induced flow is, at least for the non-dissipative dynamics, of significant importance.

5.3.4. The impact on saturated gravity-wave trains

In the following, the impact of ω -modulation on the saturation of GW trains within a monochromatic tidal wave, $u = U \sin(Mz - \Omega t)$, is investigated, ignoring any induced mean flows. This might be appropriate in the region of GW breaking where momentum diffusion is dominant. Diurnal forces due to saturation for a conventional vertical-column parameterization are compared to the forces obtained with a time-dependent solution of eq. (5.5). For medium-frequency gravity waves, the zonal wind amplitude $|u'|$ is set back to the saturation threshold $|u'_s| = |c - u|$ above the level of convective instability [*Fritts*, 1984]. This also applies for time-dependent flows. Saturation leads to a flux of zonal GW pseudo-momentum

$$F = \frac{\rho_r}{2} \frac{k}{N} \hat{c}_h^3 = \frac{\rho_r}{2} \frac{k_0}{N} (c - u)^3, \quad (5.24)$$

where the horizontal intrinsic phase velocity $\hat{c}_h = \pm(c - u)$, the zonal wavenumber $k = \pm k_0$ and the reference density ρ_r are used. In the case of small U and $c_0 \gg U$, where in the conventional approach no critical levels are encountered and saturation is not disrupted due to strong wind shears which may overcome the effect of density decrease, the resulting force is $f_\lambda = -\partial_z F / \rho_r$.

In the conventional approach, the GW phase velocity is assumed to be constant, i.e. $c = \pm c_0$, and thus, the saturation flux becomes

$$F_{conv}^\pm = \frac{\rho_r}{2} \frac{k_0}{N} (\pm c_0^3 - 3c_0^2 U \sin(Mz - \Omega t) + \dots), \quad (5.25)$$

where terms nonlinear in U are not given explicitly. The diurnal force exerted on the mean flow due to the damping of counter-propagating gravity waves is

$$f_{conv,T} = -\frac{1}{\rho_r} \frac{\partial}{\partial z} (F_{conv}^+ + F_{conv}^-) \quad (5.26)$$

$$= -\frac{3c_0^2 U k_0}{N} \left(\frac{1}{H_\rho} \sin \Phi_T - M \cos \Phi_T \right), \quad (5.27)$$

with the tidal phase $\Phi_T = Mz - \Omega t$ and the density scale height $H_\rho = -(\partial_z \ln \rho_r)^{-1}$. Terms nonlinear in U/c_0 have been neglected.

On the other hand, by taking realistic GW propagation into account, the periodic change in the background wind induces a modulation of frequency and hence zonal phase velocity (see eq. (5.6) and (5.7)). This effect reduces the variation of the intrinsic horizontal phase velocity, the saturation pseudo-momentum flux and hence the diurnal force due to gravity waves. Utilizing eq. (5.6), the diurnal force

$$f_T = f_{conv,T} \left(1 - \frac{\delta c}{U} \right) \quad (5.28)$$

is obtained. Recall that $\delta c < U$, which is ensured by eq. (5.7). Therefore, the diurnal GW force f_T is reduced due to phase velocity variations compared to the conventional approach. Note also that no critical layer is encountered by the gravity waves in the time-dependent approach. The localized deposition of GW pseudo-momentum at the conventional critical layer is smoothed out by the effects of frequency modulation.

5.3.5. Vertical column thinking and phase velocity modulation in realistic flows

Large-scale circulation models need to apply GW parameterizations [McLandress, 1998]. Horizontal gradients of the background medium are neglected, leading via eq. (4.109) and (4.110) to a conserved horizontal wavenumber k_h . The time-dependence of the transient large-scale motion is also neglected in the vertical column. GW trains are assumed to feel a stationary background and adjust instantaneously to a given wind field. In this sense, perturbations in the GW field propagate infinitely fast to the levels above. The advective time scale associated with a vertically propagating GW field, however, can be on the order of a day. This does not mean that the scale-separation assumption that the GW times-scale $\sim \hat{\omega}^{-1}$ should be significantly less than a day is violated. In eq. (5.7), the ratio $c_{gz,0}/C$ can be interpreted as the ratio between the background time scale and the GW advective time scale and it, rather than $\hat{\omega}/\Omega$, directly affects the variation of GW phase velocities and diurnal forces.

The distribution of travel times of rays from the lower boundary at $\hat{z}_B = 20$ km to $\hat{z} = 88$ km is shown in figure 5.4 for the “full” experiment (see section 5.2.2). In this experiment, the GW ensemble (described in section 5.2.1) is propagated using the full set of ray-tracing equations (4.108)-(4.112) through a background flow consisting of a

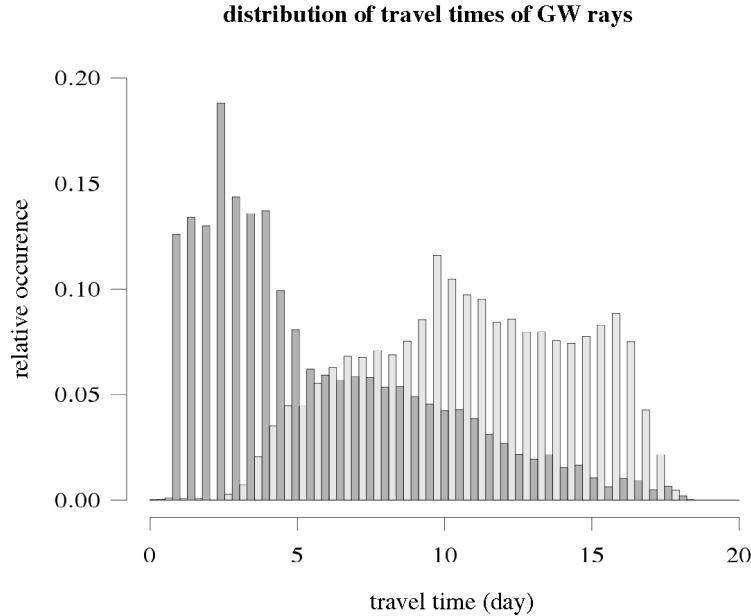


Figure 5.4.: Distribution of travel times from the lower boundary $\hat{z}_B = 20$ km to about $\hat{z} = 88$ km for a composite of slower GW members 1 to 8 (light gray) and faster GW members 9 to 14 (darker gray) from the “full” experiment. A bin size of half a day was chosen. Each pair of two neighboring bars shares the same bin for ease of comparison.

temporal-mean flow and the average diurnal variations obtained from the climate model HAMMONIA, as explained in section 3.1. For the histograms in fig. 5.4, the travel time of each ray calculated from the difference between its initial launch time and the time it reaches at a certain position is obtained. The travel time of ray points close to HAMMONIA grid points are taken and, depending on their distance to that grid point, a weighted average is constructed. Then the travel times at all horizontal grid points at the $\hat{z} = 88$ km level for all times of the last two days of integration are collected. The initially slower part of the GW ensemble, i.e. GW members 1 to 8 from tab. 5.1, has a broad almost uniform distribution in the range of 8 to 18 days. The faster GW members 9 to 14 contribute to a long tailed distribution with a maximum between 2 and 3 days. Some parts of the GW field with long travel times are captured near critical layers.

The experience obtained from the vertical column model has guided the conventional thinking of GW - mean flow interaction. The horizontal phase velocity of the gravity waves c_h is assumed to be constant with height. Its value is compared to the vertical profile of the horizontal background wind in the GW direction, $u_h = \mathbf{u} \cdot \mathbf{k}_h / k_h$. The difference between the two, i.e. the intrinsic horizontal phase velocity $\hat{c}_h = c_h - u_h$, is to a good approximation directly proportional to the vertical GW wavelength (see medium-frequency approximation (4.20)). When \hat{c}_h approaches its minimum, the vertical structure of the GW contracts and turbulent diffusion becomes much more effective. The saturation momentum flux (5.24) is proportional to \hat{c}_h^3 whose vertical variations determine the GW force on the background medium. Hence, a consistent estimate of \hat{c}_h is of major importance.

Fig. 5.5 shows the zonal phase velocity for GW ensemble member 12 (see tab. 5.1) at $\lambda = 0$ and $\varphi = 15\text{N}$ from the “noREF” experiment at four different times. In the “noREF” experiment, horizontal refraction and propagation of GW rays is switched off and rays move only vertically through a basic state which includes temporal mean and

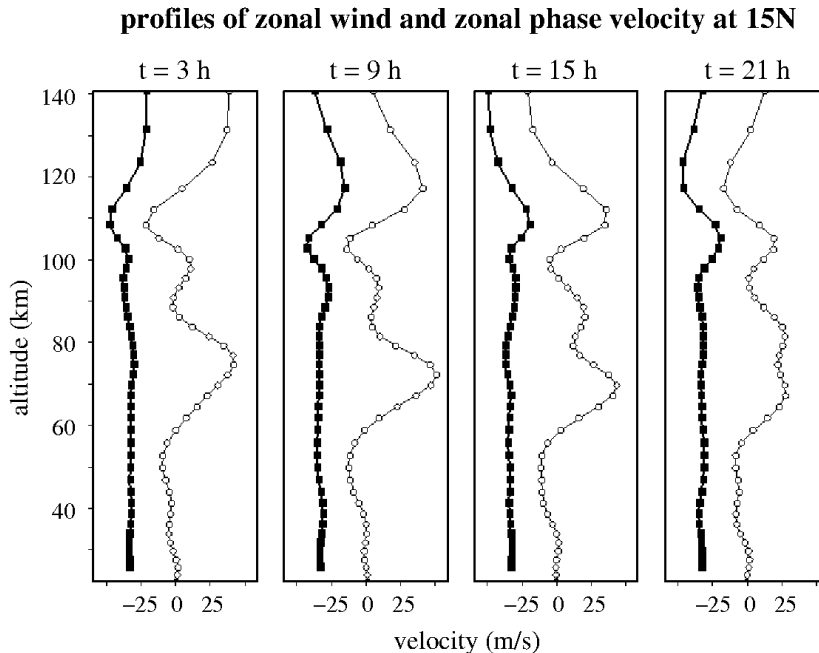


Figure 5.5.: Vertical profiles of the zonal background wind (open circles) and zonal GW phase velocity (filled squares) of the westward GW ensemble member 12 at $\lambda = 0$ and $\varphi = 15\text{N}$ for the “noREF” experiment at four different times.

average diurnal fields from HAMMONIA in representative January conditions. The GW member 12 moves westward and thus has favorable conditions within the mean zonal west-wind. But in contrast to the conventional vertical-column thinking, the zonal phase velocity c is not constant with altitude! Beyond 80 km, phase velocity variations δc grow to 15 m/s. As horizontal refraction is not allowed here, the variations in phase velocity result only from frequency modulation due to the periodic changes of the diurnal tides.

For the eastward-propagating GW member 9, a temporal snapshot of c_h and u_h at time $t = 0$ and day 16 is given in fig. 5.6 at $\varphi = 15\text{S}$ for the “full” experiment. This experiment again includes effects of horizontal propagation and refraction. The

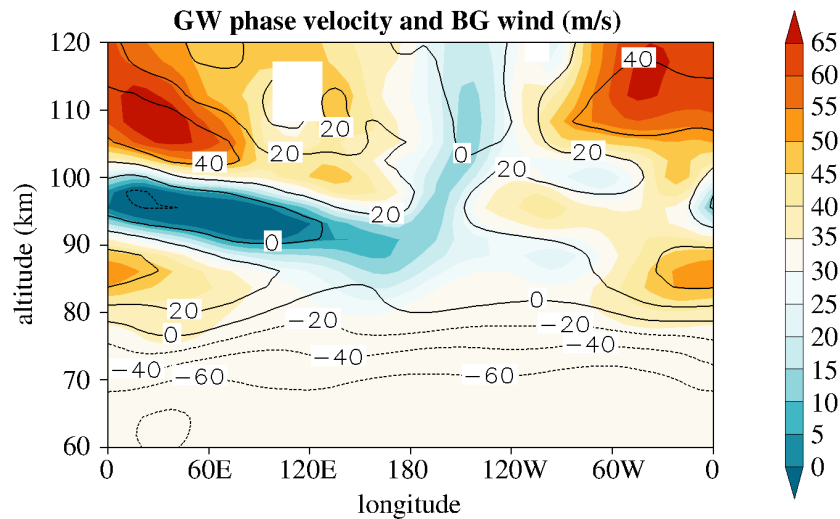


Figure 5.6.: The horizontal phase velocity c_h (colors) and the horizontal background wind u_h in the wave direction (contours) at $t = 0$ of day 16 and $\varphi = 15\text{S}$ for the eastward propagating GW member 9 in the “full” experiment.

variations in c_h match surprisingly well to the variations in u_h in the mesopause region. This suggests a “cooperation” between frequency modulation and horizontal refraction due to tides. The amplitude of the c_h -modulation becomes more than 30 m/s, larger than the initial phase velocity in the lower thermosphere. Hence, there is no single (constant) phase velocity which can be attributed to the GW field when the temporal variation of the thermal tide is present. Furthermore, due to the large ω -variations, negative values of c_h occur at the minima of the tidal winds.

The frequency modulation due to diurnal tides is one of the most important effects that reduces the strength of the GW-tide interaction. This has already been suggested in the context of the simple example (see eq. 5.28) and is inspected in more detail with respect to realistic diurnal tides in section 5.5.

5.4. Horizontal propagation and refraction of gravity-wave fields

5.4.1. Motivation

It might also be of interest to investigate horizontal deviations from the vertical-column approximation and its impact on the GW-tide interaction. First, the horizontal motion of GW fields can lead to large horizontal displacements. This happens especially in

regions where vertical GW motion is slowed down. The average horizontal GW propagation alters the mean distribution of GW fields, changing the time-mean force due to gravity waves but also the diurnal GW force. Second, the horizontal inhomogeneities of the temporal mean flow and of the diurnal variations induce changes in the horizontal wavenumbers k and l as given by eq. (4.109) and (4.110), called “horizontal refraction”. The temporally averaged horizontal refraction leads to changes in the GW fields which indirectly affect the diurnal tides. In the next sections, simple examples of the mean refraction effects are introduced. A discussion of the horizontal refraction in more realistic flows is provided in section 5.4.4

5.4.2. Mechanisms of mean horizontal refraction

Dunkerton [1984] and *Eckermann* [1992] showed that it is mainly the meridional gradients of the mean zonal wind $[\bar{u}]$ that cause horizontal refraction of GW fields in the middle atmosphere. Gravity waves propagating against the jet are refracted into its maximum. This is easily shown using the ray-tracing eq. (4.110). Typical northern winter conditions for the upper stratosphere are assumed. A change in l due to the zonal mean wind is

$$d_t l|_{mean} = -k \partial_\varphi [\bar{u}] / a_E, \quad (5.29)$$

where for simplicity curvature effects have been excluded. In the winter hemisphere, gravity waves propagating against the jet, i.e. with $k < 0$, have the best propagation conditions. On the northern flank of $[\bar{u}]$, the wind increases with decreasing φ , thus $\partial_\varphi [\bar{u}] < 0$, and similarly $\partial_\varphi [\bar{u}] > 0$ on the southern flank. Hence, $d_t l|_{mean} < 0$ north of the jet and > 0 south of the jet. An initially zonally aligned GW packet is refracted into the winter west-wind jet. In the summer hemisphere, the east-wind jet with $[\bar{u}] < 0$ supports eastward GW motion with $k > 0$, but because the Doppler shift $k[\bar{u}]$ is negative, as in the winter hemisphere, the same arguments apply. The effect is illustrated in fig. 5.7(a).

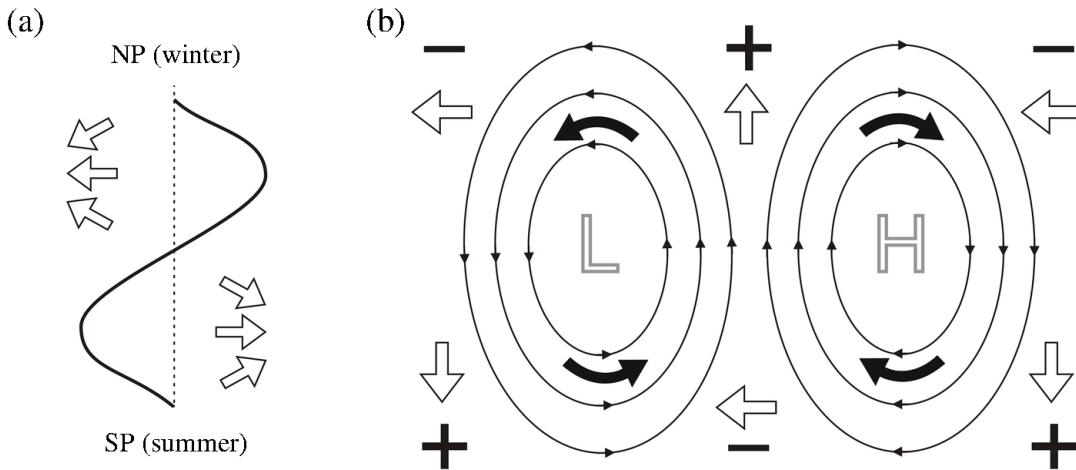


Figure 5.7.: Schematic view of the mean refraction of the horizontal wavenumber vector. Panel (a): $[\bar{u}]$ (solid line) induces refraction of k_h (open arrows) into the jet. Panel (b): Horizontal velocity due to vorticity (bold arrows) and strain deformation (plus/minus signs) of a stationary planetary wave (solid streamlines) control the tendencies in k (horizontal open arrows) and l (vertical open arrows).

A rough estimate of the change in meridional wavenumber $|\Delta l|$ of an initially zonally directed GW field is presented in the following. Within a latitude interval of around 20° , equivalent to a meridional distance of $\Delta y \approx 2000$ km, the mean zonal wind $[\bar{u}]$ increases (decreases) by about $\Delta u \approx 60$ m/s from high- to mid-latitudes in the northern (southern) hemisphere upper stratosphere. Hence, the meridional gradient is $|\partial_\varphi[\bar{u}]/a_E| \approx |\Delta u/\Delta y| \approx 3 \times 10^{-5} \text{ s}^{-1}$. A GW with initial values of $k = 2\pi (300 \text{ km})^{-1}$, $c = 20$ m/s, and thus $c_{gz} \approx 0.4$ m/s, excited in the lower atmosphere, propagates into the jet region. Due to the meridional wind gradients, it is refracted into the jet core, while the vertical wind shear leads to a decrease of its vertical wavenumber $|m|$. The GW packet accelerates up to a maximum vertical group velocity of ≈ 6 m/s. The wave field goes along a path of minimal travel time (analogous to Fermat's principle in geometric optics). Between 30 km and 70 km altitude, i.e. $\Delta z \approx 40$ km, the average group velocity is about $\bar{c}_{gz} \approx 3$ m/s. Thus, the GW is there for $\Delta t \approx \Delta z/\bar{c}_{gz} \approx 1.3 \times 10^4 \text{ s}$, or about 4 hours. During this time interval, meridional refraction is most effective and induces a cumulative change of $|\Delta l|/k \approx |\partial_\varphi[\bar{u}]/a_E|\Delta t \approx 40\%$ above the jet. Note that since the zonal wavenumber remains constant ($\partial_\lambda[\bar{u}] = 0$, ignoring the metric correction in eq. (4.109)), the horizontal wavenumber $k_h = |k|\sqrt{1 + \Delta l^2/k^2}$ increases by about $|\Delta k_h/k| \approx \Delta l^2/(2k^2) \approx 8\%$. In addition to the effect on k_h , there is horizontal redistribution of the GW field.

Planetary Rossby waves in the winter stratosphere also affect the horizontal refraction of gravity waves [*Dunkerton and Butchart*, 1984]. A simple wave field in a channel with Cartesian geometry is shown in fig 5.7(b). The planetary wave is described by a simple stream function $\psi = -\Psi \sin(x) \sin(y)$, where x and y have been scaled by the channel size and Ψ is an arbitrary constant. For $u = -\partial_y \psi$ and $v = \partial_x \psi$, the divergence and shear deformation of the horizontal background wind are zero, so gravity waves in this simple planetary wave are only affected by the vorticity $\zeta = 2 \partial_x v$ and the strain deformation $\vartheta = 2 \partial_x u$. Following *Bühler* [2009], the tendencies of k and l are

$$d_t \mathbf{k}_h|_{pw} = -\mathbf{S} \cdot \mathbf{k}_h, \quad (5.30)$$

with the wind-shear tensor

$$\mathbf{S} = \begin{pmatrix} \partial_x u & \partial_x v \\ \partial_y u & \partial_y v \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \vartheta & \zeta \\ -\zeta & -\vartheta \end{pmatrix}. \quad (5.31)$$

Thus, the planetary wave vorticity leads to a rotation of \mathbf{k}_h , via $d_t k|_\zeta = -\zeta l$ and $d_t l|_\zeta = \zeta k$, in the sense of the background vorticity. Cyclonic vorticity leads to anti-clockwise turning of \mathbf{k}_h and anticyclonic vorticity leads to clockwise turning of \mathbf{k}_h . The strain deformation induces via $d_t k|_\vartheta = -\vartheta k$ and $d_t l|_\vartheta = \vartheta l$ a change in the magnitude of the corresponding wavenumbers. For instance, positive strain increases the magnitude of l . Both effects are summarized in fig. 5.7(b). Initially westward propagating gravity waves crossing the planetary wave trough in high latitudes are refracted to the south downstream of the ridge. Gravity waves from lower latitudes are refracted northward upstream of the ridge (down- and upstream with respect to $[u] > 0$).

The beta-effect (the second term on the right-hand side of eq. (4.110)) is usually small and causes a decrease of l in the northern and an increase of l in the southern hemisphere mid-latitudes. Lastly, geometric effects due to the spherical shape of earth induce additional GW refraction (the last terms in eq. (4.109) and (4.110)). In a hypothetical isothermal earth at rest, the angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{k}A$ of a non-dissipating GW packet is constant along its path. This is shown in appendix A.5 for a shallow atmosphere. The wave packet is forced to move on a great circle [*Dunkerton*

and Butchart, 1984; Hasha *et al.*, 2008]. For an initially zonally aligned GW at middle or higher latitudes, the geometric refraction causes equatorward motion.

5.4.3. The impact on saturated gravity-wave trains

In section 5.3.4, the impact of frequency modulation on saturated GW trains was discussed. A related topic is raised here with emphasis on the impact of horizontal (meridional) refraction on saturated wave trains. Again, the conventional vertical-column approach for the simple example described above is compared to a solution which takes horizontal gradients of the background flow into account. In the following, the GW train with initial values of $k = 2\pi (300 \text{ km})^{-1}$ and $c = 20 \text{ m/s}$ moving through a typical January zonal mean wind $[\bar{u}]$ is considered again. If the influence of the mean meridional wind on the Doppler shift is neglected, i.e. $|l[\bar{v}]| \ll |k[\bar{u}]|$, then the intrinsic frequency $\hat{\omega}$ remains almost unaffected by meridional refraction. But as the cumulative change in l leads to an increase in the horizontal wavenumber of $\Delta k_h \approx 8 \%$, the intrinsic horizontal phase velocity \hat{c}_h is also reduced by 8 %. If the breakdown of the GW is described by saturation, then the corresponding flux of GW pseudo-momentum is given by eq. (5.24), i.e. $F \propto \hat{c}_h^3$ (again assuming the medium-frequency approximation). Compared to the conventional saturation flux F_{conv} , the value of F is reduced due to refraction by a factor of $3\Delta k_h/|k| \approx 24 \%$. Therefore, if the vertical dependence of this additional factor is ignored, the real zonal force f_λ is also diminished by horizontal refraction compared to the force $f_{conv,\lambda}$ calculated within the vertical-column approach, i.e.

$$f_\lambda = f_{conv,\lambda} \left(1 - 3 \frac{\Delta k_h}{k_h} \right). \quad (5.32)$$

The force reduction due to horizontal gradients is mainly a temporal average effect, but it reduces the diurnal GW force as well.

5.4.4. Horizontal refraction in realistic flows

Stationary mean flows

Here, the impact of realistic but stationary mean flows on the propagation of GW rays is investigated. Simplified ray-tracing experiments including the effects of horizontal propagation and refraction analogous to the complex “full” experiment from tab. 5.2 have been performed.

The impact of the zonal-mean flow on meridional propagation and refraction is studied. In climatological January conditions, with $[\bar{u}]$ and $[\bar{v}]$ obtained from HAMMONIA, rays of the four different GW members 9, 10, 12 and 13 (see tab. 5.1), are calculated and plotted in fig. 5.8. In fig. 5.8(a), one sees that the initially eastward propagating GW member 9 has the best conditions in the summer east-wind jet. There, the rays are mainly vertically aligned. In the mesopause region, the summer wind reversal induces a deceleration of vertical GW motion, and the GW rays refract into the meridional direction, i.e. into the tropics. In the winter hemisphere, gravity waves are mainly filtered. The impact of Earth’s curvature on ray propagating can also be seen by comparing the former ray calculations to a simulation without curvature terms (gray rays). Curvature effects are important mainly in high latitudes where great circle propagation leads to equatorward motion of the group rays. As seen for three rays at high northern latitudes, there exists some chance for member 9 to circumvent the west-wind jet. For member 10, which is initially north-eastward directed, the picture is quite similar. Larger meridional displacements of rays appear, with some of the rays initially in the southern hemisphere

Rays in a zonal-mean zonal wind (m/s)

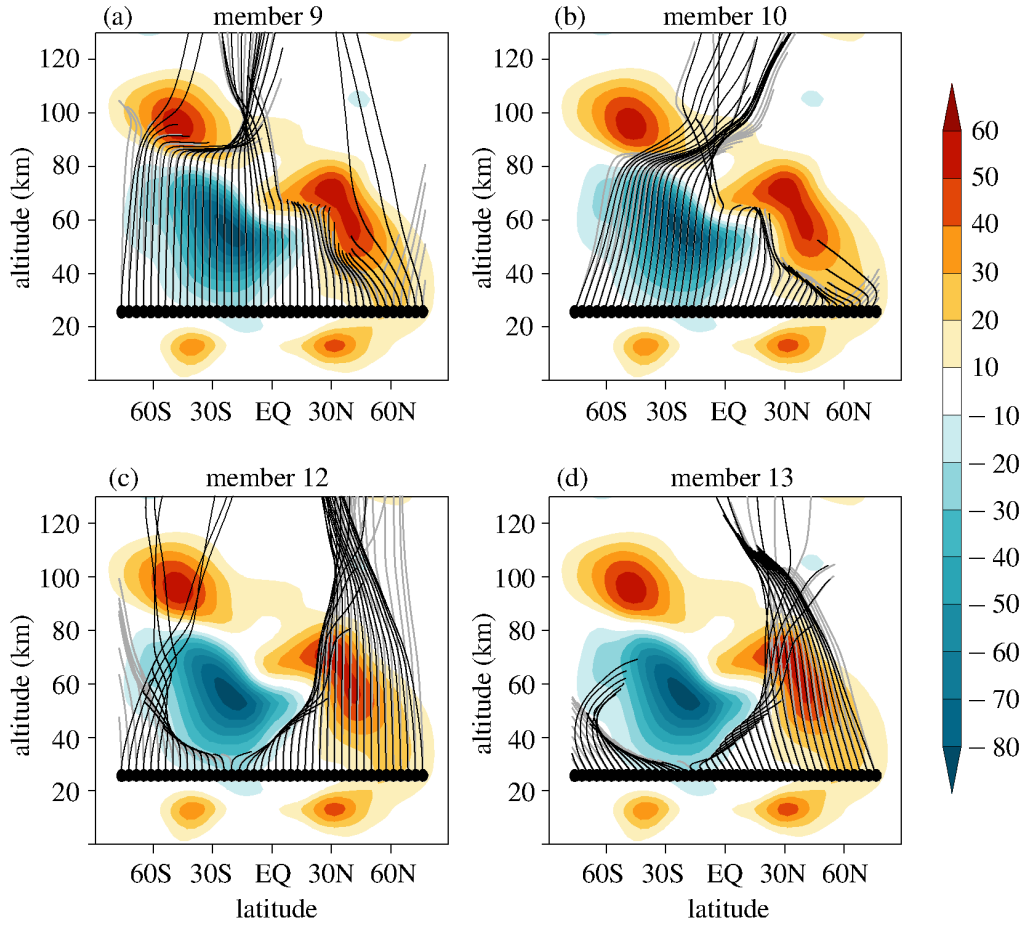


Figure 5.8.: Rays (black lines) of GW members (a) 9, (b) 10, (c) 12 and (d) 13, see tab. 5.1, in a climatological January flow $[\bar{u}]$ (colors). Each ray is started at a meridional grid point and above 20 km. For comparison, ray calculation have been repeated with the curvature terms in (4.109) and (4.110) switched off (gray lines).

reaching the northern mid-latitudes in the lower thermosphere. For members 12 and 13 which are initially west- and south-westward directed, respectively, the situation is almost mirrored, but as the east-wind jet has a larger meridional extent than the west-wind jet, the region of ray propagation in the winter mesosphere is much narrower. Above the middle atmosphere jets, crossing of rays at one or more points happens. These are caustics due to the meridional focusing of GW rays, as mentioned by *Dunkerton* [1984], and indicate that a locally monochromatic description of the gravity waves is not appropriate anymore.

The impact of stationary planetary waves on realistic GW propagation is illustrated in fig. 5.9. Another ray-tracing experiment was performed in the climatological HAMMONIA background, but now including zonal variations. Rays of the initially south-westward directed GW member 13 (see tab. 5.1) are plotted for different launch longitudes and latitudes. The main message of the picture is that GW motion is extremely complex in this wavy background. Rays are strongly curved and change their propagation direction within a small geographical region. Gravity waves from high latitudes have the tendency to move equatorward. Mid-latitude gravity waves behave quite ir-

Horizontal ray propagation in a stationary planetary wave

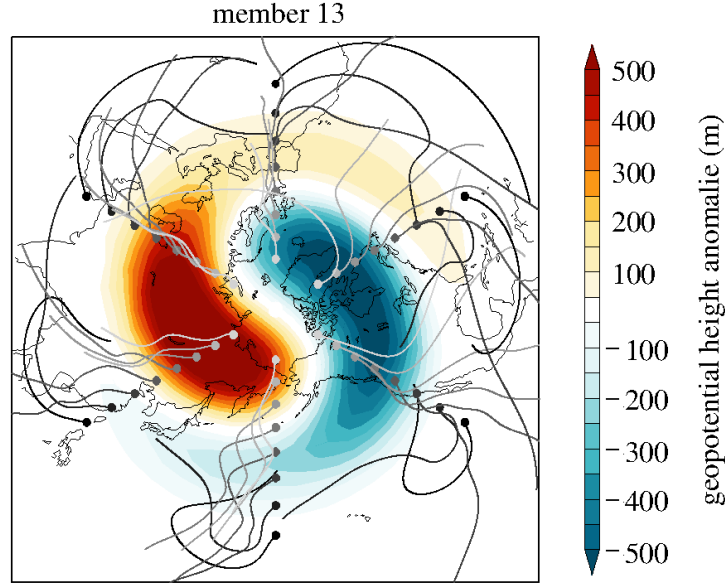


Figure 5.9.: A stationary planetary wave in the stratosphere and its effects on GW rays. The zonal geopotential height disturbance (colors) at 50 km in intervals of 50 m and several group rays of GW member 13 in the northern hemisphere are plotted. The initial ray positions at 6 different longitudes and 8 different latitudes (different shades of gray) are marked by filled circles and their projections on the 50 km surface are shown.

regularly in their meridional movement. A net northward motion seems to occur and is caused by the mean refraction due to the zonally averaged flow (compare to fig. 5.8).

Background with diurnal tides

Next, effects of the mean horizontal refraction in a background flow including diurnal tides are discussed. Results from the “full” experiment are analyzed in that context. As the background conditions are time-dependent, it should be kept in mind that rays and stream lines are not identical. This should be taken into account in the following analysis of GW fields. A collection of several zonally and temporally averaged GW quantities is shown in fig. 5.10 for GW members 9, 10, 12 and 13. The arrows show the mean group velocities $[\bar{c}_{g\varphi}]$ and $100[\bar{c}_{gz}]$, illustrating the mean streaming of GW fields. The shadings indicate the zonally and temporally averaged geographical distribution of the initial latitudinal position the GW field had at the lower boundary $\hat{z}_B = 20$ km where the rays were initialized. This is a passive tracer in the full three-dimensional group flow without any sources and sinks and illustrates the displacement of each part of the GW field. The last quantity overlaid in the plots is the zonally and temporally averaged background wind in the wave direction $[\bar{u}_h]$. Negative contours indicate GW propagation against the wind and positive contours GW propagation with the wind.

GW members 9 and 10 are respectively eastward and north-eastward aligned at the lower boundary $\hat{z}_B = 20$ km. They have favorable propagation conditions in the southern hemisphere stratosphere. In the jet core at about 30S, the mean group-velocity vectors are mainly vertically aligned. At the jet edges, parts of the GW fields are refracted into the jet core. In the summer mesopause region, GW fields are slowed down due

**initial ray latitude (deg), mean group velocity (m/s)
and mean wind in wave direction (m/s)**

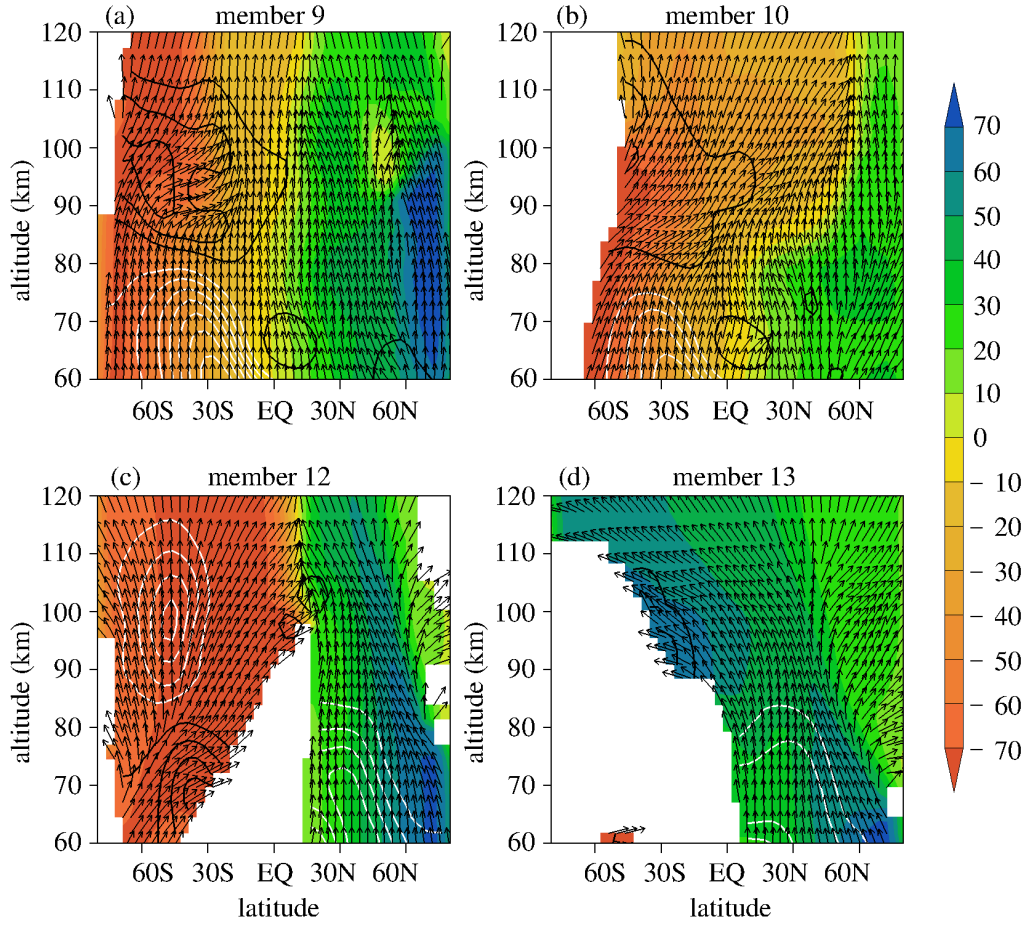


Figure 5.10.: Meridional projection of the temporally and zonally averaged group velocities ($[\bar{c}_{g\varphi}]$, $100[\bar{c}_{gz}]$) (arrows) for several GW ensemble members (a) 9, (b) 10, (c) 12 and (d) 13. The horizontal background wind in the wave direction $[\bar{u}_h]$ is plotted in contours with an interval of 10 m/s. Positive contours are black and negative white. The initial meridional position of the GW field at the lower boundary $\hat{z}_B = 20$ km, where the rays have been initialized, is shown using shading with an interval of 10.

to reversing background winds. The GW fields avoid the positive jet core and are refracted into mid-latitudes. Especially for GW member 10, this refraction leads to an irreversible growth of the meridional wavenumber above 90 km and to large meridional displacements. In the northern hemisphere, the GW fields can propagate vertically through the minima of planetary wave structures.

GW members 12 and 13 are respectively westward and south-westward aligned at the lower boundary. The westerly-wind vortex of the northern winter hemisphere provides most favorable propagation conditions. In the jet core, the group velocities are mainly vertical. At the wind reversal, GW fields are refracted in the meridional direction. For GW member 13, the mean latitude positions are interchanged in the lower thermosphere. Parts of the GW field initially in the northern mid-latitudes have moved southward to the equatorial region and even into the southern hemisphere, whereas parts of the GW field initially from the subtropics have propagated northwards. As discussed before,

due to the modulation of stratospheric winds by planetary waves, zonally dependent wave guides can develop. The easterly-wind jet in the southern hemisphere to a large extent prohibits propagation of GW member 12 and 13. Interestingly, some chance exists for parts of the high-latitude GW field to circumvent the jet core (GW member 12). Above the critical jet, GW fields are refracted southward and spread over a large horizontal domain. The considerable horizontal expansion of GW fields, as seen e.g. for GW member 12 at 100 km and between 80S and 10N, as well as for GW member 13 above 110 km and between 80S and 30N, also influences the amplitudes of the GW fields via eq. (4.112). The corresponding change in GW amplitudes is not incorporated in most previous ray-tracing work [Marks and Eckermann, 1995; Hasha et al., 2008; Song and Chun, 2008] which commonly apply the assumption of a constant vertical flux $F_A = c_{gz}A$ of wave action density A .

The results shown in fig. 5.10 are compared to the ray calculation in the zonally averaged flow shown in fig. 5.8. Clearly, the domain where GW motion is possible is greatly widened. Temporal and zonal variations in the background conditions lead to a smooth mean distribution of GW fields. Accentuated “shadow” regions beyond which no ray motion is possible, e.g. the “shadow” of the east-wind jet in figs. 5.8(c) and 5.8(d), tend to shrink. The increased variability in the background conditions causes a stronger mixing of GW rays. A comparison of ray-tracing calculations with the geometric effects switched off illustrates that geometric refraction can lead to reduced poleward GW motion and increases the probability of the intrusion of high-latitude parts of the GW field into lower latitudes.

The temporally averaged field of the horizontal wavenumber vector at 80 km is given in fig 5.11. The initially eastward directed GW member 12 and the initially westward directed GW member 9 are plotted in the northern and southern hemisphere, respectively. Around 30S and 30N, each of the corresponding GW members are roughly zonally aligned. In the tropics/subtropics, the horizontal wavenumber vectors \mathbf{k}_h are refracted poleward, and in the mid- to high latitudes \mathbf{k}_h points equatorward.

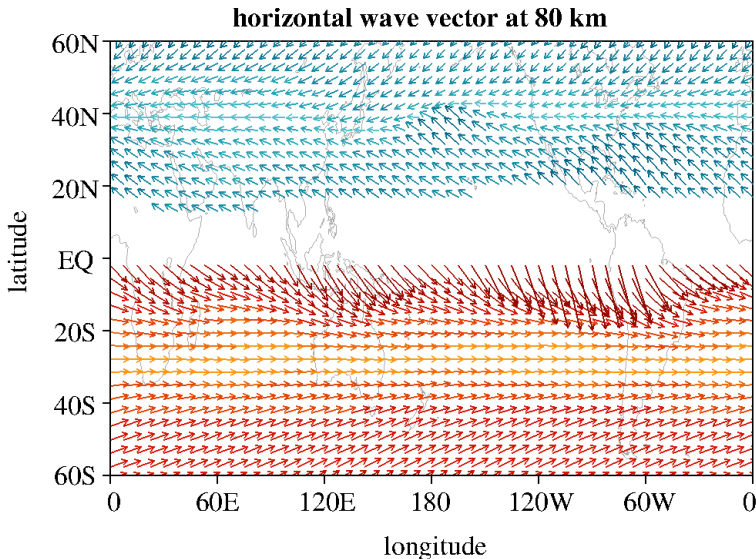


Figure 5.11.: The temporally averaged horizontal wavenumber vector at 80 km for GW members 9 (red arrows) and 12 (blue arrows) plotted only until the equator. Darker colors indicate larger magnitudes of the meridional wavenumber.

section 5.4.2, it is mainly the meridional gradients in $[\bar{u}]$ that are responsible for that behavior.

In the following, the statistical nature of the meridional GW motion is investigated. The relative occurrence of meridional displacements of the four GW members is plotted in fig. 5.12. GW fields appearing in the altitude range from 60 to 120 km and covering one daily cycle have been analyzed globally. GW members 9 and 12, which are zonally

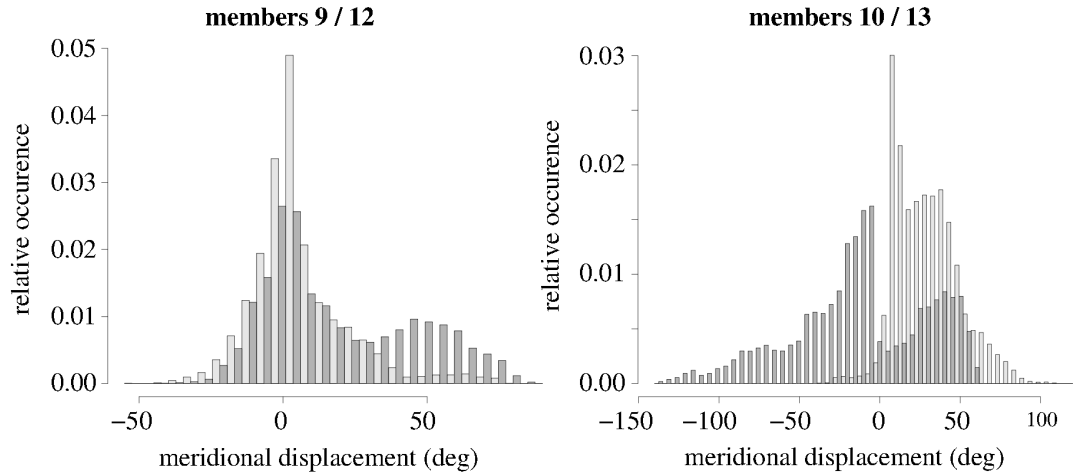


Figure 5.12.: Distribution of meridional displacements for GW members 9 (left, light gray), 12 (left, dark gray), 10 (right, light gray) and 13 (right, dark gray). The bin size is 5 with each two neighboring bars sharing the same bin as in fig. 5.4.

aligned at the lower boundary \hat{z}_B , have broad distributions of meridional displacements with the most probable values around zero. Displacements larger than 30 appear, and a secondary maximum arises for GW member 12 at about 50. This is connected to the intrusion of polar gravity waves into lower latitudes in the southern hemisphere (see also fig. 5.8(c) and fig. 5.10(c)). GW members 10 and 13 are initially counter-propagating pairs with north-eastward and south-westward direction. Because they initially have a preferential meridional direction, they also show larger displacements than GW members 9 and 12. Essentially, the distributions are mainly one-sided with a maximum value at zero and a slow decay in the direction of preferred propagation. For the main branches of the distributions, median displacements of 26 and -27 occur for members 10 and 13, respectively. Surprisingly, displacements larger than 100 are also possible which shows the possibility of interhemispheric GW propagation. Additionally, a secondary maximum appears for GW member 13 at about 50. By inspection of fig. 5.8(d) and fig. 5.10(d), this can be attributed to the poleward intrusion of subtropical GW fields into the northern hemisphere.

5.5. Gravity-wave forces on the tide

5.5.1. Mean gravity-wave forces

Before investigating the periodic GW forces, which are one major focus of this study, changes in the temporal mean GW force are inspected. As discussed by *Andrews et al.* [1987], the relevant GW forcing of the mean flow, in the present case the temporally averaged flow plus diurnal tides, is given by the divergence of the GW pseudo-momentum flux rather than the GW momentum flux itself. The main difference between the two arises for slowly vertically propagating inertia-gravity waves. These waves produce a

Stokes drift counterbalanced by an Eulerian-mean flow locally attached to the waves [Bühler, 2009]. Hence, some parts of the force inferred from the divergence of the momentum flux are needed to sustain the local Eulerian circulation and do not change the background conditions. To avoid these artificial flow phenomena, the average flow should be interpreted as Lagrangian-average motion for which the corresponding forces were derived in section 2.2.3. The vertical flux of zonal pseudo-momentum is [Fritts and Alexander, 2003]

$$F = \rho_r \langle \zeta \partial_x \phi' \rangle_w = \hat{c}_{gz} k A = \rho_r \langle u' w' \rangle_w \left(1 - \frac{f^2}{\hat{\omega}^2} \right), \quad (5.33)$$

where here ζ is the vertical displacement of the fluid by GW motion. Again the primes denote GW perturbations averaged over reasonable scales via the bracket operator, and the equal signs hold only in the WKB limit (as discussed in section 4.3.3 and calculated in appendix A.3). Therefore, the wave stress on the (Lagrangian) mean flow is reduced by a factor of $(1 - f^2/\omega^2)$.

In neglecting horizontal variations in the GW fields, the horizontal force due to GW stresses is [Fritts and Alexander, 2003]

$$\mathbf{f}_{gw,h}^L \approx -\frac{1}{\rho_r} \partial_z (\hat{c}_{gz} \mathbf{k}_h A), \quad (5.34)$$

but the impact of horizontal inhomogeneities in the background conditions on the diurnal GW force are of special interest here. Hence, the more complete form (2.72) which includes the full divergence of the flux tensor of GW pseudo-momentum [see Grimshaw, 1975a]

$$\mathbf{f}_{gw,h}^L = -\frac{1}{\rho_r} \nabla \cdot (\hat{c}_g \mathbf{k}_h A) \quad : \text{Lagrangian-mean horizontal GW force} \quad (5.35)$$

is used. Compared to eq. (2.72) in section 2.2.3, the work against the frictional stresses is omitted as wave motion under large wave Reynolds number conditions is assumed.

In the following, an ensemble mean force is calculated by an arithmetic average over all 14 GW members introduced in section 5.2.1. The temporally and zonally averaged zonal GW force is shown in fig. 5.13. Three experiments, (a) “full”, (b) “noREF” and (c) “TS”, are compared. In the mesopause region at about 80 km to 85 km, the typical dipole structure is visible, with a negative forcing peak in winter and a positive peak in summer. In the “TS” simulation, the peak values are about -55 and 48 m/s per day which is a factor 1.5 to 2 smaller than typical GW forcing values in realistic GCM simulations [compare e.g. Alexander et al., 2010; Richter et al., 2010]. Besides this deficit, it is instructive to quantify the impact of tidal time-dependence and mean horizontal gradients on the time-mean GW force using the 3 experiments. There is no significant change between the GW forces in the “TS”-experiment (fig. 5.13(c)) and the “noREF”- experiment (fig. 5.13(b)) when focusing on the mesopause region. Hence, the frequency modulation has no significant impact on the mean force there. On the other hand, the GW force is diminished in the “full”-simulation with values of about -40 and 36 m/s per day due to the impact of horizontal refraction. This is a reduction of about 17 % and 35 % in the southern and northern hemispheres, respectively, and, as explained in section 5.4.3, a temporal mean effect resulting from a cumulative change in the horizontal wavenumber k_h . For the “full”-simulation, force-weighted hemispheric averages [Preusse et al., 2009] of $|\Delta k|/k_h$, $|\Delta l|/k_h$ and $|\Delta k_h|/k_h$ were calculated with

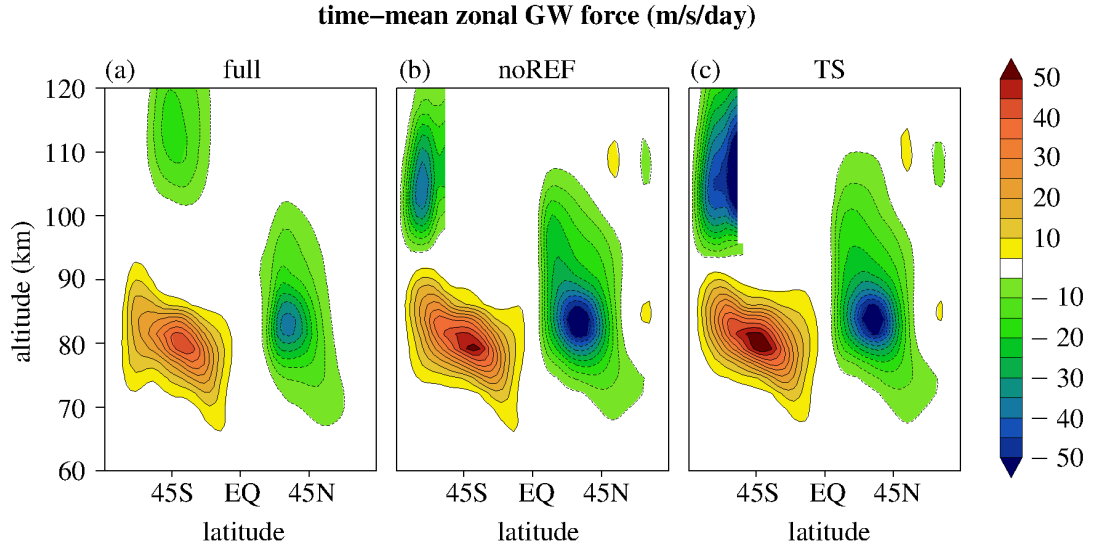


Figure 5.13.: Temporally and zonally averaged zonal GW force f_λ in the (a) “full”, (b) “noREF” and (c) “TS” experiment. Contour interval is $5 \text{ ms}^{-1} \text{ day}^{-1}$ and negative values represented by dashed lines.

northern hemispheric values of 47 %, 11 % and 12 % respectively. There, changes in l seem to dominate changes in k_h . Using eq. (5.32), the increase in k_h can explain a reduction of the temporal mean force of about 36 %. In the southern hemisphere, however, the situation is not such as clear.

Different ray simulations excluding either tides or planetary waves were also performed (not shown). For the reduction of the temporally mean force, equivalent values were found even if temporal and zonal variations of the background flow were excluded. Furthermore, runs with tides only were performed but no clear impact of the tidal gradients on the GW saturation was found.

5.5.2. Periodic forces due to wave stresses

Since GW fields in the MLT region are periodically modulated by tidal winds, they produce a periodic force acting back on the diurnal tides. For instance, consider the zonal force $f_\lambda = \mathbf{f}_{gw,h}^L \cdot \mathbf{e}_\lambda$ given in fig. 5.14. The temporal behavior of f_λ at two specific geographic locations, $\lambda = 60\text{E}$, $\varphi = 30\text{S}$ and $\lambda = 60\text{E}$, $\varphi = 30\text{N}$, was taken from the “full” experiment. In the southern hemisphere, eastward propagating gravity waves lead to a positive pseudo-momentum deposition and thus a positive force at about 80 km. In the southern lower thermosphere, westward moving gravity waves which circumvent the east-wind jet (see fig. 5.8(c) and fig. 5.10(c)) induce a negative force. In the northern hemisphere, westward propagating gravity waves deposit their negative pseudo-momentum between 80 km and 100 km. The temporal mean force is strongly disturbed by the impact of tides on GW saturation. In tidal phases where the background wind becomes positive, eastward propagating gravity waves are destabilized. The vertical GW wavelength and the volume occupied by the wave field shrinks. Due to the first effect, the GW fields are more sensitive to turbulent diffusion, the latter induces an additional increase in GW amplitudes. Breakdown of the gravity waves results below the maxima of the tidal wind. For westward moving gravity waves the opposite holds. The tidal winds cause an undulation of the total GW pseudo-momentum deposition in which the diurnal GW force component can be of the order of the temporal

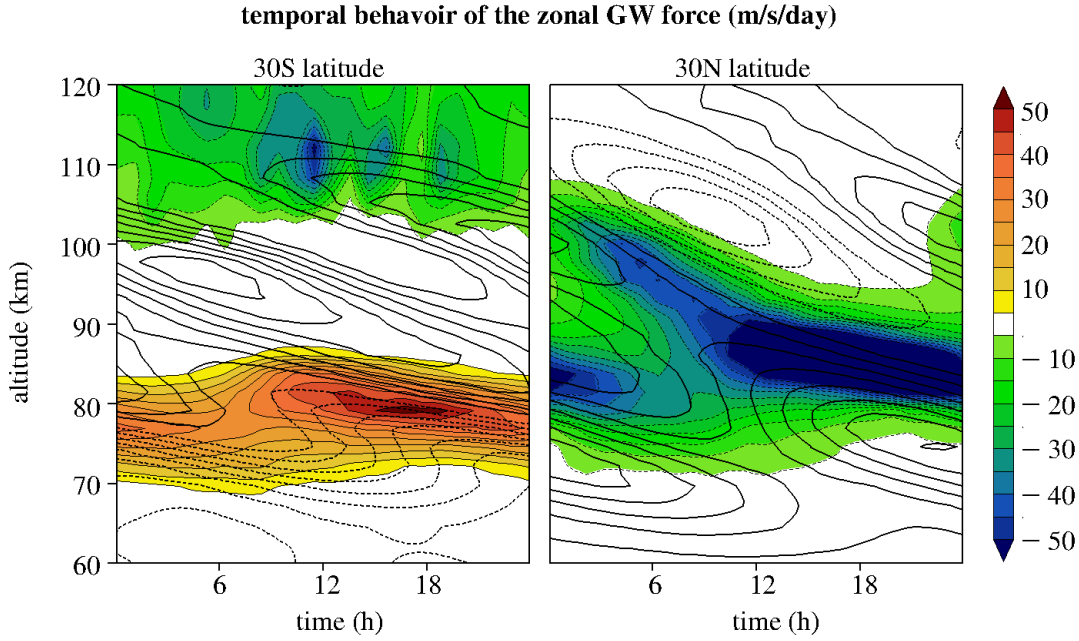


Figure 5.14.: Temporal behavior of the zonal GW force f_λ (color) from the “full” experiment at $\lambda = 60$ E and $\varphi = 30$ S (left) and $\varphi = 30$ N (right) in intervals of $5 \text{ ms}^{-1} \text{ day}^{-1}$. The zonal background wind is plotted in contours with an interval of 10 m/s.

mean force. To summarize, a net negative force results when the tidal wind enters its negative phase, and a net positive force emerges when the tidal wind enters its positive phase. This suggests that the diurnal GW force is positively correlated with the local tidal acceleration. The effect is discussed more quantitatively in section 5.5.4, where equivalent Rayleigh friction coefficients are introduced.

The zonally-averaged amplitude of the diurnal zonal force is calculated (compare to eq. (3.4)) and shown in fig. 5.15 for the three experiments “full”, “noREF” and “TS”. The results from the “TS” experiment, fig. 5.15(c), which mimics the effect of a conventional Lindzen GW parameterization, are chosen as a basis for comparison. In the northern winter hemisphere a pronounced subtropical maximum with a peak force of 40 m/s per day can be seen whereas in the southern summer hemisphere and high latitudes the forcing peaks at about 80 m/s per day. The overall structure of the forcing amplitudes of the “TS” experiment compares quite well to past investigations by several authors with very different assumptions on GW source parameters, even though a highly simplified GW ensemble is used here. For instance, *Miyahara and Forbes* [1991] presented in their fig. 6 a GW forcing field with a strong subtropical winter maximum at 88 km of about 100 m/s per day and a weak subtropical summer maximum. Also, the strong high-latitude forcing is at about 110 km. The asymmetry between subtropical winter and summer maxima at about 90 km was even stronger in the work of *Meyer* [1999] (his fig. 10) with peaks at 60 m/s per day (winter) and 10 m/s per day (summer). *Ortland and Alexander* [2006] (their fig. 6-10) found symmetric configurations with subtropical peaks between 90 and 100 km in the range of 40 to 60 m/s per day but used equinox conditions for the time mean background. In a non-linear simulation with resolved gravity waves, *Watanabe and Miyahara* [2009] (their fig. 9) found lower peaks at about 70 km, also with a more pronounced subtropical winter maximum of 30 m/s per day. Thus, at least the amplitudes of the zonal forces of the present study are in the range of past investigations. Note that all of the mentioned studies focused only on amplitudes

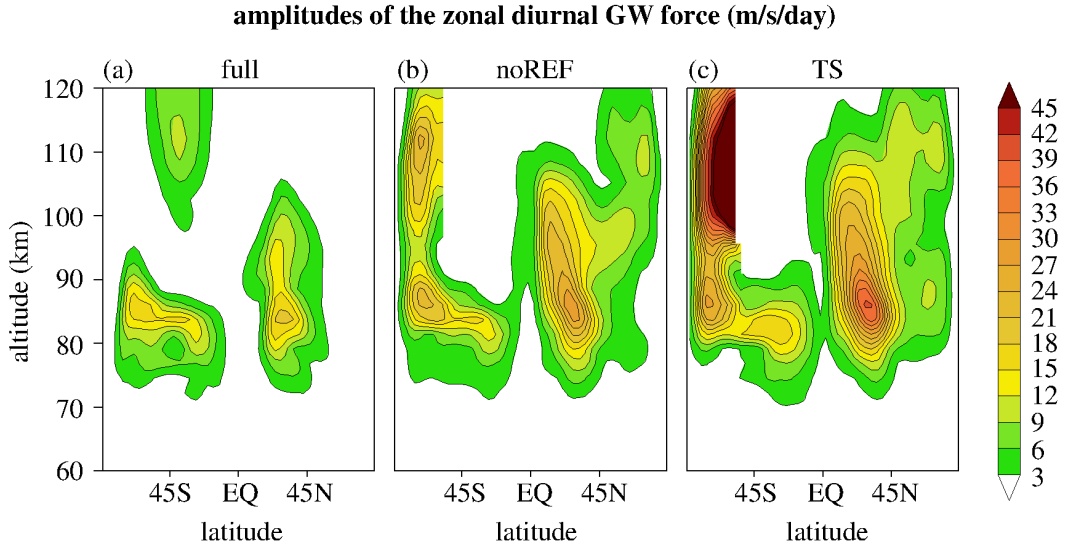


Figure 5.15.: Amplitudes of the zonal diurnal GW force f_λ in $\text{ms}^{-1} \text{day}^{-1}$ for the (a) “full”, (b) “noREF” and (c) “TS” experiments. Contour interval is $3 \text{ms}^{-1} \text{day}^{-1}$.

of the diurnal migrating components of the forces whereas a composite of migrating and non-migrating forcing components is shown here.

For the “noREF” experiment (fig. 5.15(b)), the total forcing amplitude is decreased. The northern maximum is reduced to 28 m/s per day which is about 30% less than the “TS” value. In the southern hemisphere, dramatic differences arise in the high-latitude maximum with a reduction of up to 70%. Relations (5.5) and (5.28) are used to roughly estimate this frequency-induced reduction. When averaged globally over the altitude range of 60 km to 120 km, the mean tidal vertical phase velocity is $C \approx -0.34 \text{ m/s}$, and when additionally averaged over the GW ensemble, the mean GW vertical group velocity is $c_{gz,0} \approx 0.52 \text{ m/s}$. Eqns. (5.5) and (5.28) imply a reduction of 40% in the average diurnal force. Considering the crudeness of the assumptions made, this estimate is quite good.

The forcing amplitudes are further decreased in the “full” experiment (fig. 5.15(a)). The asymmetry between the subtropical winter and summer maxima is reduced. Peak winter values close to 20 m/s per day are only 50% of the “TS” winter maximum. In addition, the high-latitude peak at about 110 km has moved equatorward to 30S. Its value is reduced to about 90% of the conventional estimate. In high latitudes, the forcing is diminished due to the effect of mean horizontal refraction. The globally and spectrally averaged reduction of the diurnal, zonal GW force is 65% for the “full” experiment compared to the “TS” experiment.

The amplitude of the meridional diurnal force is shown in fig. 5.16. Its structure is similar to the zonal forcing shown in fig. 5.15, but its maxima reach only about 40% of f_λ , likely the result of the anisotropy of the chosen GW ensemble. With reference to satellite observations, it was discussed by *Lieberman et al.* [2010] that the meridional GW force can be as much as twice as large as the zonal GW force. This feature is not reproduced with the simple toy GW ensemble. In the “noREF” and “full” experiments (fig. 5.16(a) and 5.16(b)) the winter peak is reduced 30% of its “TS” value. The subtropical summer forcing changes slightly between the three experiments. Interestingly, the high-latitude summer peak at about 60S and 110 km in the “TS” simulation totally vanishes in the “noREF” experiment and reappears in the “full” simulation at 30S. This indicates that, owing to horizontal refraction, initially zonally aligned parts of the GW fields turn in the

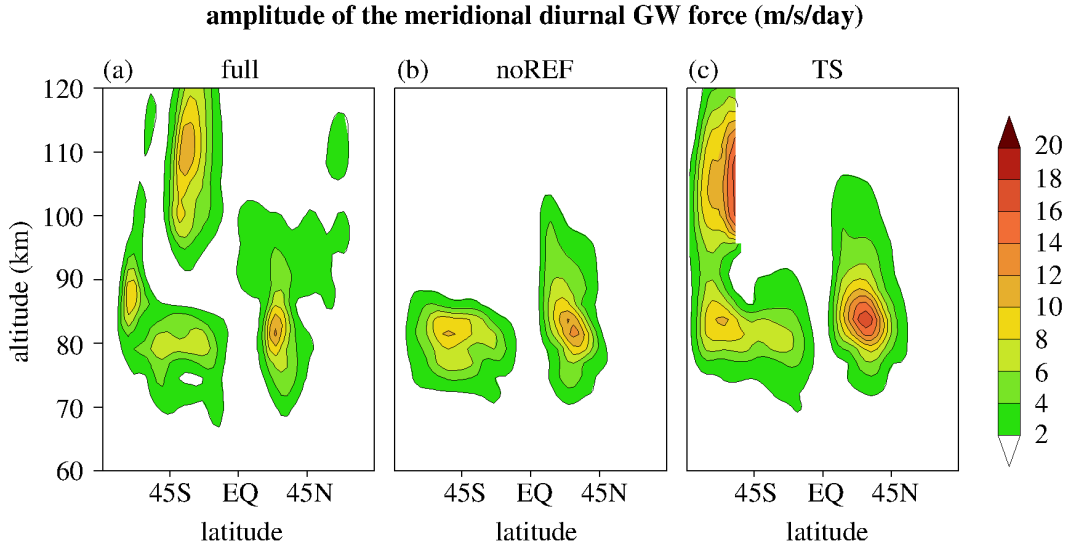


Figure 5.16.: Same as fig.5.15, but for the amplitude of the meridional GW force on the diurnal tide with a contour interval of $2 \text{ ms}^{-1} \text{ day}^{-1}$.

meridional direction and dominate the meridional forcing amplitude in some regions.

5.5.3. Direct effects of horizontal refraction on the diurnal forcing

As discussed by *Preusse et al.* [2009], horizontal refraction of vertically propagating gravity waves may lead to changes in the induced forcing. In the following, the complete GW forcing (5.35) is divided into contributions connected to horizontal inhomogeneities of the pseudo-momentum flux tensor, turning and stretching of the horizontal wavenumber vector, and dissipation of wave action. The last is the dominant contribution, but the corrections due to the others are of special interest here. First, the zonal force can be divided into horizontal and vertical divergence parts

$$\mathbf{f}_{gw,h}^L \cdot \mathbf{e}_\lambda = f_\lambda = f_{\lambda,hor} + f_{\lambda,vert}, \quad (5.36)$$

where

$$f_{\lambda,hor} = -\frac{1}{\rho_r} \left\{ \frac{\partial_\lambda}{a_E \cos \varphi} (\hat{c}_{g\lambda} k A) + \frac{\partial_\varphi}{a_E} (\hat{c}_{g\varphi} k A) - \frac{\tan \varphi}{a_E} (\hat{c}_{g\lambda} l + \hat{c}_{g\varphi} k) A \right\} \quad (5.37)$$

and

$$f_{\lambda,vert} = -\frac{1}{\rho_r} \partial_z (\hat{c}_{gz} k A). \quad (5.38)$$

Note that metric terms arise because the terms $\mathbf{e}_\lambda \mathbf{e}_\varphi$, etc. in the representation of the pseudo-momentum tensor $\hat{\mathbf{c}}_g \mathbf{k}_h A$ depend on the coordinate system.

Usually, only $f_{\lambda,vert}$ is considered as forcing [*Fritts and Alexander, 2003; Preusse et al., 2009*]. This term is divided into the different contributions mentioned above. With the azimuth angle of GW direction in the horizontal plane α and writing $k = k_h \cos \alpha$, the decomposition

$$f_{\lambda,vert} = f_{\lambda,turn} + f_{\lambda,str} + f_{\lambda,diss} \quad (5.39)$$

is defined, where force contributions owing to horizontal turning of gravity waves

$$f_{\lambda,turn} = -\frac{\partial_z(\cos \alpha)}{\rho_r} k_h \hat{c}_{gz} A, \quad (5.40)$$

horizontal stretching of the GW field

$$f_{\lambda,str} = -\frac{\partial_z(k_h)}{\rho_r} \cos \alpha \hat{c}_{gz} A \left(1 + k_h \partial_{k_h}(\ln \hat{c}_{gz}) \right), \quad (5.41)$$

and convergence of vertical flux of wave action

$$f_{\lambda,diss} = -\frac{\partial_z(\hat{c}_{gz} A) |_{k_h}}{\rho_r} k_h \cos \alpha \quad (5.42)$$

appear.

The first force $f_{\lambda,turn}$ gives a local forcing contribution when the horizontal wavenumber vector \mathbf{k}_h turns with height due to refraction induced by horizontal gradients of the background wind and by geometric effects. *Preusse et al.* [2009] calculated the zonally and temporally averaged turning force utilizing their GW source spectrum. They found that this term is mostly of second order with magnitudes not larger than 5% of the dominant contribution.

The second force $f_{\lambda,str}$ arises due to changes in the horizontal wavenumber and results from horizontal stretching or shrinking of GW structures due to horizontal inhomogeneities of the background conditions. The stretching force includes contributions associated with (i) changes in the magnitude of the horizontal pseudo-momentum $k_h A$ and (ii) vertical changes in the vertical group velocity \hat{c}_{gz} only due to refraction. The latter may be understood as refraction-induced variations in the volume occupied by the GW fields. The factor $k_h \partial_{k_h}(\ln \hat{c}_{gz})$ varies between -2 for the reflection limit $|m|/k_h \rightarrow 0$ and 2 for the critical layer limit $|m|/k_h \rightarrow \infty$. Therefore, by the action of GW volume compression, the force due to horizontal contraction of the GW field can be enhanced by a factor of 3! To the knowledge of the author, the magnitude of this term has not been considered separately in any published literature such as *Preusse et al.* [2009]. The last force $f_{\lambda,diss}$ gives the convergence of the wave action flux not connected to horizontal refraction. This forcing contribution arises when GW fields dissipate.

As before, diurnal amplitudes of the several forces are shown. Fig. 5.17 gives an overview of magnitudes of the various force contributions. Obviously, the force due to dissipation of vertical wave action flux $f_{\lambda,diss}$ in fig. 5.17(d) is dominant for the interaction between gravity waves and thermal tides, with major contributions in the whole domain. The force due to the horizontal divergence of the flux of zonal pseudo-momentum (fig. 5.17(a)) has two maxima in the mesopause region with peaks of about 1 m/s per day and an additional peak in the southern thermosphere of about 2 m/s per day. The forcing due to turning of \mathbf{k}_h (fig. 5.17(b)) peaks in the subtropical winter hemisphere at about 80 km and in the subtropical summer hemisphere at about 105 km with a peak value of 1 m/s per day. The force due to stretching (fig. 5.17(c)) is about two to three times larger than the turning force. The subtropical winter maximum peaks at a surprisingly high 5 m/s per day. Hence, the horizontal deformation of the GW field seems to be more important than realized in the past.

Horizontal averages of the amplitudes of $f_{\lambda,hor}$, $f_{\lambda,turn}$, $f_{\lambda,str}$ and $f_{\lambda,diss}$ relative to f_{λ} are plotted in fig. 5.18. In the mesosphere, the convergence of the wave action flux contributes about 150% of the full forcing amplitude. This overestimation is reduced

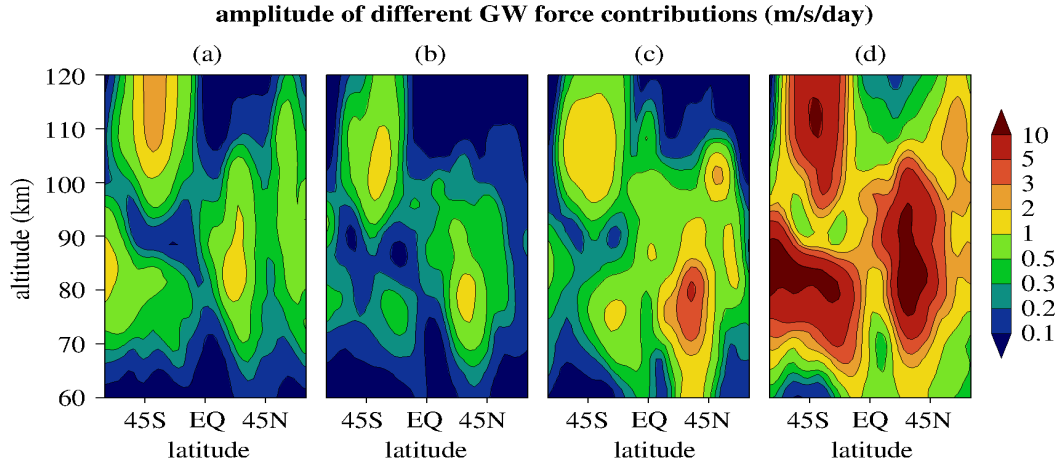


Figure 5.17.: Diurnal amplitudes of different zonal force contributions: (a) $f_{\lambda,hor}$, (b) $f_{\lambda,turn}$, (c) $f_{\lambda,str}$ and (d) $f_{\lambda,diss}$. A non-linear color scale has been used with contours at 0.1, 0.2, 0.3, 0.5, 1, 2, 3, 5 and 10 in m/s per day.

mainly by the effect of GW field stretching which contributes around 60% below 75 km. It is found that in this region the GW horizontal wavelengths shrink with altitude, i.e. k_h increases. The growth in k_h partially compensates the dissipative reduction in wave action flux. The process is related to the phenomenon known as wave capture [Bühler and McIntyre, 2005; Bühler, 2009]. However, the induced mean flow and its interaction with the GW fields are not considered here. The relative contributions due to the horizontal divergence of zonal pseudo-momentum flux and the horizontal turning of the wavenumber vector range between 10% and 25%.

As discussed, changes in \mathbf{k}_h due to background flow inhomogeneities induce direct forces acting locally on the background flow, but the effect of permanent changes in the GW field is much larger than local effects. Changes accumulate as parts of the wave field propagate through the middle atmosphere. This can be seen by the large difference between the forcing amplitude derived from the different simulations, i.e. when comparing “full” and “noREF” experiments as in fig. 5.15. These differences are not fully explained by taking only local force contributions due to horizontal refraction into account. The cumulative changes in the wavenumbers are important.

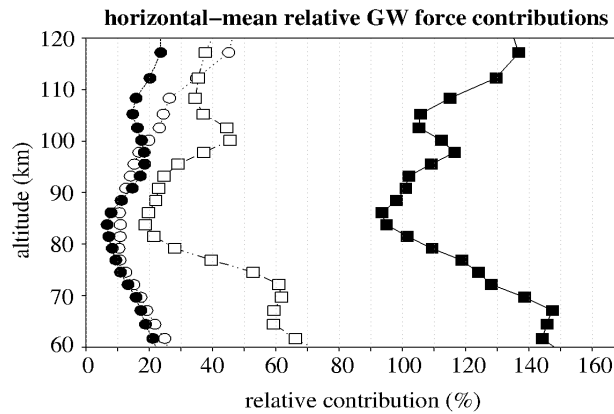


Figure 5.18.: Horizontal average between 60S and 60N of the contributions of $f_{\lambda,hor}$ (open circles), $f_{\lambda,turn}$ (filled circles), $f_{\lambda,str}$ (open squares) and $f_{\lambda,diss}$ (filled squares) relative to the full zonal force of fig. 5.15(a).

5.5.4. Equivalent Rayleigh friction coefficients

Equivalent Rayleigh friction coefficients (ERFs) have been introduced in the context of GW-tide interaction by *Miyahara and Forbes* [1991]; *Forbes et al.* [1991] and further discussed by *McLandress* [2002]. With the help of ERFs, the effects of gravity waves can be incorporated into a linear tidal model [*Miyahara and Forbes*, 1991; *Forbes et al.*, 1991; *Ortland*, 2005b]. However, the concept also has diagnostic value for non-linear simulations with parametrized gravity waves [*McLandress*, 2002] as well as simulations with resolved gravity waves [*Watanabe and Miyahara*, 2009].

Previous studies mainly focused on the GW effect on the migrating tidal components whereas non-migrating parts have been ignored. Here both effects are discussed in a zonally averaged manner. With the definition of the real part γ_R and imaginary part γ_I of the ERFs

$$\gamma_R = -U^{-2} [\overline{f_\lambda u_T}] , \quad (5.43)$$

$$\gamma_I = -\Omega^{-1} U^{-2} [\overline{f_\lambda \partial_t u_T}] , \quad (5.44)$$

the diurnal force is approximated by

$$f_\lambda \approx -\gamma_R u_T - \frac{\gamma_I}{\Omega} \partial_t u_T , \quad (5.45)$$

where again the brackets and overbar denote zonal and temporal average, respectively. u_T is the zonal diurnal wind and U denotes the zonally-average diurnal wind amplitude (see eq. (3.4)).

Note that $[\overline{f_\lambda u_T}]$ is the average tidal kinetic energy tendency induced by zonal GW forcing. Since $[\overline{f_\lambda u_T}] < 0$ is equivalent to $\gamma_R > 0$, positive real parts of the ERFs indicate regions of decrease in tidal kinetic energy and therefore damping of the tides. The imaginary part of the ERF acts on the tidal phase structure. For $\gamma_I < 0$, a decrease in tidal vertical wavelength is observed (see discussion by *McLandress* [2002]; *Ortland* [2005a]). A reduction of tidal vertical wavelength is seen as a very robust result in previous investigations, whereas the GW effect on tidal amplitudes is controversial [*Ortland and Alexander*, 2006, and references therein].

The real parts of the ERFs are shown in fig. 5.19 for the “full”, “noREF” and “TS” simulations. For the reference simulation “TS” in fig. 5.19(c), large positive peaks up to $60 \times 10^{-6} s^{-1}$ occur. The maxima correspond to values of 2 to 5 per day which are a factor of 3 to 5 larger than values reported by *Forbes et al.* [1991] and *McLandress* [2002], but in line with *Miyahara and Forbes* [1991].

Fig. 5.19(c) shows the typical way a Lindzen saturation parameterization acts on γ_R . Consider, for instance, a vertical profile at 45S. Negative values of γ_R are encountered below about 78 km and positive values above that altitude. The mean onset of GW breaking is around 75 km where for this profile the small negative peak appears. Negative γ_R can lead to an increase in tidal amplitudes which is typical for gravity waves which approach their critical levels [*Lu and Fritts*, 1993; *Mayr et al.*, 1999]. This also holds if the onset of convective instability of breaking gravity waves is dominated by tidal winds. In altitudes above the onset of breaking, the saturation is controlled rather by the density decrease than by the increase in tidal wind amplitudes. This is explained with the help of the simple example of section 5.3.2 in the limit $c_0 \gg U$ and eq. (5.27). Assuming additionally that tidal amplitudes slowly increase in the vertical with $H_U = (\partial_z \ln U)^{-1} > 0$, the real part of the ERF is [*Lu and Fritts*, 1993]

$$\gamma_{R,conv} = \frac{3c_0^2 k_0}{N} \left(\frac{1}{H_\rho} - \frac{1}{H_U} \right) . \quad (5.46)$$

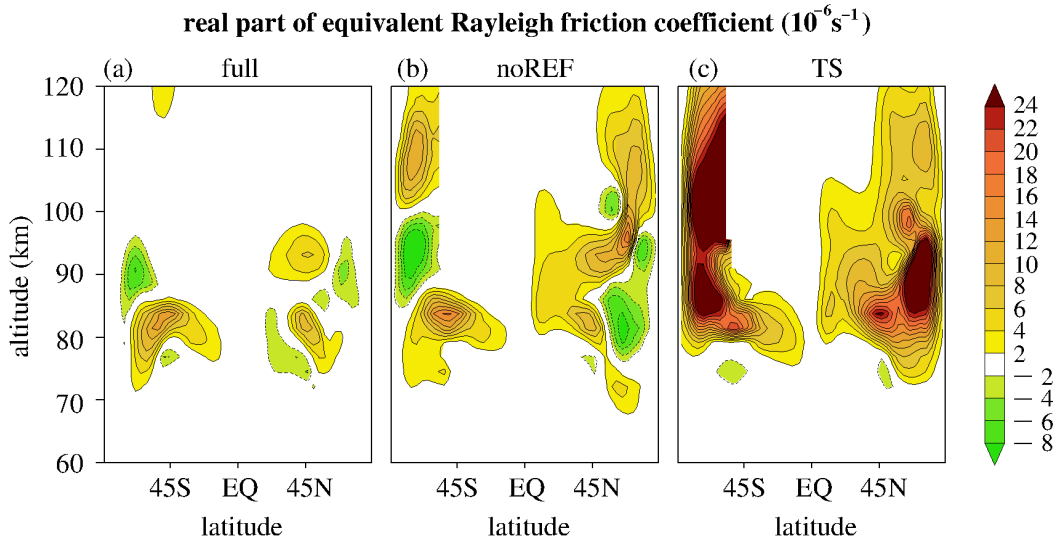


Figure 5.19.: Real part of the equivalent Rayleigh friction coefficient of the zonal tidal wind for the (a) “full”, (b) “noREF” and (c) “TS” experiment in $10^{-6}s^{-1}$ with an interval of $2 \times 10^{-6}s^{-1}$. Negative values are shaded.

Clearly, density and wind changes act in complementary ways on the GW saturation. Positive values of γ_R are consistently obtained above the onset of GW breaking. This is in line with sensitivity studies reported by *McLandress* [1997] and contributes to the discussion raised by *Akmaev* [2001] on the effect of Lindzen-type saturations.

For the “noREF” experiment (fig. 5.19(b)), the magnitude of γ_R is reduced. The latitude-altitude structure is wave-like with a vertical wavelength comparable to the tidal wavelength. In fig. 5.19(a), the magnitude of the ERF is further reduced. Compared to the “noREF” experiment, the influence of γ_R is drastically lowered in high latitudes and in the thermosphere. Fig. 5.19(a) corresponds surprisingly well to the non-linear simulation with resolved gravity waves by *Watanabe and Miyahara* [2009], even though here a simple GW ensemble was used and no feedback between gravity waves and tides was taken into account.

The imaginary parts γ_I of the ERFs are shown in fig. 5.20. For the “TS” experiment in fig. 5.20(c), two distinct negative peaks appear at 45N, 85 km and 60S, 100 km with respective magnitudes around -27 and -60 in units of $10^{-6}s^{-1}$. In the “noREF” experiment, the mid-latitude winter maximum is reduced to 40%. In the summer mid-latitude mesopause region there is an additional negative peak of about $-25 \times 10^{-6}s^{-1}$. The dominant thermospheric negative maximum in the “TS” simulation has changed its sign in the “noREF” simulation. Finally, in fig. 5.20(a), γ_I for the “full” ray-tracing experiment is shown. For γ_I from the “full” experiment any significant impact on the lower thermosphere has disappeared. Two negative maxima in the mid-latitude mesopause region are present with peaks around -20 and -15 in $10^{-6}s^{-1}$ in the southern and northern hemispheres, respectively.

Horizontal averages of γ_R and γ_I are comparable in magnitude to the values published by *McLandress* [2002]. For the “full” simulation, one distinct peak appears between 80 and 85 km in both γ_R and γ_I with respective values of 3 and -7 in $10^{-6}s^{-1}$.

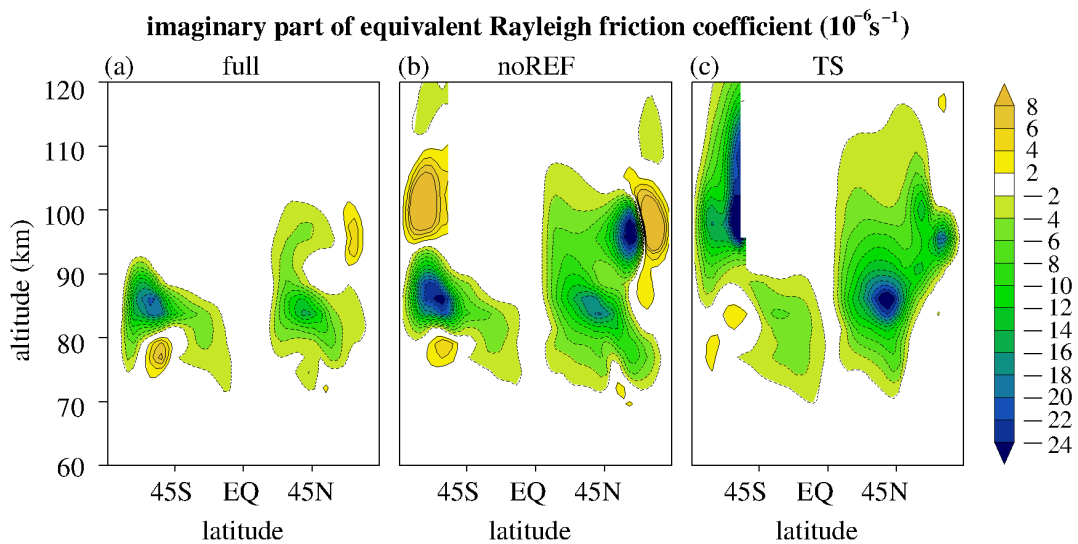


Figure 5.20.: Same as fig. 5.19, but for the imaginary part.

5.6. Summary

- The simple GW ensemble from *Becker and Schmitz* [2003] was chosen as a “toy” configuration to investigate the impacts of temporal and horizontal dependence of the HAMMONIA tides on GW propagation and dissipation.
- Each of the 14 GW members were integrated separately within three different experiments:
 - (i) full ray tracing (“full”),
 - (ii) only vertical ray tracing with time-dependence (“noREF”),
 - (iii) and the conventional GW parameterization setting (“TS”).
- The modulation of GW observed frequencies results from the time-dependence of the diurnal tides.
- Illustrative examples are discussed for which it was shown that
 - (i) the frequency variations (and corresponding phase velocity variations) depend on the ratio between mean GW group velocity and tidal phase speed,
 - (ii) the variations in the saturated pseudo-momentum flux decrease leading to lower diurnal GW forcing amplitudes, and that
 - (iii) less critical filtering appears in the interaction with tidal winds.
- The horizontal refraction of GW fields is induced by horizontal inhomogeneities of the HAMMONIA winds and can directly or indirectly change the GW influence on thermal tides.
- The mean wind jets in the stratosphere refract gravity waves propagating against the wind into the jet streams.

- Planetary waves in the northern winter hemisphere lead to complex, zonally dependent GW motion and meridional transport of GW fields.
- Zonal inhomogeneities and temporal variability enhance the mixing of GW rays and lead to smoother fields of GW properties with smaller “shadow” regions behind critical jets.
- Large meridional displacements of gravity waves can be found for the given GW ensemble.
- Periodic modulation of GW pseudo-momentum fluxes causes a diurnal GW force.
- Frequency modulation and horizontal refraction are each responsible for an overall reduction of the GW forcing amplitude by about 30%, and in some regions by up to 90%.
- The direct effects of the horizontal refraction on the diurnal forcing are in the order of 10% to 20% in main forcing altitudes and partially compensate each other.
- Possible effects of the GW force on the tides depend on the phase relation between them. For this, equivalent friction coefficients are defined:
 - (i) the real part acts on tidal amplitudes
 - (ii) while the imaginary part acts on the tidal phase structure.
- Compared to the conventional simulations, the damping of the tidal amplitudes (which results from the Lindzen saturation assumption and is considered controversial in the literature) is significantly decreased.
- The shrinking of tidal phase structure is found for negative imaginary parts of the friction coefficients and is more confined to the mesopause region in the full ray-tracing simulation.

6. Conclusions and Outlook

6.1. Conclusions

The interaction between gravity waves and thermal tides in the middle atmosphere was investigated. It is believed that gravity-wave breakdown to small-scale turbulent structures is the dominant mechanism for gravity-wave forcing on the mean flow [Fritts and Alexander, 2003]. As this breaking process is periodically modulated by tidal waves, a periodic force results which acts back on middle-atmosphere tides. Hence, a large range of scales spanning from local turbulent eddies to global circulations is involved in the interaction process of which only a small part is resolved in most modern climate models like HAMMONIA. The way gravity waves and tides interact, and also the exchange with the “third partner”-turbulence, are only approximately described using effective subgrid-scale parameterizations.

For this parameterization problem, a consistent theoretical framework was provided. After an investigation of the acoustic and available potential energies of compressible fluid dynamics, the anelastic theory was introduced for the description of gravity-wave motion. A dynamical system with three inherent scales was assumed. The first is connected to the fast turbulent eddies, the second to the gravity-wave motion and the last to the large-scale variations of the basic state flow including thermal tides. Application of a scale-selective filter achieved a reduced large-scale system in which effects of subgrid-scales were included. Mean turbulent forces resulted from turbulent stresses, and heating rates appeared due to mean dissipation of mechanical energy, turbulent fluxes of entropy and sensible heat. For subgrid-scale gravity waves, two different filter strategies were investigated: the Eulerian average and the Lagrangian average. The Eulerian average consists of a local average fixed in space and time. Eulerian-mean gravity-wave forces result from the divergence of the gravity-wave momentum-flux tensor. In contrast, the Lagrangian average is an average along fluid trajectories. Lagrangian-mean gravity-wave forces arise from the divergence of the so-called gravity-wave pseudo-momentum flux tensor which is related to the work done by fluid displacements against the perturbation pressure. For gravity waves affected by Earth’s rotation, the two descriptions are different, but the Lagrangian-mean framework seems to be more appropriate. However, as discussed, the choice of the average depends on the modeler’s taste regarding which kind of gravity-wave parameterization is useful and thus which interpretation of the mean flow is favored. With these preparations, a solid basis was created for the discussion of the impacts of gravity waves on the tidal flow.

The “first partner” in the interaction process, the diurnal tides, was introduced with the help of simulation data from the complex chemistry-climate model HAMMONIA. Diurnal variations in winds and temperature were calculated from a monthly-mean daily cycle. In the mesopause region, the tidal winds have magnitudes up to 50 m/s and are the dominant variability pattern. The extreme wind variations have a strong influence on any type of perturbation moving through them. It was also discussed that non-migrating tidal components, which do not follow the apparent motion of the sun, have a large impact on the tropical middle atmosphere.

For the “second partner” in the interaction process, the gravity waves, linear small-amplitude dynamics were discussed. The reader was led step-by-step to the general ray-

tracing method for gravity waves. First, a typical textbook example of gravity waves was given. With respect to a resting isothermal basic state, dispersion of gravity-wave fields and its typical approximations were discussed. Then an extension of gravity-wave motion in general moving basic states was provided. Via a multiple-scale analysis, the asymptotic regimes for gravity-wave propagation were derived including the non-hydrostatic, the medium-frequency and the inertia-gravity wave regimes. The essential properties of all regimes were collected into a hybrid scheme. Via Wentzel-Kramers-Brillouin theory, asymptotic solutions in the form of gravity-wave packets or modulated wave trains were found. Local dispersion, polarization and wave action relations hold even if the background flow varies slowly compared to the gravity-wave scales. The previous relations form the basis for the ray-tracing method in which each part of the gravity-wave field is traced along its group velocity. The gravity-wave properties change depending on the variations in the background flow. The possible impact of gravity waves on the background flow was discussed with the help of generalized Eliassen-Palm theorems. It was shown, that even if the gravity waves are not affected by (i) viscous damping (ii) local heating and (iii) critical filtering, they can force an induced flow via wave transience and horizontal refraction of the wave field.

The ray-tracing method was applied to the gravity wave-tide interaction. The main objective was an evaluation of the assumptions usually made by single-column gravity-wave parameterizations. A global analysis of gravity-wave fields was performed with a zonally dependent climatological mean flow and diurnal tides taken from HAMMONIA. To quantify the impact of temporal and horizontal variability, a small and highly simplified gravity-wave ensemble was used as a “toy” configuration for the investigation of gravity-wave propagation. Three ray-tracing experiments, (i) “TS” (time slicing), (ii) “noREF” (no horizontal refraction) and (iii) “full”, with increasing complexity were performed. In all three, the background conditions were unaffected by gravity-wave forces. A successive reduction of imposed assumptions gave the opportunity to compare the ray-tracing results with the conventional approach.

When the time-dependence of the thermal tides is included in the description of gravity-wave propagation, observed gravity-wave frequencies are modulated [Eckermann and Marks, 1996; Walterscheid, 2000]. Also, the horizontal gravity-wave phase velocity c_h is periodically changed such that the waves avoid their conventional critical level. Since c_h follows the shape of the background wind, the diurnal gravity-wave forcing is reduced. When horizontal refraction is included, meridional gradients of the climatological background refract gravity-wave fields into the wind jets in such a way that the waves minimize their travel time. Global wave guides are thus formed, occasionally leading to large meridional displacements of parts of the gravity-wave field and to an interhemispheric exchange of wave energy.

The frequency modulation and refraction of horizontal wavenumbers of the gravity-wave fields can substantially reduce the diurnal gravity-wave forcing on the thermal tide. The inclusion of horizontal propagation especially decreases the forcing in polar regions and in the lower thermosphere. Direct forces due to horizontal refraction are not dominant but may lead to second order changes of the diurnal force in some regions. Hence, the apparent differences between the simulations with different restrictions on horizontal refraction and time-dependence are not explained by local forcing effects. Cumulative changes, which each part of the gravity-wave fields undergoes during its propagation, are however of significance.

With the help of Rayleigh friction coefficients, the possible effect of the diurnal gravity-wave force on the diurnal tides was estimated. For the conventional gravity-wave parameterization with the Lindzen saturation assumption, the real part of the equivalent

Rayleigh friction coefficient is mainly positive and is therefore expected to produce a damping of tidal amplitudes. This situation changes when the temporal and horizontal dependence of the background conditions is taken into account. In the more complex ray-tracing simulations, which also use the simple saturation approach for parameterizing turbulence, the forcing is more confined to the mesopause region with much smaller coefficients. Furthermore, there are alternating areas of positive and negative influence on the tide. For the imaginary part of the Rayleigh friction coefficient, results from previous investigations are confirmed [Ortland and Alexander, 2006, and reference therein]. Two negative peaks are found in the mesopause region, which as shown by those authors, decreases the vertical wavelength of the thermal tides.

The idealizations in this study certainly limit the generality of the conclusions that can be drawn regarding the realistic impact of gravity waves on the tide. Nevertheless, it seems to support the view that gravity-wave parameterization should not be used blindly. All assumptions have to be tested for each target problem.

6.2. Outlook

The motto of further improvements to this study would be convergence to more realism with respect to the interaction between gravity waves and thermal tides. Several important points concerning this issue are:

- coupling of gravity-wave source parameters to physical processes in a general climate model
- inclusion of the direct feedback between gravity-wave forces and diurnal variations
- sophisticated description of the wave breaking process
- non-linear interactions between multiple unresolved waves
- deviations from the locally monochromatic wave

An extremely simple gravity-wave source ensemble was used as “toy” configuration for this study. Only a handful gravity-wave fields with very similar horizontal scales were allowed to propagate through the tidal background. For improvement, it would be beneficial to couple the gravity-wave source parameters to the physical processes in a general climate model. *Song and Chun* [2005] formulated the theoretical properties of a spectrum of convectively generated gravity waves. This spectrum was applied by *Song and Chun* [2008] within a ray-tracing study and implemented into the American climate model WACCM. In another study with WACCM, *Richter et al.* [2010] performed experiments including two different non-orographic gravity-wave source mechanisms within a Lindzen gravity-wave parameterization. The first was also associated with deep moist convection, where the gravity waves are excited when the convection parameterization is active (see also *Beres et al.* [2005]). The second was a frontal gravity-wave excitation connected to the frontogenesis function. The overall spatial and temporal structure of both excitation mechanisms is very intermittent with the first more confined to low latitudes and the latter dominant in mid-latitudes. One interesting question concerning the gravity wave-tide interaction arises: How does the temporal modulation of gravity-wave excitation interfere with the tidal waves in the middle atmosphere?

Offline ray-tracing simulation were performed in which the tidal variation was not influenced by the calculated gravity-wave forces. The background data were taken from

a HAMMONIA simulation in which the impact of gravity waves had already been parameterized with the Hines parameterization. This leads to some inconsistency since the derived and applied gravity-wave forces differ. Furthermore, it is not clear to what extent the gravity-wave forces in the complex ray-tracing model will change if the tides interactively react to them. Hence, it is of great interest to couple the ray-tracing model RAPAGI to a tidal model, be it linear or non-linear, in order to investigate the effect of the flow induced by gravity-wave forcing.

One further important simplification made in the current study concerns the parameterization of turbulence induced by gravity-wave breaking, which was kept as simple as possible and described by the Lindzen saturation hypothesis. It is clear that the breaking process is much more complicated. In a series of studies it was shown that gravity waves can break even if their amplitudes are substantially below the convective instability threshold and that the turbulent decay can extract much more energy from the gravity-wave field than assumed by the saturation assumption [Achatz, 2007, and references therein]. Besides the large uncertainties concerning the impact of turbulence on gravity waves, there are already more sophisticated gravity-wave damping schemes available in the literature [see Marks and Eckermann, 1995, and references therein]. In future, a more sophisticated turbulence scheme could be implemented in RAPAGI and applied to global ray simulations. Furthermore, the estimates of the turbulent diffusion coefficients from the saturation approach may be improper. For that reason, the calculations of diurnal heating rates due to gravity-wave dissipation and heat fluxes were postponed until a more sophisticated turbulence scheme is implemented.

McLandress [1998] distinguished between monochromatic and spectral gravity-wave parameterizations. The complex ray-tracing approach falls into the first category. For a more spectral flavor of ray tracing, one could couple several spectral gravity-wave components diffusively. Following Medvedev and Klaassen [1995], a nonlinear wave diffusion can be used in which the influence of other spectral gravity-wave components can destroy the principal wave field. However, it is not obvious whether the large amount of spectral information required for the nonlinear damping is computationally feasible within global ray tracing.

In addition, one might think about the study of more complex objects than locally monochromatic gravity waves. These may result from more realistic gravity-wave source shapes or from shape-deforming mechanisms (wave focusing, wave reflexion, critical filtering etc.). Ray tracing might also be useful in that case if it is assumed that the gravity-wave pattern is composed of a large number of spectral components. The tracing would be done in a phase space or so-called augmented space [Hayes, 1970; Eckermann and Marks, 1996; Hertzog et al., 2001] and integration over all spectral information is required to compute effects on the tides. The wave action can be distributed over phase space quite irregularly, with splitting and merging of several wave parts possible, but the phase-space volume occupied by the wave pattern is constant [e.g. Hertzog et al., 2002; Broutman et al., 2004]. However, even if this method is able to produce finite estimates of wave action density at a caustic, it still relies on the scale-separation assumption and breaks down if the scales of certain spectral components become close to the background scales.

It might be annoying that the number of open questions raised by this study is much larger than the answers proposed, but this seems to be notoriously the case in scientific research!

A. Appendix

A.1. Approximations of the Lagrangian-mean anelastic dynamics

In the following, the approximation steps which led in section 2.2.3 from the original set (2.68), (2.69) to the final set (2.70), (2.71) are provided. It is assumed that the displacement is small compared to the unresolved wave scales and that the displacement and the wave scales are also small compared to the resolved scales. For vertical wave scale \hat{h} , vertical background scale \hat{H} and the vertical displacement ζ , the following symbolic hierarchy is assumed

$$|\zeta| \ll \hat{h} \ll \hat{H}. \quad (\text{A.1})$$

First of all, the Lagrangian-mean momentum equation is repeated:

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla \right) \langle \mathbf{v} \rangle_w^L + \underbrace{\langle 2\boldsymbol{\Omega}^\xi \times \mathbf{v}^\xi \rangle_w}_{\textcircled{1}} = & - \underbrace{\langle (\nabla\phi)^\xi \rangle_w}_{\textcircled{2}} + \underbrace{\langle b^\xi \mathbf{e}_z^\xi \rangle_w}_{\textcircled{3}} \\ & + \underbrace{\frac{1}{\tilde{\rho}_r} \nabla \cdot \langle \mathbf{K}^T \cdot (\mathbf{F}^\xi + \mathbf{R}_t^\xi) \rangle_w}_{\textcircled{4}}, \end{aligned} \quad (\text{A.2})$$

The four numbered terms are linearized with respect to $\boldsymbol{\xi}$ and then discussed separately. For $\textcircled{1}$, the angular velocity is expanded to obtain $\boldsymbol{\Omega}^\xi = \boldsymbol{\Omega} + \boldsymbol{\xi} \cdot \nabla \boldsymbol{\Omega}$. The gradient of $\boldsymbol{\Omega}$ includes two effects: First, the meridional dependence of the components $\Omega_v = \Omega \sin \varphi$ and $\Omega_h = \Omega \cos \varphi$, the so-called beta-effect, and second, due to the spherical nature of Earth also the unit vectors change with position, hence also with displacement, and additional metric terms appear. Both are neglected to arrive at

$$\textcircled{1} \approx \left\langle 2(\boldsymbol{\Omega} + \boldsymbol{\xi} \cdot \nabla \boldsymbol{\Omega}) \times (\langle \mathbf{v} \rangle_w^L + \mathbf{v}^\ell) \right\rangle_w \approx 2\boldsymbol{\Omega} \times \langle \mathbf{v} \rangle_w^L. \quad (\text{A.3})$$

From the lifted pressure gradient term $\textcircled{2}$, important impacts of the subgrid-scale arise:

$$\textcircled{2} \approx - \langle \nabla\phi + \boldsymbol{\xi} \cdot \nabla \nabla\phi \rangle_w \approx -\nabla \langle \phi \rangle_w - \frac{1}{\tilde{\rho}_r} \nabla \cdot (\tilde{\rho}_r \langle \boldsymbol{\xi} \nabla \phi' \rangle_w), \quad (\text{A.4})$$

where in the second step the linearized continuity equation (2.63), i.e. $\nabla \cdot (\tilde{\rho}_r \boldsymbol{\xi}) = 0$, was used. Note that instead of $\langle \phi \rangle_w^L$ the term $\langle \phi \rangle_w$ appears here which is in the author's personal opinion more convenient. The corresponding Stokes correction $\langle \phi \rangle_w^L - \langle \phi \rangle_w$ is negligible at the given order. For the buoyancy, there is a coupling between the perturbation and the metric which is approximated by

$$\textcircled{3} \approx \left\langle (\langle b \rangle_w^L + b^\ell) (\mathbf{e}_z + \boldsymbol{\xi} \cdot \nabla \mathbf{e}_z) \right\rangle_w \approx \langle b \rangle_w^L \mathbf{e}_z \quad (\text{A.5})$$

Lastly, for the frictional stresses (using just \mathbf{F}^ξ),

$$\left\langle \mathbf{K}^T \cdot \mathbf{F}^\xi \right\rangle_w \approx \left\langle ((1 + \nabla \cdot \boldsymbol{\xi}) \mathbf{1} - (\nabla \boldsymbol{\xi})^T) \cdot (\mathbf{F} + \boldsymbol{\xi} \cdot \nabla \mathbf{F}) \right\rangle_w \quad (\text{A.6})$$

$$\approx \langle \mathbf{F} \rangle_w + \langle (\nabla \cdot \boldsymbol{\xi}) \mathbf{F}' \rangle_w + \langle \boldsymbol{\xi} \cdot \nabla \mathbf{F}' \rangle_w - \left(\langle \mathbf{F}' \cdot \nabla \boldsymbol{\xi} \rangle_w \right)^T \quad (\text{A.7})$$

$$= \langle \mathbf{F} \rangle_w + \nabla \cdot \langle \boldsymbol{\xi} \mathbf{F}' \rangle_w - \left(\nabla \cdot \langle \mathbf{F}' \boldsymbol{\xi} \rangle_w \right)^T + \langle \boldsymbol{\xi} \nabla \cdot \mathbf{F}' \rangle_w, \quad (\text{A.8})$$

where $\mathbf{F} = \mathbf{F}^T$ was required and the transformation of lifted surface elements (2.60) was used. In the last line, the two flux-terms are negligible compared to the last term, as they involve the background scale whereas the last term incorporates wave scales. Hence, the following approximation might be convenient:

$$\textcircled{4} \approx \frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\langle \mathbf{F} \rangle_w + \langle \mathbf{R}_t \rangle_w + \langle \boldsymbol{\xi} \nabla \cdot (\mathbf{F}' + \mathbf{R}'_t) \rangle_w \right). \quad (\text{A.9})$$

The frictional stress tensors \mathbf{F} and \mathbf{R}_t might be chosen as linear functions of $\nabla \mathbf{v}$. Then $\langle \mathbf{F} \rangle_w$ and $\langle \mathbf{R} \rangle_w$ are linear functions of $\nabla \langle \mathbf{v} \rangle_w$ and not of $\nabla \langle \mathbf{v} \rangle_w^L$. But $\langle \mathbf{v} \rangle_w^L$ are the prognostic variables of the Lagrangian-mean system and a corresponding reformulation is desirable. However, it seems to be quite uncritical to also neglect the resulting Stokes correction within the parameterization of friction. From the above set of approximations the Lagrangian-mean momentum equation (2.70) derives. The thermodynamic equation (2.69) is also restated:

$$\begin{aligned} \left(\partial_t + \langle \mathbf{v} \rangle_w^L \cdot \nabla \right) \langle h_{an} \rangle_w^L = & - \left(g + \langle b \rangle_w^L \right) \langle w \rangle_w^L - \underbrace{\langle b^\ell w^\ell \rangle_w}_{\textcircled{5}} \\ & - \underbrace{\frac{1}{\tilde{\rho}_r} \nabla \cdot \left\langle \mathbf{K}^T \cdot (\mathbf{q}^\xi + \mathbf{J}_t^\xi) \right\rangle_w}_{\textcircled{6}} + \underbrace{\langle \epsilon \rangle_w^L}_{\textcircled{7}} + \langle Q \rangle_w^L. \end{aligned} \quad (\text{A.10})$$

The vertical buoyancy flux is

$$\textcircled{5} \approx - \langle (b' + \boldsymbol{\xi} \cdot \nabla b) (w' + \boldsymbol{\xi} \cdot \nabla w) \rangle_w \approx - \langle b' w' \rangle_w, \quad (\text{A.11})$$

where in the last step the terms $\langle b' \boldsymbol{\xi} \rangle_w \cdot \nabla \langle w \rangle_w$ and $\langle w' \boldsymbol{\xi} \rangle_w \cdot \nabla \langle b \rangle_w$ were neglected because the ratio between background scales and displacement is assumed to be small. For the heat fluxes in $\textcircled{6}$ the results from $\textcircled{4}$ can be used as \mathbf{q} and \mathbf{J}_t are tensors of order one, i.e.

$$\textcircled{6} \approx \frac{1}{\tilde{\rho}_r} \nabla \cdot \left(\langle \mathbf{q} \rangle_w + \langle \mathbf{J}_t \rangle_w + \langle \boldsymbol{\xi} \nabla \cdot (\mathbf{q}' + \mathbf{J}'_t) \rangle_w \right). \quad (\text{A.12})$$

The rate of dissipation of mechanical energy is approximated by

$$\tilde{\rho}_r \langle \epsilon \rangle_w^L \approx \nabla \langle \mathbf{v} \rangle_w \cdot \left(\langle \mathbf{F} \rangle_w + \langle \mathbf{R}_t \rangle_w \right) + \langle \nabla \mathbf{v}' \cdot (\mathbf{F}' + \mathbf{R}'_t) \rangle_w, \quad (\text{A.13})$$

where the impact of the displacement, i.e. the terms $\boldsymbol{\xi} \cdot \nabla \nabla \mathbf{v}$ and $\boldsymbol{\xi} \cdot \nabla (\mathbf{F} + \mathbf{R}_t)$, was neglected. Furthermore, it seems not to be crucial to replace $\langle \mathbf{v} \rangle_w \rightarrow \langle \mathbf{v} \rangle_w^L$ in the mean dissipation rate. Then the Lagrangian-mean enthalpy equation (2.71) results. At this stage, it is not obvious whether the resulting Lagrangian-mean system consistently conserves any type of energy and therefore further evaluation is needed to resolve this issue.

A.2. Non-dimensionalization of linear gravity-wave dynamics

A.2.1. Linear continuity equation

With the density scale height $H_\rho = -(\partial_z \ln \rho_r)^{-1}$, the continuity equation (4.34) is written as

$$\nabla_h \cdot \mathbf{u}' + \frac{\partial w'}{\partial z} - \frac{w'}{H_\rho} = 0. \quad (\text{A.14})$$

In the following, all non-dimensional variables are indicated with an asterisk. Using the scales discussed in section 4.3.2, the continuity eq. transforms to

$$\nabla_h^* \cdot \mathbf{u}^* + \frac{\partial w^*}{\partial z^*} - \varepsilon \frac{w^*}{H_\rho^*} = 0, \quad (\text{A.15})$$

where $\varepsilon = \hat{h}/\hat{H}$ is the small two-scale ratio and $H_\rho^* = H_\rho/\hat{H}$.

A.2.2. Linear momentum balance

For the linear momentum balance (4.35), horizontal and vertical dynamics are considered separately. Using scales listed in the tables 4.2-4.4 and discussed in section 4.3.2, the magnitude of each term is compared to the inertial force. The effects of the spherical nature of the earth are neglected for the scaling of GW variations. The equations of linear motion are then

$$D_t u' - 2\Omega_v v' + 2\Omega_h w' = -\mathbf{v}' \cdot \nabla u - \partial_x \phi' + f'_{R,x}, \quad (\text{A.16})$$

$$D_t v' + 2\Omega_v u' = -\mathbf{v}' \cdot \nabla v - \partial_y \phi' + f'_{R,y}, \quad (\text{A.17})$$

$$D_t w' - 2\Omega_h u' = -\partial_z \phi' + b + f'_{R,z}, \quad (\text{A.18})$$

where $\Omega_h = \Omega \cos \varphi$ and $\Omega_v = \Omega \sin \varphi$ are the horizontal and vertical components of the rotation vector, respectively, and $f'_{R,x}$, $f'_{R,y}$ and $f'_{R,z}$ are the components of the friction force

$$\mathbf{f}'_R = (\nu + \nu_t) \left\{ \nabla^2 \mathbf{v}' + \nabla \left(\frac{w'}{H_\rho} \right) \right\} + \frac{\partial_z (\rho_r (\nu + \nu_t))}{\rho_r} (\partial_z \mathbf{v}' + \nabla w'). \quad (\text{A.19})$$

The different terms in the equations of linear motion are compared to the inertial forces, which scale like

$$|D_t u'| \sim |D_t v'| \sim \varepsilon \hat{l}/\hat{t}^2 \quad \text{and} \quad |D_t w'| \sim \varepsilon \hat{h}/\hat{t}^2, \quad (\text{A.20})$$

where the mean Strouhal number was set to $\text{Sr}_m = \varepsilon^{-1}$. The local time derivative $\partial_t u'$ is also dominated by \hat{u}/\hat{t} . The advection by the horizontal wind is

$$|\mathbf{u} \cdot \nabla_h u'| \sim |\mathbf{u} \cdot \nabla_h v'| \sim \hat{\mathbf{U}} \hat{u}/\hat{l} = \hat{u}^2/(\varepsilon \hat{l}) = \hat{u}/\hat{t} \sim |\partial_t u'| \sim |\partial_t v'|, \quad (\text{A.21})$$

where $\hat{\mathbf{U}} = \hat{u}/\varepsilon$ and $\hat{u} = \varepsilon \hat{l}/\hat{t}$ was used. Using $\hat{W} = \hat{a}_B \text{Ro}_m \hat{\mathbf{U}}$, the advection by the vertical wind is

$$|w \partial_z u'| \sim |w \partial_z v'| \sim \hat{W} \hat{u}/\hat{h} = \hat{a}_B \text{Ro}_m \hat{\mathbf{U}} \hat{u}/\hat{h} = \hat{u}/\hat{T} = \varepsilon_t \hat{u}/\hat{t} \quad (\text{A.22})$$

and therefore at least one order of magnitude smaller than the other terms. The same is true for the advection of w' . Additionally, for the Coriolis forces

$$\frac{|2\Omega_v v'|}{|D_t u'|} \sim \frac{|2\Omega_v u'|}{|D_t v'|} \sim \Omega \hat{t} = \frac{1}{\text{Ro}_w}, \quad (\text{A.23})$$

$$\frac{|2\Omega_h w'|}{|D_t u'|} \sim \hat{\mathbf{a}} \Omega \hat{\mathbf{t}} = \frac{\hat{\mathbf{a}}}{\text{Ro}_w}, \quad (\text{A.24})$$

$$\frac{|2\Omega_h u'|}{|D_t w'|} \sim \hat{\mathbf{a}}^{-1} \Omega \hat{\mathbf{t}} = \frac{1}{\hat{\mathbf{a}} \text{Ro}_w} \quad (\text{A.25})$$

are obtained where the scaling of wave Rossby number $\text{Ro}_w = (\Omega \hat{\mathbf{t}})^{-1}$ depends on the GW regime under consideration. Next, for the force due to background wind gradients, the vertical shear is distinguished from the horizontal gradients:

$$\frac{|\mathbf{u}' \cdot \nabla_h u|}{|D_t u'|} \sim \frac{|\mathbf{u}' \cdot \nabla_h v|}{|D_t v'|} \sim \frac{\hat{\mathbf{l}}}{\hat{\mathbf{L}}} = \varepsilon_x = \frac{\hat{\mathbf{a}}_B}{\hat{\mathbf{a}}} \varepsilon, \quad (\text{A.26})$$

$$\frac{|w' \partial_z u|}{|D_t u'|} \sim \frac{|w' \partial_z v|}{|D_t v'|} \sim \varepsilon, \quad (\text{A.27})$$

where $\hat{\mathbf{u}} \approx \varepsilon \hat{\mathbf{U}}$ was used. The scaled pressure gradients behave like

$$\frac{|\partial_x \phi'|}{|D_t u'|} \sim \frac{|\partial_y \phi'|}{|D_t v'|} \sim 1, \quad (\text{A.28})$$

$$\frac{|\partial_z \phi'|}{|D_t w'|} \sim \frac{1}{\hat{\mathbf{a}}^2} \quad (\text{A.29})$$

and the buoyancy force scales like

$$\frac{|b'|}{|D_t w'|} \sim \frac{1}{\hat{\mathbf{a}}^2}. \quad (\text{A.30})$$

For the viscous force, the Laplacian of the wind perturbation is the dominant contribution. The other terms depend on the vertical gradients of background density and viscosity. With the help of the wave Reynolds number $\text{Re}_w = \hat{\mathbf{h}}^2 / (\hat{\mathbf{v}} \hat{\mathbf{t}})$, the viscous damping scales like

$$\frac{|(\nu + \nu_t) \partial_{zz} \mathbf{u}'|}{|D_t \mathbf{u}'|} \sim \frac{|(\nu + \nu_t) \partial_{zz} w'|}{|D_t w'|} \sim \frac{\hat{\mathbf{v}} \hat{\mathbf{t}}}{\hat{\mathbf{h}}^2} = \frac{1}{\text{Re}_w}, \quad (\text{A.31})$$

$$\frac{|(\nu + \nu_t) \nabla_h^2 \mathbf{u}'|}{|D_t \mathbf{u}'|} \sim \frac{|(\nu + \nu_t) \nabla_h^2 w'|}{|D_t w'|} \sim \frac{\hat{\mathbf{v}} \hat{\mathbf{t}}}{\hat{\mathbf{l}}^2} = \frac{\hat{\mathbf{a}}^2}{\text{Re}_w}. \quad (\text{A.32})$$

For the case of strong turbulent diffusion, e.g. in the mesopause region, the turbulent viscosity can be in the order of $\nu_t \approx 100 \text{ m}^2/\text{s}$. In this case, Re_w is ≈ 1 for medium-frequency waves, but ≈ 100 for non-hydrostatic waves. Altogether, the dimensionless equations of motion are

$$\begin{aligned} \frac{Du^*}{Dt^*} = & -\varepsilon_t W^* \frac{\partial u^*}{\partial z^*} + \frac{2}{\text{Ro}_w} (v^* \sin \varphi - \hat{\mathbf{a}} w^* \cos \varphi) \\ & - \varepsilon_x \mathbf{u}^* \cdot \nabla_h^* U^* - \varepsilon w^* \frac{\partial U^*}{\partial z^*} - \frac{\partial \phi^*}{\partial x^*} + \frac{f_{R,x}^*}{\text{Re}_w}, \end{aligned} \quad (\text{A.33})$$

$$\begin{aligned} \frac{Dv^*}{Dt^*} = & -\varepsilon_t W^* \frac{\partial v^*}{\partial z^*} - \frac{2}{\text{Ro}_w} u^* \sin \varphi \\ & - \varepsilon_x \mathbf{u}^* \cdot \nabla_h^* V^* - \varepsilon w^* \frac{\partial V^*}{\partial z^*} - \frac{\partial \phi^*}{\partial y^*} + \frac{f_{R,y}^*}{\text{Re}_w}, \end{aligned} \quad (\text{A.34})$$

$$\hat{\mathbf{a}}^2 \frac{Dw^*}{Dt^*} = -\varepsilon_t \hat{\mathbf{a}}^2 W^* \frac{\partial w^*}{\partial z^*} + \frac{2 \hat{\mathbf{a}}}{\text{Ro}_w} u^* \cos \varphi - \frac{\partial \phi^*}{\partial z^*} + b^* + \hat{\mathbf{a}}^2 \frac{f_{R,z}^*}{\text{Re}_w}, \quad (\text{A.35})$$

where the frictional force is

$$\mathbf{f}_R^* = (\nu^* + \nu_t^*) \left\{ \frac{\partial^2 \mathbf{v}^*}{\partial z^{*2}} + \hat{\mathbf{a}}^2 \nabla_h^* \cdot (\nabla_h^* \mathbf{v}^*) \right\} + O(\varepsilon), \quad (\text{A.36})$$

and D/Dt^* is the advective derivative with the horizontal background wind (U^*, V^*) .

A.2.3. Linear thermodynamics

The scaling of the linear buoyancy equation (4.36) is investigated in detail. The buoyancy tendency scales like

$$|D_t b'| \sim \frac{\hat{\mathbf{b}}}{\hat{\mathbf{t}}}. \quad (\text{A.37})$$

The restoring motion due to the background stratification

$$\frac{|w' N^2|}{|D_t b'|} \sim 1 \quad (\text{A.38})$$

is the dominant mechanism for GW dynamics. The ratio between the mean and perturbation buoyancy scales is

$$\frac{\hat{\mathbf{B}}}{\hat{\mathbf{b}}} = \varepsilon^{-1} \quad (\text{A.39})$$

where $\hat{\mathbf{u}} = \varepsilon \hat{\mathbf{U}}$ and $\hat{\mathbf{U}} = \hat{\mathbf{L}} \Omega / \text{Sr}_m$ with the mean Strouhal number Sr_m . The advection of mean buoyancy by horizontal wind perturbations scales like

$$\frac{|\mathbf{u}' \cdot \nabla_h b|}{|D_t b'|} \sim \text{Sr}_m^{-1} \text{Ro}_w^{-1} = \varepsilon_x, \quad (\text{A.40})$$

again using $\text{Sr}_m = \varepsilon^{-1}$ and $\varepsilon_x = \varepsilon \text{Ro}_w^{-1}$, whereas terms involving buoyancy advection by the vertical wind scale like

$$\frac{|w' \partial_z b|}{|D_t b'|} \sim \frac{|w' b / H_\theta|}{|D_t b'|} \sim \text{Sr}_m^{-1} = \varepsilon \quad \text{and} \quad \frac{|w b' / H_\theta|}{|D_t b'|} \sim \frac{\varepsilon_t}{\text{Sr}_m} = \varepsilon \varepsilon_t. \quad (\text{A.41})$$

The heating terms behave like

$$\frac{|g \epsilon' / (c_p T_r)|}{|D_t b'|} \sim \frac{\varepsilon^2}{\text{Re}_w}, \quad \frac{|g Q' / (c_p T_r)|}{|D_t b'|} \sim \frac{1}{\text{He}}, \quad (\text{A.42})$$

$$\frac{|\text{Pr}^{-1}(\nu + \nu_t) \partial_{zz} b'|}{|D_t b'|} \sim \frac{1}{\text{Pr Re}_w}, \quad \frac{|\text{Pr}^{-1}(\nu + \nu_t) \nabla_h^2 b'|}{|D_t b'|} \sim \frac{\hat{\mathbf{a}}^2}{\text{Pr Re}_w}. \quad (\text{A.43})$$

The heating number He is the ratio of local enthalpy change to diabatic heating induced by latent heat release, interaction with solar or terrestrial radiation and other external heat sources. Therefore, the dimensionless thermodynamic equation is

$$\begin{aligned} \frac{D b^*}{D T^*} + w^* (N^*)^2 &= -\varepsilon_t W^* \frac{\partial b^*}{\partial z^*} - \varepsilon_x \mathbf{u}^* \cdot \nabla_h^* B^* \\ &- \varepsilon \left(w^* \frac{\partial B^*}{\partial z^*} + \frac{w^* B^*}{H_\theta^*} + \varepsilon_t \frac{W^* b^*}{H_\theta^*} \right) + \frac{R H_p^*}{c_p} \frac{Q^*}{\text{He}} + \frac{f_q^*}{\text{Re}_w \text{Pr}} + O(\varepsilon^2), \end{aligned} \quad (\text{A.44})$$

where the convergence of the heat fluxes

$$f_q^* = (\nu^* + \nu_t^*) \left(\frac{\partial^2 b^*}{\partial z^{*2}} + \hat{\mathbf{a}}^2 \nabla_h^* \cdot (\nabla_h^* b^*) \right) + O(\varepsilon) \quad (\text{A.45})$$

and the pressure scale $H_p = -\partial_z \ln p_r$ were utilized.

A.3. Relations between gravity-wave variables

A.3.1. Wave energy and fluxes

In the following, relations between different wave energy contributions are derived. First, the horizontal wind perturbation $\hat{\mathbf{u}}$ is divided into components parallel to the horizontal wavenumber vector along $\mathbf{e}_{\parallel} = \mathbf{k}_h/k_h$ and parallel to the wave front (orthogonal to the wavenumber vector) along $\mathbf{e}_{\perp} = (-l, k, 0)^T/k_h$. Using eq. (4.71) and (4.72), both components with included dimensions are

$$\hat{u}_{\parallel} = \hat{\mathbf{u}} \cdot \mathbf{e}_{\parallel} = -i \frac{m}{k_h} \frac{\hat{\omega} \hat{b}}{N^2}, \quad (\text{A.46})$$

$$\hat{u}_{\perp} = \hat{\mathbf{u}} \cdot \mathbf{e}_{\perp} = -\frac{m}{k_h} \frac{f \hat{b}}{N^2}. \quad (\text{A.47})$$

Hence, rotation causes motion perpendicular to \mathbf{k}_h . The kinetic energy associated with parallel motion is ¹

$$E_{kin,\parallel} = \frac{\rho_r}{4} |\hat{u}_{\parallel}|^2 = \frac{\rho_r}{4} \frac{m^2}{k_h^2} \frac{\hat{\omega}^2 |\hat{b}|^2}{N^4}, \quad (\text{A.48})$$

with perpendicular motion is

$$E_{kin,\perp} = \frac{\rho_r}{4} |\hat{u}_{\perp}|^2 = \frac{\rho_r}{4} \frac{m^2}{k_h^2} \frac{f^2 |\hat{b}|^2}{N^4}, \quad (\text{A.49})$$

and with vertical motion is

$$E_{kin,v} = \frac{\rho_r}{4} |\hat{w}|^2 = \frac{\rho_r}{4} \frac{\hat{\omega}^2 |\hat{b}|^2}{N^4}. \quad (\text{A.50})$$

Therefore, the total kinetic energy

$$E_{kin} = E_{kin,\parallel} + E_{kin,\perp} + E_{kin,v} = \frac{\rho_r}{4} (\hat{\omega}^2 |\mathbf{k}|^2 + f^2 m^2) \frac{|\hat{b}|^2}{k_h^2 N^4} \quad (\text{A.51})$$

is, due to the impact of rotation, larger than the available potential energy

$$E_{pot} = \frac{\rho_r}{4} \frac{|\hat{b}|^2}{N^2}, \quad (\text{A.52})$$

Interestingly, equipartition between kinetic energy due to parallel motion and available potential energy holds if the kinetic energy due to perpendicular motion is counted with the available potential energy [Bühler, 2009], i.e.

$$E_{kin,\parallel} + E_{kin,v} = E_{pot} + E_{kin,\perp}. \quad (\text{A.53})$$

Furthermore, in the limit $f \rightarrow 0$ kinetic and available potential energy are equal. The total wave energy is the sum of all energy terms,

$$E_w = \frac{\rho_r}{2} \frac{\hat{\omega}^2 |\mathbf{k}|^2}{k_h^2 N^4} |\hat{b}|^2. \quad (\text{A.54})$$

¹ The factor 1/4 is easily explained: For instance, if the term $\langle u'^2 \rangle_w / 2$ is expanded using $u' = \Re \{ \hat{u} e^{i\Theta} \}$, one obtains

$$\langle u'^2 \rangle_w / 2 = \langle (\hat{u} e^{i\Theta} + \hat{u}^* e^{-i\Theta})^2 \rangle_w / 8 = |\hat{u}|^2 / 4$$

The pressure flux is responsible for the intrinsic transport of wave energy, since

$$\rho_r \langle \phi' \mathbf{u}' \rangle_w = \frac{\rho_r}{2} \Re \left\{ \hat{\phi} \hat{u}_{\parallel}^* \right\} \mathbf{e}_{\parallel} = \frac{\rho_r}{2} \frac{(N^2 - \hat{\omega}^2) \hat{\omega}}{N^4 k_h} |\hat{b}|^2 \mathbf{e}_{\parallel} = \hat{c}_{gh} E_w \mathbf{e}_{\parallel}, \quad (\text{A.55})$$

$$\rho_r \langle \phi' w' \rangle_w = \frac{\rho_r}{2} \Re \left\{ \hat{\phi} \hat{w}^* \right\} = -\frac{\rho_r}{2} \frac{(N^2 - \hat{\omega}^2) \hat{\omega}}{N^4 m} |\hat{b}|^2 = \hat{c}_{gz} E_w, \quad (\text{A.56})$$

with the intrinsic group velocities $\hat{c}_{gh} = c_{gh} - u_h$, $\hat{c}_{gz} = c_{gz}$ and $c_{gh} = \mathbf{c}_g \cdot \mathbf{k}_h / k_h$. Horizontal entropy transport is induced by wind component perpendicular to \mathbf{k}_h

$$\rho_r \langle b' \mathbf{u}' \rangle_w = \frac{\rho_r}{2} \Re \left\{ \hat{b} \hat{u}_{\perp}^* \right\} \mathbf{e}_{\perp} = -\frac{\rho_r}{2} \frac{m f}{N^2 k_h} |\hat{b}|^2 \mathbf{e}_{\perp} = -\frac{m k_h f N^2}{|\mathbf{k}|^2 \hat{\omega}^2} E_w \mathbf{e}_{\perp} \quad (\text{A.57})$$

whereas the vertical entropy transport is zero

$$\rho_r \langle b' w' \rangle_w = 0, \quad (\text{A.58})$$

because no damping of gravity waves was included in the lowest order of the WKB expansion. The momentum flux tensor is easily rewritten in terms of \mathbf{e}_{\parallel} and \mathbf{e}_{\perp} to get

$$\begin{aligned} \mathbf{R}_w &= \rho_r \langle \mathbf{v}' \mathbf{v}' \rangle_w \\ &= \rho_r \left(\langle u_{\parallel}'^2 \rangle_w \mathbf{e}_{\parallel} \mathbf{e}_{\parallel} + \langle u_{\perp}'^2 \rangle_w \mathbf{e}_{\perp} \mathbf{e}_{\perp} + \langle w'^2 \rangle_w \mathbf{e}_z \mathbf{e}_z + \langle u_{\parallel}' w' \rangle_w (\mathbf{e}_{\parallel} \mathbf{e}_z + \mathbf{e}_z \mathbf{e}_{\parallel}) \right). \end{aligned} \quad (\text{A.59})$$

The symmetric parts are connected to the kinetic energies in their respective directions. The only off-diagonal part is the vertical flux of horizontal momentum

$$\rho_r \langle u_{\parallel}' w' \rangle_w = -\frac{\rho_r}{2} \frac{m \hat{\omega}^2}{N^2 k_h} |\hat{b}|^2 = -\frac{m k_h}{|\mathbf{k}|^2} E_w. \quad (\text{A.60})$$

For waves affected by rotation it has been discussed in *Andrews et al.* [1987] that parts proportional to the entropy flux must be removed from the momentum flux to obtain the more meaningful Eliassen-Palm flux (EP flux)

$$F_{\text{EP}} = -\rho_r \left(\langle u_{\parallel}' w' \rangle_w - f \langle b' u_{\perp}' \rangle_w / N^2 \right) = -\rho_r \langle u_{\parallel}' w' \rangle_w \left(1 - \frac{f^2}{\hat{\omega}^2} \right). \quad (\text{A.61})$$

Hence the EP flux carried by gravity waves is less for inertia gravity waves.

A.3.2. Wave action and pseudo-momentum

Two extremely important wave quantities are the wave action density

$$A = \frac{E_w}{\hat{\omega}} \quad (\text{A.62})$$

and the horizontal pseudo-momentum

$$\mathcal{P}_h = k_h A. \quad (\text{A.63})$$

The wave action density A is a quadratic measure of wave amplitude. It will be shown that A is conserved if no wave dissipation is active. The pseudo-momentum \mathcal{P}_h can be exchanged with the mean momentum *as if* the waves carried that amount of momentum [McIntyre, 1980]. It is an important quantity for discussing circulation changes due wave fields in a mean flow [Bühler, 2009].

The EP flux (A.61) can be rewritten with the help of eq. (4.80) as

$$F_{\text{EP}} = -\hat{c}_{gz} k_h \frac{E_w}{\hat{\omega}} = -F_h, \quad (\text{A.64})$$

where $F_h = c_{gz} \mathcal{P}_h$ is the vertical flux of horizontal pseudo-momentum. Therefore, the EP flux concept, successfully applied to e.g. Rossby wave dynamics, easily translates to the concept of wave pseudo-momentum. The mean work done by GW displacements against GW pressure perturbations strongly influences the Lagrangian-mean flow. The resulting wave stress tensor (see eq. 2.72) is

$$\rho_r \langle \boldsymbol{\xi} \nabla \phi' \rangle_w \approx \frac{\rho_r}{2} \Re \left\{ \hat{\boldsymbol{\xi}} (-i\mathbf{k}) \hat{\phi}^* \right\} = \frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{v}} \hat{\phi}^* \right\} \frac{\mathbf{k}}{\hat{\omega}}, \quad (\text{A.65})$$

where in the second step $\hat{\mathbf{u}} = -i\hat{\omega} \hat{\boldsymbol{\xi}}$ was used. It was shown in eq. (A.55) and (A.56) that the pressure flux is equal to the intrinsic energy flux for GW packets in the WKB approximation. Hence the resulting wave stress on the Lagrangian-mean flow is due mainly to the flux of the pseudo-momentum vector $\mathbf{k}A$, i.e.

$$\rho_r \langle \boldsymbol{\xi} \nabla \phi' \rangle_w \approx \hat{c}_g \mathbf{k} A \quad (\text{A.66})$$

In the next section, a prognostic equation for A is derived which determines the evolution of \mathcal{P}_h .

A.4. Conservation of wave action

It was shown by *Grimshaw* [1975a, b] that the wave energy equation (4.94) with its sources, i.e. shear and buoyancy production, can be reformulated as a conservation law for the wave action. It is repeated here with special emphasis on the metric of the spherical earth which extends the analysis in Cartesian coordinates performed by *Grimshaw* [1975a, b]. For inertia-gravity waves, constraints on the background flow are derived.

First, the mean flow dynamics are reviewed. The horizontal momentum equation is given by

$$D_t \mathbf{u} + f \mathbf{e}_z \times \mathbf{u} = -\nabla \phi + \mathbf{f}_h. \quad (\text{A.67})$$

The effect of wave, turbulent and molecular stresses are collected into \mathbf{f}_h . For small mean Rossby numbers $\text{Ro}_m \ll 1$, a dominant geostrophic balance between Coriolis and pressure gradient forces exists. The momentum advection $D_t \mathbf{u}$ is incorporated into an ageostrophic forcing $\mathbf{f}_{ag} = \mathbf{f}_h - D_t \mathbf{u}$ so that

$$f \mathbf{e}_z \times \mathbf{u} = -\nabla \phi + \mathbf{f}_{ag}. \quad (\text{A.68})$$

The term \mathbf{f}_{ag} is responsible for $O(\text{Ro}_m)$ corrections to the geostrophic mean flow. Using hydrostatic balance $\partial_z \phi = b$, the vertical shear $\partial_z \mathbf{u}$ can be related to the large-scale buoyancy gradient ∇b , i.e. the thermal wind relation is

$$f \mathbf{e}_z \times \partial_z \mathbf{u} = -\nabla b + \partial_z \mathbf{f}_{ag}. \quad (\text{A.69})$$

Applying the curl to eq. (A.68), the divergence of the mean horizontal wind is induced by the beta effect and the curl of the ageostrophic forces

$$\nabla_h \cdot \mathbf{u} = -\frac{v\beta}{f} + \frac{(\nabla_h \times \mathbf{f}_{ag}) \cdot \mathbf{e}_z}{f}, \quad (\text{A.70})$$

where $\beta = \partial_\varphi f / a_E$. In the following, several steps are performed to systematically derive a wave action equation.

1st Step: The buoyancy production $\rho_r \langle b' \mathbf{u}' \rangle_w \cdot \nabla_h b / N^2$ is reformulated in terms of shear production.

2nd Step: The horizontal part of the wave stress tensor $\rho_r \langle \mathbf{u}' \mathbf{u}' \rangle_w$ is rewritten in terms of pressure flux and residual terms.

3rd Step: The ray equation for $\hat{\omega}$ is derived and related to the sum of shear and buoyancy production.

Calculations in the 1st Step:

The horizontal entropy flux is perpendicular to the wave direction,

$$\mathbf{e}_\perp \cdot \nabla_h b = -f \mathbf{e}_\perp \cdot (\mathbf{e}_z \times \partial_z \mathbf{u}) + \mathbf{e}_\perp \cdot \partial_z \mathbf{f}_{ag} \quad (\text{A.71})$$

$$= -f \mathbf{e}_\parallel \cdot \partial_z \mathbf{u} + \mathbf{e}_\perp \cdot \partial_z \mathbf{f}_{ag}. \quad (\text{A.72})$$

Thus, the buoyancy production is

$$\frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{u}} \hat{b}^* \right\} \cdot \frac{\nabla b}{N^2} = -\frac{f}{N^2} \frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{u}}_\perp \hat{b}^* \right\} (\mathbf{e}_\parallel \cdot \partial_z \mathbf{u}) + \frac{\rho_r}{2N^2} \Re \left\{ \hat{\mathbf{u}}_\perp \hat{b}^* \right\} (\mathbf{e}_\perp \cdot \partial_z \mathbf{f}_{ag}). \quad (\text{A.73})$$

The Eliassen-Palm flux relation (A.61) and its connection to the pseudo-momentum flux is used to obtain

$$\begin{aligned} \rho_r \langle \mathbf{u}' w' \rangle_w \cdot \partial_z \mathbf{u} + \rho_r \langle \mathbf{u}' b' \rangle_w \cdot \frac{\nabla_h b}{N^2} &= \frac{E_w}{\hat{\omega}} \mathbf{k}_h \hat{c}_{gz} \cdot \partial_z \mathbf{u} \\ &+ \rho_r \langle \mathbf{u}' b' \rangle_w \cdot \frac{\partial_z \mathbf{f}_{ag}}{N^2}. \end{aligned} \quad (\text{A.74})$$

The relation above is quite important as it shows that the exchange of mechanical energy in background shear occurs through a superposition of vertical flux of horizontal momentum and horizontal entropy flux. In other words, for gravity waves affected by rotation the entropy flux partially cancels the momentum flux. Furthermore, the coupling to ageostrophic background forces builds an additional wave energy source.

Calculations in the 2nd Step:

For the horizontal stress tensor $\rho_r \langle \mathbf{u}' \mathbf{u}' \rangle_w$, the lowest order relation

$$-i \hat{\omega} \hat{\mathbf{u}} = -f \mathbf{e}_z \times \hat{\mathbf{u}} - i \mathbf{k}_h \hat{\phi} \quad (\text{A.75})$$

is used. When $\hat{\mathbf{u}}$ is rewritten as $\hat{\mathbf{u}} = \hat{\mathbf{u}}_{rot} + \hat{\mathbf{u}}_\phi$ with $\hat{\mathbf{u}}_{rot} = -if \mathbf{e}_z \times \hat{\mathbf{u}} / \hat{\omega}$ and $\hat{\mathbf{u}}_\phi = \mathbf{k}_h \hat{\phi} / \hat{\omega}$, for the second term one obtains

$$\frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{u}}_\phi \hat{\mathbf{u}}^* \right\} \cdot \nabla_h \mathbf{u} = \frac{\rho_r}{2} \frac{\mathbf{k}_h}{\hat{\omega}} \Re \left\{ \hat{\phi} \hat{\mathbf{u}}^* \right\} \cdot \nabla_h \mathbf{u} = \frac{E_w}{\hat{\omega}} \mathbf{k}_h \hat{c}_{gh} \cdot \nabla_h \mathbf{u}, \quad (\text{A.76})$$

where eq. (A.55) and the intrinsic horizontal group velocity \hat{c}_{gh} were used. The term $\hat{\mathbf{u}}_{rot} = -if / \hat{\omega} (-\hat{v}, \hat{u}, 0)^T$ contributes to

$$\frac{\rho_r}{2} \Re \left\{ \hat{\mathbf{u}}_{rot} \hat{\mathbf{u}}^* \right\} \cdot \nabla_h \mathbf{u} = \frac{\rho_r}{2} \frac{f}{\hat{\omega}} \Im \left\{ \hat{u} \hat{v}^* \right\} (\mathbf{e}_\lambda \mathbf{e}_\lambda + \mathbf{e}_\varphi \mathbf{e}_\varphi) \cdot \nabla_h \mathbf{u}, \quad (\text{A.77})$$

where $\Re \{ i |\hat{u}|^2 \} = \Re \{ i |\hat{v}|^2 \} = 0$ and $\Re \{ i \hat{v} \hat{u}^* \} = -\Im \{ \hat{v} \hat{u}^* \} = \Im \{ \hat{u} \hat{v}^* \}$ were utilized. With the help of eq. (4.71), (4.76) and (A.54),

$$\frac{\rho_r}{2} \frac{f}{\hat{\omega}} \Im \left\{ \hat{u} \hat{v}^* \right\} = \frac{\rho_r}{2} \frac{f^2 m^2}{k_h^2 N^4} |\hat{b}|^2 = \frac{f^2 m^2}{\hat{\omega}^2 |\mathbf{k}|^2} E_w \quad (\text{A.78})$$

is obtained. The diagonal part of $\nabla_h \mathbf{u}$ is $\nabla_h \cdot \mathbf{u}$ where the metric terms have been carefully taken into account. Hence finally,

$$\rho_r \langle \mathbf{u}' \mathbf{u}' \rangle_w \cdot \nabla_h \mathbf{u} = \frac{E_w}{\hat{\omega}} \mathbf{k}_h \hat{c}_{gh} \cdot \nabla_h \mathbf{u} + \frac{f^2 m^2}{\hat{\omega}^2 |\mathbf{k}|^2} E_w \nabla_h \cdot \mathbf{u}. \quad (\text{A.79})$$

Calculations in the 3rd Step:

The intrinsic GW frequency obeys $\hat{\omega} = \omega - \mathbf{k}_h \cdot \mathbf{u}$. Hence, the ray equation for $\hat{\omega}$ is

$$d_t \hat{\omega} = d_t \omega - d_t \mathbf{k}_h \cdot \mathbf{u} - \mathbf{k}_h \cdot d_t \mathbf{u}. \quad (\text{A.80})$$

For the $d_t \mathbf{k}_h$ -equation, eq. (4.99) with the decomposition (4.100) are used to get

$$d_t \mathbf{k}_h = -c_{gi} \nabla_h \mathbf{e}_i \cdot \mathbf{k} - \frac{\partial \omega}{\partial \Lambda_n} \nabla_h \Lambda_n, \quad (\text{A.81})$$

where Λ_n are u , v and f for $n = 1, 2, 3$, and $\mathbf{e}_1 = \mathbf{e}_\lambda$ and $\mathbf{e}_2 = \mathbf{e}_\varphi$. Using (4.104) and (4.89),

$$k \nabla_h u + l \nabla_h v = \nabla_h \mathbf{u} \cdot \mathbf{k}_h + \frac{\tan \varphi}{a_E} (kv - lu) \mathbf{e}_\lambda \quad (\text{A.82})$$

and $k \hat{c}_{g\varphi} = l \hat{c}_{g\lambda}$ is exploited to obtain

$$d_t \mathbf{k}_h = -\nabla_h \mathbf{u} \cdot \mathbf{k}_h - \frac{f m^2}{\hat{\omega} |\mathbf{k}|^2} \beta \mathbf{e}_\varphi. \quad (\text{A.83})$$

Therefore, with $d_t \omega = \mathbf{k}_h \cdot \partial_t \mathbf{u}$,

$$d_t \hat{\omega} = -\mathbf{k}_h \hat{c}_g \cdot \nabla \mathbf{u} + \frac{f m^2}{\hat{\omega} |\mathbf{k}|^2} \beta v, \quad (\text{A.84})$$

in which obviously no metric corrections appear explicitly.

Combination of energy sources

The wave exchanges energy with the background flow via shear and buoyancy production,

$$S_E = -\rho_r \langle \mathbf{u}' \mathbf{v}' \rangle_w \cdot \nabla \mathbf{u} - \rho_r \langle \mathbf{u}' \mathbf{b}' \rangle_w \cdot \frac{\nabla_h b}{N^2}. \quad (\text{A.85})$$

As shown above, the non-dissipative source/sink term S_E can be reformulated as

$$S_E = -\frac{E_w}{\hat{\omega}} \mathbf{k}_h \hat{c}_g \cdot \nabla \mathbf{u} - \rho_r \langle \mathbf{u}' \mathbf{b}' \rangle_w \cdot \frac{\partial_z \mathbf{f}_{ag}}{N^2} - \frac{f^2 m^2}{\hat{\omega}^2 |\mathbf{k}|^2} E_w \nabla_h \cdot \mathbf{u}. \quad (\text{A.86})$$

Using eq. (A.70), the last source term is

$$-\frac{f^2 m^2}{\hat{\omega}^2 |\mathbf{k}|^2} E_w \nabla_h \cdot \mathbf{u} = \frac{E_w}{\hat{\omega}} \frac{f m^2}{\hat{\omega} |\mathbf{k}|^2} (\beta v - (\nabla_h \times \mathbf{f}_{ag}) \cdot \mathbf{e}_z). \quad (\text{A.87})$$

All parts are combined together and eq. (A.84) is used, the wave energy source reduces to

$$S_E = \frac{E_w}{\hat{\omega}} d_t \hat{\omega} + \frac{f}{\hat{\omega}^2} \frac{m}{|\mathbf{k}|^2} E_w \{ k_h (\mathbf{e}_\perp \cdot \partial_z \mathbf{f}_{ag}) - m (\nabla_h \times \mathbf{f}_{ag}) \cdot \mathbf{e}_z \}. \quad (\text{A.88})$$

In the limit of small mean Rossby number $\text{Ro}_m \sim \varepsilon$, the energy source due to ageostrophic motion is negligible and at $O(\varepsilon)$:

$$S_E = \frac{E_w}{\hat{\omega}} d_t \hat{\omega}. \quad (\text{A.89})$$

With $A = E_w/\hat{\omega}$, the conservation of wave action (4.97) is obtained. For inertia-gravity waves, the scale-separation assumption is problematic when the interaction with diurnal tides is considered.

A.5. Conservation of GW angular momentum in a resting, isothermal basic state

In the following, a resting, isothermal basic state with $\mathbf{u} = 0$ and $N = \text{const}$ is considered. It is further assumed that the GW properties are homogeneous in time and space, so that $\nabla \cdot \mathbf{c}_g = 0$ holds, and the beta-effect is omitted. Then, as discussed above, the wave action density is constant along a ray ($d_t A = 0$) for non-dissipative wave motion. In a shallow atmosphere, the GW angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{k} A$, where \mathbf{r} denotes the position vector from the center of earth, is approximated by

$$\mathbf{L} = a_E \mathbf{e}_z \times \mathbf{k} A = a_E (-l \mathbf{e}_\lambda + k \mathbf{e}_\varphi) A. \quad (\text{A.90})$$

The change of \mathbf{L} along the ray is given by

$$d_t \mathbf{L} = a_E (-d_t(l \mathbf{e}_\lambda) + d_t(k \mathbf{e}_\varphi)) A. \quad (\text{A.91})$$

For the first term,

$$-d_t(l \mathbf{e}_\lambda) = -d_t l \mathbf{e}_\lambda - l d_t \mathbf{e}_\lambda = -d_t l \mathbf{e}_\lambda - l \hat{\mathbf{c}}_g \cdot \nabla \mathbf{e}_\lambda \quad (\text{A.92})$$

$$= \frac{k \tan \varphi}{a_E} \hat{\mathbf{c}}_{g\lambda} \mathbf{e}_\lambda - \frac{l \tan \varphi}{a_E} \hat{\mathbf{c}}_{g\lambda} \mathbf{e}_\varphi \quad (\text{A.93})$$

results, and for the second,

$$d_t(k \mathbf{e}_\varphi) = d_t k \mathbf{e}_\varphi + k d_t \mathbf{e}_\varphi = d_t k \mathbf{e}_\varphi + k \hat{\mathbf{c}}_g \cdot \nabla \mathbf{e}_\varphi \quad (\text{A.94})$$

$$= \frac{k \tan \varphi}{a_E} \hat{\mathbf{c}}_{g\varphi} \mathbf{e}_\varphi - \frac{k \tan \varphi}{a_E} \hat{\mathbf{c}}_{g\lambda} \mathbf{e}_\lambda, \quad (\text{A.95})$$

where the ray equations (4.109) and (4.110) as well as the gradients of the unit vectors (4.104) were used. Therefore,

$$d_t \mathbf{L} = a_E (k \hat{\mathbf{c}}_{g\varphi} - l \hat{\mathbf{c}}_{g\lambda}) \frac{\tan \varphi}{a_E} = 0, \quad (\text{A.96})$$

where $k \hat{\mathbf{c}}_{g\varphi} = l \hat{\mathbf{c}}_{g\lambda}$, is true for gravity waves. The GW angular momentum is constant along the rays for gravity waves within a resting, isothermal basic state.

A.6. RAPAGI: the numerical implementation

The RAY parameterization of Gravity-wave Impacts (RAPAGI) is a fast numerical model for solving the ray tracing equations on a spherical globe. For the direct use of GCM data, it is favorable to identify the position \mathbf{x} of the wave parcel with spherical coordinates λ , φ and an altitude \hat{z} defined as the globally averaged geopotential height on surfaces of constant vertical hybrid coordinate η . As each change of z along the ray is expressed as

$$d_t z = \partial_t z + (d_t \lambda) \partial_\lambda z + (d_t \varphi) \partial_\varphi z + (d_t \hat{z}) \partial_{\hat{z}} z, \quad (\text{A.97})$$

the evolution of a ray point is given by

$$d_t \lambda = \frac{c_{gx}}{a_E \cos \varphi}, \quad (\text{A.98})$$

$$d_t \varphi = \frac{c_{gy}}{a_E}, \quad (\text{A.99})$$

$$d_t \hat{z} = \frac{c_{gz} - \partial_t z - \mathbf{c} \cdot \nabla_h z}{\partial_z z}, \quad (\text{A.100})$$

where the components of group velocity \mathbf{c}_g are given in eqns. (4.78)-(4.80). This facilitates inter-model communication. The partial derivatives in eqns. (4.108)- (4.111) are given in a coordinate system with geometric altitude z , while they are usually calculated from the large-scale flow in generalized coordinates $\{\lambda, \varphi, \hat{z}(\eta)\}$. The transformation between the two

$$\nabla_h \mathbf{u}|_z = \nabla_h \mathbf{u}|_{\hat{z}} - \nabla_h z|_{\hat{z}} \frac{\partial_z \mathbf{u}}{\partial_z z}, \quad (\text{A.101})$$

$$\partial_z \mathbf{u} = \frac{\partial_{\hat{z}} \mathbf{u}}{\partial_{\hat{z}} z}, \quad (\text{A.102})$$

and similar expressions for t were taken into account in the ray-tracing simulations.

The time-integration of eqns. (4.108) - (4.111) was done in two stages. First, an estimate $\{\omega_{n+1}^*, \mathbf{k}_{n+1}^*\}$ for time $(n+1)\Delta t$ is obtained using the Heun scheme with a fixed time step of $\Delta t = 5$ min, for which convergence has been verified regarding the given GW ensemble. Second, an optimization technique is used to adaptively change all ray properties until the dispersion relation is satisfied. For $\beta_i \ll 1$, the corrected estimates

$$\omega_{n+1} = \omega_n + \Delta\omega (1 + \beta_0) \quad \text{with} \quad \Delta\omega = \omega_{n+1}^* - \omega_n, \quad (\text{A.103})$$

$$k_{i,n+1} = k_{i,n} + \Delta k_i (1 + \beta_i) \quad \text{with} \quad \Delta k_i = k_{i,n+1}^* - k_{i,n} \quad (\text{A.104})$$

fulfill dispersion relation (4.76). In the optimization step, the functional

$$\mathcal{G} = \frac{1}{2} \sum_{i=0}^3 \beta_i^2 + \hat{\beta} (\omega(\mathbf{k}_{n+1}, \mathbf{\Lambda}_{n+1}) - \omega_{n+1}) \quad (\text{A.105})$$

is minimized. The variation of \mathcal{G} with respect to β_i gives

$$\beta_0 = \hat{\beta} \Delta\omega, \quad (\text{A.106})$$

$$\beta_i = -\hat{\beta} c_{gi,n+1} \Delta k_i \quad (\text{A.107})$$

for $i = 1, 2, 3$. Inserting (A.106) and (A.107) in the dispersion relation results in a non-linear equation for the Lagrangian multiplier $\hat{\beta}$ which is solved numerically via the Newton method. Therefore, in the two-stage scheme, the additional information gained by the ω -equation (4.108) is used to correct numerical errors and stabilize the method.

Each time step, new ray points are injected at $\hat{z}_B = 20$ km and after a warming time of one day most of the model domain in which GW propagation is possible is filled with ray points. Ray points are randomly removed when their number exceeds 32 in a grid box of the large-scale model. All background quantities are interpolated to the ray position via a linear polygonal interpolation. A distance-weighted interpolation and running median average is used to obtain smooth GW properties on the large-scale mesh. In particular, the forcing terms in (4.112) are calculated this way on the mesh and interpolated back to the ray positions.

For ray integrations, no explicit test of WKB validity is performed. Only rays which cross the extreme thresholds of 100 km vertical wavelength and 10 days intrinsic period are removed from the model run. As noted by *Sartelet* [2003], ray theory performs remarkably well even if the scale separation assumption is not satisfied.

Bibliography

- Achatz, U. (2007), Gravity-wave breaking: Linear and primary nonlinear dynamics, *Adv. Space Res.*, 40(6), 719 – 733.
- Achatz, U., N. Grieger, and H. Schmidt (2008), Mechanisms controlling the diurnal solar tide: Analysis using a GCM and a linear model, *J. Geophys. Res.*, 113, A08,303.
- Achatz, U., R. Klein, and F. Senf (2010), Gravity waves, scale asymptotics and the pseudo-incompressible equations, *J. Fluid Mech.*, 663, 120–147.
- Achatz, U., F. Senf, and N. Grieger (2012), Solar tides in the middle atmosphere: Interactions with the zonal- mean flow, planetary waves and gravity waves, in *Climate And Weather of the Sun-Earth System (CAWSES): Highlights from a priority program*, edited by F.-J. Lübken, Springer, Dordrecht, The Netherlands.
- Akmaev, R. A. (2001), Simulation of large-scale dynamics in the mesosphere and lower thermosphere with the Doppler-spread parameterization of gravity waves 2. Eddy mixing and the diurnal tide, *J. Geophys. Res.*, 106, 1205–1214.
- Akmaev, R. A. (2007), On the energetics of mean-flow interactions with thermally dissipating gravity waves, *J. Geophys. Res.*, 112, D11,125.
- Alexander, M., et al. (2010), Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models, *Quart. J. R. Met. Soc.*, 136(650), 1103–1124.
- Andrews, D. (1981), A note on potential energy density in a stratified compressible fluid, *J. Fluid Mech.*, 107, 227–236.
- Andrews, D. G., and M. E. McIntyre (1978), An exact theory of nonlinear waves on a Lagrangian-mean flow, *J. Fluid Mech.*, 89, 609–646.
- Andrews, D. G., and M. E. McIntyre (1978), On wave-action and its relatives, *J. Fluid Mech.*, 89(04), 647–664.
- Andrews, D. G., J. R. Holton, and C. B. Leovy (1987), *Middle atmosphere dynamics.*, Academic Press, New York.
- Bannon, P. (1996), On the anelastic approximation for a compressible atmosphere, *J. Atmos. Sci.*, 53(23), 3618–3628.
- Bannon, P. R. (2004), Lagrangian Available Energetics and Parcel Instabilities, *J. Atmos. Sci.*, 61(14), 1754–1767.
- Bannon, P. R. (2005), Eulerian Available Energetics in Moist Atmospheres, *J. Atmos. Sci.*, 62(12), 4238–4252.
- Becker, E. (2001), Symmetric Stress Tensor Formulation of Horizontal Momentum Diffusion in Global Models of Atmospheric Circulation., *J. Atmos. Sci.*, 58, 269–282.

- Becker, E. (2003a), Frictional Heating in Global Climate Models, *Mon. Weather Rev.*, *131*(3), 508–520.
- Becker, E. (2003b), Energetics of turbulent momentum diffusion and gravity wave breakdown in general circulation models of the atmosphere, Habilitation, Rostock University.
- Becker, E. (2004), Direct heating rates associated with gravity wave saturation, *J. Atmos. Solar-Terr. Phys.*, *66*, 683–696.
- Becker, E. (2012), Dynamical Control of the Middle Atmosphere, *Space Sci. Rev.*, pp. 1–32, 10.1007/s11214-011-9841-5.
- Becker, E., and U. Burkhardt (2007), Nonlinear Horizontal Diffusion for GCMs, *Mon. Weather Rev.*, *135*(4), 1439–1454.
- Becker, E., and G. Schmitz (2002), Energy Deposition and Turbulent Dissipation Owing to Gravity Waves in the Mesosphere., *J. Atmos. Sci.*, *59*, 54–68.
- Becker, E., and G. Schmitz (2003), Climatological Effects of Orography and Land-Sea Heating Contrasts on the Gravity Wave-Driven Circulation of the Mesosphere., *J. Atmos. Sci.*, *60*, 103–118.
- Beres, J. H., R. R. Garcia, B. A. Boville, and F. Sassi (2005), Implementation of a gravity wave source spectrum parameterization dependent on the properties of convection in the Whole Atmosphere Community Climate Model (WACCM), *J. Geophys. Res.*, *110*, D10,108.
- Boville, B. A., and C. S. Bretherton (2003), Heating and Kinetic Energy Dissipation in the NCAR Community Atmosphere Model, *J. Climate*, *16*(23), 3877–3887.
- Bretherton, F. P. (1969), Momentum transport by gravity waves, *Quart. J. R. Met. Soc.*, *95*, 213–243.
- Bretherton, F. P., and C. J. R. Garrett (1968), Wavetrains in Inhomogeneous Moving Media, *Proc. R. Soc., London A*, *302*, 529–554.
- Broutman, D. (1984), The focusing of short internal waves by an inertial wave, *Geophys. Astrophys. Fluid Dyn.*, *30*(3), 199–225.
- Broutman, D., and R. H. J. Grimshaw (1988), The energetics of the interaction between short small-amplitude internal waves and inertial waves, *J. Fluid Mech.*, *196*, 93–106.
- Broutman, D., and W. R. Young (1986), On the interaction of small-scale oceanic internal waves with near-inertial waves, *J. Fluid Mech.*, *166*, 341–358.
- Broutman, D., J. W. Rottman, and S. D. Eckermann (2004), Ray Methods for Internal Waves in the Atmosphere and Ocean, *Annu. Rev. Fluid Mech.*, *36*, 233–253.
- Bühler, O. (2009), *Waves and Mean Flows*, Cambridge University Press.
- Bühler, O., and M. E. McIntyre (1998), On non-dissipative wave mean interactions in the atmosphere or oceans, *Journal of Fluid Mechanics*, *354*, 301–343.
- Bühler, O., and M. E. McIntyre (2005), Wave capture and wave-vortex duality, *J. Fluid Mech.*, *534*, 67–95.

- Burkhardt, U., and E. Becker (2006), A Consistent Diffusion Dissipation Parameterization in the ECHAM Climate Model, *Mon. Weather Rev.*, *134*, 1194–1204.
- Chapman, S., and R. Lindzen (1970), *Atmospheric tides. Thermal and gravitational*, Reidel, Dordrecht, Holland.
- Davies, T., A. Staniforth, N. Wood, and J. Thuburn (2003), Validity of anelastic and other equation sets as inferred from normal-mode analysis, *Quart. J. R. Met. Soc.*, *129*, 2761–2775.
- Dunkerton, T. J. (1981), Wave Transience in a Compressible Atmosphere. Part I: Transient Internal Wave, Mean-Flow Interaction., *J. Atmos. Sci.*, *38*, 281–297.
- Dunkerton, T. J. (1982), Stochastic parameterization of gravity wave stresses, *J. Atmos. Sci.*, *39*, 1711–1725.
- Dunkerton, T. J. (1984), Inertia-gravity waves in the stratosphere, *J. Atmos. Sci.*, *41*, 3396–3404.
- Dunkerton, T. J. (1989), Theory of internal gravity wave saturation, *Pageoph.*, *130*, 373–397.
- Dunkerton, T. J., and N. Butchart (1984), Propagation and selective transmission of internal gravity waves in a sudden warming, *J. Atmos. Sci.*, *41*, 1443–1460.
- Durrán, D. (1999), *Numerical methods for wave equations in geophysical fluid dynamics*, Springer, New York.
- Durrán, D., and A. Arakawa (2007), Generalizing the Boussinesq approximation to stratified compressible flow, *C. R. Mec.*, *335*(9-10), 655–664.
- Durrán, D. R. (1989), Improving the Anelastic Approximation., *J. Atmos. Sci.*, *46*, 1453–1461.
- Durrán, D. R. (2008), A physically motivated approach for filtering acoustic waves from the equations governing compressible stratified flow, *J. Fluid Mech.*, *601*, 365–379.
- Eckermann, S. D. (1992), Ray-Tracing Simulation of the Global Propagation of Inertia Gravity Waves Through the Zonally Averaged Middle Atmosphere, *J. Geophys. Res.*, *97*(D14), 15,849–15,866.
- Eckermann, S. D., and C. J. Marks (1996), An idealized ray model of gravity wave-tidal interactions, *J. Geophys. Res.*, *101*, 21,195–21,212.
- Forbes, J. M., J. Gu, and S. Miyahara (1991), On the interactions between gravity waves and the diurnal propagating tide, *Planet. Space Sci.*, *39*, 1249–1257.
- Fritts, D. C. (1984), Gravity wave saturation in the middle atmosphere - A review of theory and observations, *Rev. Geophys. Space Phys.*, *22*, 275–308.
- Fritts, D. C., and M. J. Alexander (2003), Gravity wave dynamics and effects in the middle atmosphere, *Rev. Geophys.*, *41*, 1003.
- Fritts, D. C., and R. A. Vincent (1987), Mesospheric momentum flux studies at Adelaide, Australia - Observations and a gravity wave-tidal interaction model, *J. Atmos. Sci.*, *44*, 605–619.

- Gossard, E. E., and W. H. Hooke (1975), *Waves in the atmosphere: Atmospheric infrasound and gravity waves - Their generation and propagation*, vol. 2, 470 pp., Elsevier Scientific Publishing Co., Amsterdam.
- Grimshaw, R. (1984), Wave action and wave-mean flow interaction, with application to stratified shear flows, *Annu. Rev. Fluid Mech.*, *16*, 11–44.
- Grimshaw, R. H. J. (1975a), Nonlinear internal gravity waves in a rotating fluid, *J. Fluid Mech.*, *71*, 497–512.
- Grimshaw, R. H. J. (1975b), Internal gravity waves: critical layer absorption in a rotating fluid, *J. Fluid Mech.*, *70*, 287–304.
- Hasha, A., O. Bühler, and J. Scinocca (2008), Gravity Wave Refraction by Three-Dimensionally Varying Winds and the Global Transport of Angular Momentum, *J. Atmos. Sci.*, *65*, 2892–2906.
- Hayes, W. D. (1970), Kinematic Wave Theory, *Proc. R. Soc., London A*, *320*, 209–226.
- Hertzog, A., C. Souprayen, and A. Hauchecorne (2001), Observation and backward trajectory of an inertio-gravity wave in the lower stratosphere, *Ann. Geophys.*, *19*, 1141–1155.
- Hertzog, A., C. Souprayen, and A. Hauchecorne (2002), Eikonal simulations for the formation and the maintenance of atmospheric gravity wave spectra, *J. Geophys. Res.*, *107*, 4145.
- Holm, D. D. (1999), Fluctuation effects on 3D Lagrangian mean and Eulerian mean fluid motion, *Physica D*, *133*, 215–269.
- Holm, D. D. (2002), Averaged Lagrangians and the mean effects of fluctuations in ideal fluid dynamics, *Physica D*, *170*, 253–286.
- Holton, J. (2004), *An introduction to dynamic meteorology*, Elsevier Academic Press, San Diego, London.
- Holton, J. R. (1982), The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere, *J. Atmos. Sci.*, *39*, 791–799.
- Jones, W. L. (1967), Propagation of internal gravity waves in fluids with shear flow and rotation, *J. Fluid Mech.*, *30*, 439–448.
- Klein, R. (2000), Asymptotic analyses for atmospheric flows and the construction of asymptotically adaptive numerical methods, *Z. angew. Math. Mech.*, *80*(11-12), 765–785.
- Klein, R. (2008), An unified approach to meteorological modelling based on multiple-scales asymptotics, *Adv. Geoscience*, *15*, 23–33.
- Klein, R., and O. Pauluis (2012), Thermodynamic Consistency of a Pseudo-Incompressible Approximation for General Equations of State, *J. Atmos. Sci.*, (*in print*).
- Lange, H.-J. (2002), *Physik des Wetters und des Klimas*, Dietrich Reimer Verlag, Berlin.

- Lieberman, R. S., D. M. Riggin, D. A. Ortland, S. W. Nesbitt, and R. A. Vincent (2007), Variability of mesospheric diurnal tides and tropospheric diurnal heating during 1997-1998, *J. Geophys. Res.*, *112*, D20,110.
- Lieberman, R. S., D. A. Ortland, D. M. Riggin, Q. Wu, and C. Jacobi (2010), Momentum budget of the migrating diurnal tide in the mesosphere and lower thermosphere, *J. Geophys. Res.*, *115*, D20,105.
- Lighthill, J. (1978), *Waves in fluids*, Cambridge University Press.
- Lindzen, R. (1966), On the theory of the diurnal tide, *Mon. Weather Rev.*, *94*(5), 295–301.
- Lindzen, R. A. (1990), *Dynamics in Atmospheric Physics*, Cambridge University Press.
- Lindzen, R. S. (1967), Thermally driven diurnal tide in the atmosphere, *Quart. J. R. Met. Soc.*, *93*(395), 18–42.
- Lindzen, R. S. (1981), Turbulence and stress owing to gravity wave and tidal breakdown, *J. Geophys. Res.*, *86*, 9707–9714.
- Lipps, F. B. (1990), On the Anelastic Approximation for Deep Convection., *J. Atmos. Sci.*, *47*, 1794–1798.
- Lipps, F. B., and R. S. Hemler (1982), A Scale Analysis of Deep Moist Convection and Some Related Numerical Calculations., *J. Atmos. Sci.*, *39*, 2192–2210.
- Liu, H., et al. (2007), Comparative study of short-term diurnal tidal variability, *J. Geophys. Res.*, *112*, D18,108.
- Lorenz, E. N. (1955), Available potential energy and the maintenance of the general circulation, *Tellus*, *7*, 157–167.
- Lu, W., and D. C. Fritts (1993), Spectral estimates of gravity wave energy and momentum fluxes. Part 3: Gravity wave-tidal interactions, *J. Atmos. Sci.*, *50*, 3714–3727.
- Marks, C. J., and S. D. Eckermann (1995), A Three-Dimensional Nonhydrostatic Ray-Tracing Model for Gravity Waves: Formulation and Preliminary Results for the Middle Atmosphere., *J. Atmos. Sci.*, *52*, 1959–1984.
- Mayr, H. G., J. G. Mengel, K. L. Chan, and H. S. Porter (1999), Seasonal variations and planetary wave modulation of diurnal tides influenced by gravity waves, *Adv. Space Res.*, *24*, 1541–1544.
- Mayr, H. G., J. G. Mengel, K. L. Chan, and H. S. Porter (2001), Mesosphere dynamics with gravity wave forcing: Part I. Diurnal and semi-diurnal tides, *J. Atmos. Solar-Terr. Phys.*, *63*, 1851–1864.
- McIntyre, M. E. (1980), An introduction to the generalized Lagrangian-mean description of wave, mean-flow interaction, *Pageoph.*, *118*, 152–176.
- McIntyre, M. E. (1981), On the ‘wave-momentum’ myth, *J. Fluid Mech.*, *106*, 331–347.
- McLandress, C. (1997), Sensitivity studies using the Hines and Fritts gravity-wave drag parameterizations, in *Gravity Wave Processes and their Parameterization in Global Climate Models*, edited by K. Hamilton, pp. 245–256, Springer, New York.

- McLandress, C. (1998), On the importance of gravity waves in the middle atmosphere and their parameterization in general circulation models., *J. Atmos. Solar-Terr. Phys.*, *60*, 1357–1383.
- McLandress, C. (2002), The Seasonal Variation of the Propagating Diurnal Tide in the Mesosphere and Lower Thermosphere. Part I: The Role of Gravity Waves and Planetary Waves., *J. Atmos. Sci.*, *59*, 893–906.
- Medvedev, A. S., and G. P. Klaassen (1995), Vertical evolution of gravity wave spectra and the parameterization of associated wave drag, *J. Geophys. Res.*, *100*, 25,841–25,854.
- Messiah, A. (1961), *Quantum mechanics*, North-Holland Pub., Amsterdam.
- Meyer, C. K. (1999), Gravity wave interactions with the diurnal propagating tide, *J. Geophys. Res.*, *104*, 4223–4240.
- Miyahara, S., and J. Forbes (1991), Interactions between gravity waves and the diurnal tide in the mesosphere and lower thermosphere, *J. Met. Soc. Japan*, *69*(5), 523–531.
- Nance, L. B., and D. R. Durran (1997), A Modeling Study of Nonstationary Trapped Mountain Lee Waves. Part I: Mean-Flow Variability., *J. Atmos. Sci.*, *54*, 2275–2291.
- Nappo, C. J. (2002), *An introduction to atmospheric gravity waves*, Academic Press, San Diego, US.
- Norton, W. A., and J. Thuburn (1999), Sensitivity of mesospheric mean flow, planetary waves, and tides to strength of gravity wave drag, *J. Geophys. Res.*, *104*, 30,897–30,912.
- Ogura, Y., and N. A. Phillips (1962), Scale Analysis of Deep and Shallow Convection in the Atmosphere., *J. Atmos. Sci.*, *19*, 173–179.
- Ortland, D. A. (2005a), Generalized Hough Modes: The Structure of Damped Global-Scale Waves Propagating on a Mean Flow with Horizontal and Vertical Shear., *J. Atmos. Sci.*, *62*, 2674–2683.
- Ortland, D. A. (2005b), A Study of the Global Structure of the Migrating Diurnal Tide Using Generalized Hough Modes., *J. Atmos. Sci.*, *62*, 2684–2702.
- Ortland, D. A., and M. J. Alexander (2006), Gravity wave influence on the global structure of the diurnal tide in the mesosphere and lower thermosphere, *J. Geophys. Res.*, *111*, A10S10.
- Pedlosky, J. (1982), *Geophysical fluid dynamics*, Springer-Verlag, New York and Berlin.
- Pedlosky, J. (2003), *Waves in the ocean and atmosphere: introduction to wave dynamics*, 260 pp pp., Springer: Berlin.
- Phillips, N. A. (1966), The Equations of Motion for a Shallow Rotating Atmosphere and the ‘Traditional Approximation’, *J. Atmos. Sci.*, *23*, 626–627.
- Plumb, R. (1983), A new look at the energy cycle, *J. Atmos. Sci.*, *40*(7), 1669–1688.
- Preusse, P., S. D. Eckermann, M. Ern, J. Oberheide, R. H. Picard, R. G. Roble, M. Riese, J. M. Russell, and M. G. Mlynczak (2009), Global ray tracing simulations of the SABER gravity wave climatology, *J. Geophys. Res.*, *114*, D08,126.

- Richter, J. H., F. Sassi, and R. R. Garcia (2010), Toward a Physically Based Gravity Wave Source Parameterization in a General Circulation Model, *J. Atmos. Sci.*, *67*, 136–156.
- Sartelet, K. N. (2003), Wave propagation inside an inertia wave. Part I: Role of time dependence and scale separation, *J. Atmos. Sci.*, *60*, 1433–1447.
- Schmidt, H., et al. (2006), The HAMMONIA Chemistry Climate Model: Sensitivity of the Mesopause Region to the 11-Year Solar Cycle and CO₂ Doubling, *J. Climate*, *19*, 3903–3931.
- Senf, F., and U. Achatz (2011), On the impact of middle-atmosphere thermal tides on the propagation and dissipation of gravity waves, *J. Geophys. Res.*, *116*, D24,110.
- Senf, F., and U. Achatz (2012), On generalized linear flows: From compressible to sound-proof dynamics, (*unpublished manuscript*).
- Shaw, T. A., and T. G. Shepherd (2007), Angular Momentum Conservation and Gravity Wave Drag Parameterization: Implications for Climate Models, *J. Atmos. Sci.*, *64*(1), 190–203.
- Shaw, T. A., and T. G. Shepherd (2009), A Theoretical Framework for Energy and Momentum Consistency in Subgrid-Scale Parameterization for Climate Models, *J. Atmos. Sci.*, *66*, 3095–3114.
- Shepherd, T. G., and T. A. Shaw (2004), The Angular Momentum Constraint on Climate Sensitivity and Downward Influence in the Middle Atmosphere, *J. Atmos. Sci.*, *61*(23), 2899–2908.
- Smith, R. (1979), The Influence of the Earth’s Rotation on Mountain Wave Drag., *J. Atmos. Sci.*, *36*, 177–180.
- Song, I., and H. Chun (2008), A Lagrangian Spectral Parameterization of Gravity Wave Drag Induced by Cumulus Convection, *J. Atmos. Sci.*, *65*, 1204–1224.
- Song, I.-S., and H.-Y. Chun (2005), Momentum Flux Spectrum of Convectively Forced Internal Gravity Waves and Its Application to Gravity Wave Drag Parameterization. Part I: Theory, *J. Atmos. Sci.*, *62*(1), 107–124.
- Sonmor, L. J., and G. P. Klaassen (2000), Mechanisms of gravity wave focusing in the middle atmosphere, *J. Atmos. Sci.*, *57*, 493–510.
- Vadas, S. L., and D. C. Fritts (2005), Thermospheric responses to gravity waves: Influences of increasing viscosity and thermal diffusivity, *J. Geophys. Res.*, *110*, D15,103.
- Walterscheid, R. L. (2000), Propagation of small-scale gravity waves through large-scale internal wave fields: Eikonal effects at low-frequency approximation critical levels, *J. Geophys. Res.*, *105*(D14).
- Watanabe, S., and S. Miyahara (2009), Quantification of the gravity wave forcing of the migrating diurnal tide in a gravity wave-resolving general circulation model, *J. Geophys. Res.*, *114*(D7), D07,110.
- Wilhelmson, R., and Y. Ogura (1972), The Pressure Perturbation and the Numerical Modeling of a Cloud, *J. Atmos. Sci.*, *29*(7), 1295–1307.

- Yuan, T., C.-Y. She, D. A. Krueger, F. Sassi, R. Garcia, R. G. Roble, H.-L. Liu, and H. Schmidt (2008), Climatology of mesopause region temperature, zonal wind, and meridional wind over Fort Collins, Colorado (41N, 105W), and comparison with model simulations, *J. Geophys. Res.*, *113*, D03,105.
- Zhong, L., L. J. Sonmor, A. H. Manson, and C. E. Meek (1995), The influence of time-dependent wind on gravity-wave propagation in the middle atmosphere, *Ann. Geophys.*, *13*, 375–394.

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides Statt, dass ich die vorgelegte Dissertation selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Fabian Senf

Leipzig, den 1. März 2012

Danksagung

Liebe Familie, ich möchte mich bei Euch dafür bedanken, dass Ihr mich immer bei all meinen Entscheidungen mit Geduld und Zuspruch unterstützt habt.

Liebe Anja, ich möchte mich bei Dir für die vielen schönen letzten Jahre bedanken und für die wunderbaren Kinder, die Du mir geschenkt hast.

Lieber Ulrich, lieber Erich, ich möchte mich bei Euch bedanken, dass Ihr mir diese Arbeit ermöglicht habt. Ihr habt mich herzlich im Leibniz-Institut aufgenommen, mir die Dynamik der mittleren Atmosphäre schmackhaft gemacht und mich hingebungsvoll bei meinen Fragen und Problemen unterstützt. Auch die Situation durch Ulrichs Wechsel nach Frankfurt haben wir gut gemeinsam gemeistert. Ich danke Euch besonders dafür, dass Ihr mir meinen Entscheidungsfreiraum gelassen und mich nicht zwischen den Stühlen sitzen lassen habt.

Lieber Sebastian, Du warst ein netter Kollege und bist mir über die Zeit ein sehr guter Freund geworden. Auf unseren Fahrradtouren durch die Kühlung hatten wir immer viel Spaß. Dafür danke ich Dir.

Liebe Kühlungsborner Theoretiker und Modellierer, Ihr habt mir einen tollen Rahmen für meine Arbeit geschaffen. Jeder hatte ein offenes Ohr und hat mir tatkräftig bei der Lösung von kleinen oder größeren Problem geholfen.

Liebe Mitarbeiter des Leibniz-Instituts, während meiner Zeit am Institut herrschte eine besondere Atmosphäre und gegenseitige Wertschätzung der wissenschaftlichen Arbeit des anderen. Ich bin gerne jeden Tag auf die Kühlung gekommen und dazu hat jeder Mitarbeiter seinen Teil beigetragen.

Lieber Mark, Deine detaillierte und fachkundige Prüfung, bei der Du Dir vielleicht manchmal die Haare gerauft hast, hat mir sehr geholfen, die Arbeit zu verbessern. Dankeschön!

Liebe Frankfurter Kollegen, ich war gern bei Euch zu Besuch. Vielen Dank für Eure Gastfreundschaft.

Lebenslauf

Persönliche Daten

Fabian Senf

Diplom-Physiker

Geb. am 6. Februar 1979 in Leipzig,
ledig, zwei Kinder

Schulbildung

1985 - 1992 157. POS Leipzig „Paul Küstner“

1992 - 1997 42. Gymnasium Leipzig „Ludwig Uhland“

Zivildienst

1997 - 1998 Pflegeheim Thekla, Leipzig

Studium

1999 - 2006 Physik-Studium an der Universität Leipzig

2002 - 2006 Meteorologie-Studium an der Universität Leipzig

2005 - 2006 Diplomarbeit am Institut für Theoretische Physik bei Prof. Dr. Behn
Thema: „Rauschgetriebene Phasenübergänge im Nichtgleichgewicht:
Asymptotische Eigenschaften gekoppelter Stratonovich-Modelle“

2007 - 2011 Promotionsstudium an der Universität Rostock im Fachbereich
Physik

Berufsweg

2002 - 2004, studentische Hilfskraft am Leipziger Institut für Meteorologie bei
2006 Prof. Dr. Metz

2007 - 2011 wissenschaftlicher Mitarbeiter am Leibniz-Institut für
Atmosphärenphysik, Kühlungsborn bei Prof. Dr. Becker

seit 2011 wissenschaftlicher Mitarbeiter am Leibniz-Institut für
Troposphärenforschung, Leipzig