

Mathematics of Planet Earth 1

Carsten Eden
Armin Iske
Editors

Energy Transfers in Atmosphere and Ocean



 Springer

Mathematics of Planet Earth

Volume 1

Series editors

Ken Golden, The University of Utah, USA

Mark Lewis, University of Alberta, Canada

Yasumasa Nishiura, Tohoku University, Japan

Malcolm Sambridge, The Australian National University, Australia

Joseph Tribbia, National Center for Atmospheric Research, USA

Jorge Passamani Zubelli, Instituto Nacional de Matemática Pura e Aplicada, Brazil

Springer's Mathematics of Planet Earth series provides a variety of well-written books of a variety of levels and styles, highlighting the fundamental role played by mathematics in a huge range of planetary contexts on a global scale. Climate, ecology, sustainability, public health, diseases and epidemics, management of resources and risk analysis are important elements. The mathematical sciences play a key role in these and many other processes relevant to Planet Earth, both as a fundamental discipline and as a key component of cross-disciplinary research. This creates the need, both in education and research, for books that are introductory to and abreast of these developments.

More information about this series at <http://www.springer.com/series/13771>

Carsten Eden · Armin Iske
Editors

Energy Transfers in Atmosphere and Ocean



 Springer

The Springer logo consists of a stylized chess knight (horse) facing left, positioned above a horizontal line.

Editors

Carsten Eden
Center for Earth System Research
and Sustainability (CEN)
University of Hamburg
Hamburg, Germany

Armin Iske
Department of Mathematics
University of Hamburg
Hamburg, Germany

ISSN 2524-4264

Mathematics of Planet Earth

ISBN 978-3-030-05703-9

<https://doi.org/10.1007/978-3-030-05704-6>

ISSN 2524-4272 (electronic)

ISBN 978-3-030-05704-6 (eBook)

Library of Congress Control Number: 2018964245

Mathematics Subject Classification (2010): 35Q35, 35Q75, 35Q86, 65Z05, 86-08, 86A05, 86A10, 86A17

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book describes a recent effort combining interdisciplinary expertise within the Collaborative Research Centre

“Energy Transfers in Atmosphere and Ocean” (TRR 181),

which was funded by the German Research Foundation (DFG).

Energy transfers between the three dynamical regimes—small-scale turbulence, internal gravity waves and geostrophically balanced motion—are fundamental to the energy cycle of both the atmosphere and the ocean. Nonetheless, they remain poorly understood and quantified and have yet to be adequately represented in today’s climate models. Since interactions between the dynamical regimes ultimately link the smallest scales to the largest ones through a range of complex processes, understanding these interactions is essential to constructing atmosphere and ocean models and to predicting the future climate.

The current lack of understanding is reflected by energetically inconsistent models with relatively large biases, but also by inconsistencies of a numerical and mathematical nature. In TRR 181, recent efforts to overcome these deficiencies are combined, and new endeavours to understand dynamical interactions and to improve the consistency of ocean and atmosphere models are fostered.

The point of departure for TRR 181 is the recognition that the energy cycle is inconsistently represented in current climate models, i.e. that the models are energetically inconsistent. A primary example of this inconsistency is the effect of the dissipation of the unresolved internal gravity wave field in the ocean, which is parameterised in standard models by mixing of density with a prescribed diffusivity. Although this diffusivity is sometimes linked to resolved parameters or to the energy input into the internal wave field, a consistent description of the energetics of the internal wave field is usually lacking. The same is true for gravity wave drag parameterisation in atmosphere models and, more generally, for nearly all parameterisations which are used today: it applies to the parameterisation of dissipation of the (available) potential energy of the turbulent balanced flow in ocean models by (isopycnal) layer thickness diffusion, the dissipation of resolved kinetic energy by hyper-viscosity in atmosphere and ocean models, and the dissipation of energy

in bottom boundary layers in both types of model. For all of these processes, the energy that is dissipated is simply lost instead of being transferred to the relevant connecting dynamical regime, or to a different form of energy; this represents a previously overlooked flaw for which solutions are only now beginning to emerge, as described in this book.

That being said, in other contexts and for other parameterisations, this “missing” energy must be artificially recreated. A prominent example is the unaccounted-for (wave-driven) supply of energy needed to mix density in the ocean by prescribing some interior diffusivity. Another example is the heating of the upper atmosphere by molecular dissipation—also related to gravity wave breaking—but the same holds true for virtually any other parameterisations and dynamical regime. In other words, our current atmosphere and ocean models fail to completely account for the mechanical energy cycle. The goal of TRR 181 is to remedy this shortcoming by connecting all parameterisations in state-of-the-art atmosphere and ocean models in an energetically and mathematically consistent way. To this end, 16 sub-projects with 29 principal investigators from applied mathematics, meteorology, and physical oceanography have been established. The following chapters provide an overview of representative specific topics covered by the sub-projects.

The processes, parameterisations, and interactions addressed range from isolated idealised model setups to fully coupled global climate models. Chapter 1 starts out by reviewing a coherent hierarchy of models for the dynamical core in which many of the multi-scale interactions (which also repeatedly feature in the later chapters of this volume) are explored. Due care is taken to fully specifying the respective scaling and simplifying assumptions and to exposing the underlying Hamiltonian structure of the inviscid equations. The chapter subsequently discusses modifications to the classical scaling regimes in the equatorial region, and the impact of different forms of viscous dissipation or eddy damping on the solution and bifurcation structure of geophysical flows. Lastly, the chapter covers the systematic derivation of stochastic parameterisations for small-scale motions, a major theoretical development of the last decade which is yet to be fully understood and to be investigated operationally.

A major point of conceptual and practical uncertainty in the modelling of the atmosphere and the ocean is the dissipation of large-scale mean or eddy balanced flow, for which several processes have been put forward. They involve interior loss of balance by either ageostrophic or symmetric instability, Lighthill radiation of gravity waves, lee wave generation at topographic obstacles, interaction with western boundaries, the interaction between gravity waves and frontogenesis, or simply bottom friction. The relative importance of these processes for both the atmosphere and the ocean, however, has yet to be quantified, despite the fact that the dissipation of balanced flow is a key element in the energy cycle. Accordingly, Chapter 2 is devoted to this topic.

Momentum exchange between the different dynamical regimes, for instance the wave drag forcing of large-scale circulation in the upper atmosphere by upward propagating and breaking gravity waves, is also important to understanding the mechanism of circulation and the energy cycle. However, wave breaking, the

interaction of waves with the mean flow and stratification, and especially wave–wave interactions (and the resulting spectral energy transfers within the wave field to smaller wavelengths and thus towards wave breaking and density mixing) are only poorly understood, but essential to grasping the effects of gravity waves on the large-scale circulation. Chapters 3 and 4 focus on the topic of gravity waves, in the ocean and atmosphere, and in models and observations.

While the turbulent energy cascade in the classical isotropic turbulence regime at high Reynolds numbers appears to offer a valid description, the assumption of an energy transport between neighbouring wave numbers in an inertial sub-range over several decades without any dissipation effects appears invalid for turbulent flow on larger scales, in particular for geostrophic turbulence. As such, basing parameterisations on this assumption is problematic. Spectral energy transport from small to large scales is present, but usually ignored in standard sub-grid closures such as harmonic, hyper- or non-linear viscosity. In particular, the use of harmonic lateral friction in ocean models for purely numerical reasons is in striking contradiction to an inverse energy cascade and underscores the need to reconsider present sub-grid-scale closures in ocean and atmosphere models. New approaches to solving these problems are described in Chapter 5.

Beyond energy, there are also other properties of the system that require consistent treatment. Amongst them is momentum conservation, which puts constraints on, e.g. eddy parameterisation in ocean models, since eddies are known to redistribute, but not to create momentum. Eddy parameterisations for the ocean and the challenges in the diagnosis of eddy effects from models and observations are discussed in Chapter 6. A further issue of fundamental physical importance is addressed in Chapter 7, which focuses on the second law of thermodynamics and how it can be consistently related to the averaged equations of motion and in particular their sub-grid closures and on how parameterisations such as hyper-viscosity can be made consistent using what we know about the directions of energy cascades from large to small scales.

The physical inconsistencies in current state-of-the-art modelling are paralleled by numerical inconsistencies and challenges. While advection conserves all properties of a fluid particle, numerical advection schemes are by definition notoriously non-conservative, since they introduce spurious mixing and dissipation, a well-known but serious shortcoming of current atmosphere and ocean models. This issue is at the heart of Chapter 8. On the one hand, this spurious mixing and dissipation needs to be reduced as much as possible using advanced numerical techniques, such as mesh adaptivity or mesh-free Lagrangian methods. On the other, the remaining inescapably spurious energy sources also need to be quantified and should be taken into account when designing new parameterisations and sub-grid closures. In turn, air–sea interaction is another fundamental but numerically challenging aspect of the climate system. Surface ocean waves play a key role in transferring momentum, energy, heat, and other properties across the air–sea boundary; yet, their interaction with large- and small-scale flows in the ocean and atmosphere is not well understood. Chapter 9 details a promising new approach to directly simulating these processes numerically.

In TRR 181, the process-oriented topics presented here are complemented by an operationally oriented synthesis focusing on two climate models currently being developed in Germany. In this way, the goal of TRR 181 is to help reduce the biases in and increase the accuracy of atmosphere and ocean models and ultimately to improve climate models and climate predictions.

Hamburg, Germany
October 2018

Carsten Eden
Armin Iske

Acknowledgements

The chapters of this work refer to projects pursued at the Collaborative Research Centre “Energy Transfers in Atmosphere and Ocean” (TRR 181), funded by the German Research Foundation (DFG). Moreover, the encouraging support of Martin Peters (Springer Heidelberg), which was the initial spark for this book, is gratefully acknowledged.

Contents

1 Multi-scale Methods for Geophysical Flows	1
Christian L. E. Franzke, Marcel Oliver, Jens D. M. Rademacher and Gualtiero Badin	
2 The Interior Energy Pathway: Inertia-Gravity Wave Emission by Oceanic Flows	53
Jin-Song von Storch, Gualtiero Badin and Marcel Oliver	
3 The IDEMIX Model: Parameterization of Internal Gravity Waves for Circulation Models of Ocean and Atmosphere	87
Dirk Olbers, Carsten Eden, Erich Becker, Friederike Pollmann and Johann Jungclaus	
4 Observations and Models of Low-Mode Internal Waves in the Ocean	127
Christian Mertens, Janna Köhler, Maren Walter, Jin-Song von Storch and Monika Rhein	
5 Toward Consistent Subgrid Momentum Closures in Ocean Models	145
Sergey Danilov, Stephan Juricke, Anton Kutsenko and Marcel Oliver	
6 Diagnosing and Parameterizing the Effects of Oceanic Eddies	193
Alexa Griesel, Julia Dräger-Dietel and Kerstin Jochumsen	
7 Entropy Production in Turbulence Parameterizations	225
Almut Gassmann and Richard Blender	
8 Reducing Spurious Diapycnal Mixing in Ocean Models	245
Knut Klingbeil, Hans Burchard, Sergey Danilov, Claus Goetz and Armin Iske	

9 Diffuse Interface Approaches in Atmosphere and Ocean—Modeling and Numerical Implementation 287
Harald Garcke, Michael Hinze and Christian Kahle

Index 309

Contributors

Gualtiero Badin

Center for Earth System Research and Sustainability (CEN)
Universität Hamburg, Hamburg, Germany
e-mail: gualtiero.badin@uni-hamburg.de

Erich Becker

Leibniz-Institut für Atmosphärenphysik e.V.
Universität Rostock, Rostock, Germany
e-mail: becker@iap-kborn.de

Richard Blender

Center for Earth System Research and Sustainability (CEN)
Universität Hamburg, Hamburg, Germany
e-mail: richard.blender@uni-hamburg.de

Hans Burchard

Leibniz Institute for Baltic Sea Research Warnemünde (IOW)
Rostock-Warnemünde, Germany
e-mail: hans.burchard@io-warnemuende.de

Sergey Danilov

Alfred Wegener Institute for Polar and Marine Research (AWI)
Bremerhaven, Germany
and Jacobs University, Bremen, Germany
e-mail: sergey.danilov@awi.de

Julia Dräger-Dietel

Center for Earth System Research and Sustainability (CEN)
Universität Hamburg, Hamburg, Germany
e-mail: julia.draeger-dietel@uni-hamburg.de

Carsten Eden

Center for Earth System Research and Sustainability (CEN)
Universität Hamburg, Hamburg, Germany
e-mail: carsten.eden@uni-hamburg.de

Christian L. E. Franzke

Center for Earth System Research and Sustainability (CEN)
Universität Hamburg, Hamburg, Germany
e-mail: christian.franzke@uni-hamburg.de

Harald Garcke

Fakultät für Mathematik, Universität Regensburg
Regensburg, Germany
e-mail: harald.garcke@mathematik.uni-regensburg.de

Almut Gassmann

Leibniz-Institut für Atmosphärenphysik e.V.
Universität Rostock, Rostock, Germany
e-mail: gassmann@iap-kborn.de

Claus Goetz

Fachbereich Mathematik, Universität Hamburg
Hamburg, Germany
e-mail: claus.goetz@uni-hamburg.de

Alexa Griesel

Center for Earth System Research and Sustainability (CEN)
Universität Hamburg, Hamburg, Germany
e-mail: alexa.griesel@uni-hamburg.de

Michael Hinze

Fachbereich Mathematik, Universität Hamburg
Hamburg, Germany
e-mail: michael.hinze@uni-hamburg.de

Armin Iske

Fachbereich Mathematik, Universität Hamburg
Hamburg, Germany
e-mail: armin.iske@uni-hamburg.de

Kerstin Jochumsen

Federal Maritime and Hydrographic Agency
Hamburg, Germany
e-mail: kerstin.jochumsen@bsh.de

Johann Jungclaus

Max-Planck-Institut für Meteorologie (MPI-M)
Hamburg, Germany
e-mail: johann.jungclaus@mpimet.mpg.de

Stephan Juricke

Jacobs University, Bremen, Germany
e-mail: s.juricke@jacobs-university.de

Christian Kahle

Zentrum Mathematik, Technische Universität München
Munich, Germany
e-mail: christian.kahle@ma.tum.de

Knut Klingbeil

Leibniz Institute for Baltic Sea Research Warnemünde (IOW)
Rostock-Warnemünde, Germany
e-mail: knut.klingbeil@io-warnemuende.de

Janna Köhler

MARUM/Institut für Umweltphysik, Universität Bremen
Bremen, Germany
e-mail: jannak@uni-bremen.de

Anton Kutsenko

Jacobs University, Bremen, Germany
e-mail: a.kutsenko@jacobs-university.de

Christian Mertens

MARUM/Institut für Umweltphysik, Universität Bremen
Bremen, Germany
e-mail: cmertens@uni-bremen.de

Dirk Olbers

Alfred Wegener Institute for Polar and Marine Research (AWI)
Bremerhaven, Germany
and MARUM/Institut für Umweltphysik, Universität Bremen
Bremen, Germany
e-mail: dirk.olbers@awi.de

Marcel Oliver

Jacobs University, Bremen, Germany
e-mail: m.oliver@jacobs-university.de

Friederike Pollmann

Center for Earth System Research and Sustainability (CEN)
Universität Hamburg, Hamburg, Germany
e-mail: friederike.pollmann@uni-hamburg.de

Jens D. M. Rademacher

Fachbereich Mathematik, Universität Bremen
Bremen, Germany
e-mail: jdmr@uni-bremen.de

Monika Rhein

MARUM/Institut für Umweltphysik, Universität Bremen
Bremen, Germany
e-mail: mrhein@uni-bremen.de

Jin-Song von Storch

Max-Planck-Institut für Meteorologie (MPI-M)
Hamburg, Germany
e-mail: jin-song.von.storch@mpimet.mpg.de

Maren Walter

MARUM/Institut für Umweltphysik, Universität Bremen
Bremen, Germany
e-mail: maren.walter@uni-bremen.de

Chapter 1

Multi-scale Methods for Geophysical Flows



Christian L. E. Franzke, Marcel Oliver, Jens D. M. Rademacher and Gualtiero Badin

Abstract Geophysical flows comprise a broad range of spatial and temporal scales, from planetary- to meso-scale and microscopic turbulence regimes. The relation of scales and flow phenomena is essential in order to validate and improve current numerical weather and climate prediction models. While regime separation is often possible on a formal level via multi-scale analysis, the systematic exploration, structure preservation, and mathematical details remain challenging. This chapter provides an entry to the literature and reviews fundamental notions as background for the later chapters in this collection and as a departure point for original research in the field.

1.1 Introduction

The climate system varies on a multitude of temporal and spatial scales which interact nonlinearly with each other. It is common that phenomena with small spatial scales also vary fast while phenomena with large spatial scales vary more slowly. For instance, small-scale turbulent eddies have spatial scales from millimeters up to a meter and exist between a few seconds and a few minutes. In the atmosphere, meso-scale phenomena like convection and gravity waves have length-scales between a few 100 meters and about 20 km and time-scales between hours and days. Synoptic

C. L. E. Franzke · G. Badin
Center for Earth System Research and Sustainability (CEN), Universität Hamburg,
Hamburg, Germany
e-mail: christian.franzke@uni-hamburg.de

G. Badin
e-mail: gualtiero.badin@uni-hamburg.de

M. Oliver (✉)
Jacobs University, Bremen, Germany
e-mail: m.oliver@jacobs-university.de

J. D. M. Rademacher
Fachbereich Mathematik, Universität Bremen, Bremen, Germany
e-mail: jdmr@uni-bremen.de

weather systems have lifetimes of a few days and spatial scales of up to 2000 km while planetary-scale teleconnection patterns can extend throughout the hemisphere and have time-scales from a week to decades.

At the most fundamental level, scale separation in geophysical flows is described by the concept of *balance*. When the Rossby number (the ratio between inertial forces and Coriolis forces) or the Froude number (the ratio of inertial flow velocity to gravity wave speed) is small, there exists a slow or *balanced* component which evolves nonlinearly and interacts only weakly with the high-frequency components. The fast motions can often, though not always, be approximately characterized as the high-frequency linear waves in the linearized equations of motion. We remark that the term “waves” is sometimes used in a loose sense for describing spatiotemporal oscillations diagnosed by Fourier wave modes, in particular in experimental studies.

A precise characterization of balance is a perennial theme in geophysical fluid dynamics; a recent review can be found in McIntyre (2015). Balance can be described in two ways: kinematically via *balance relations* which define almost-invariant objects through a phase space constraint, and dynamically as *balance models* which are closed sets of equations representing the slow dynamics on a balanced manifold, thus approximating the flow of the full system under balanced initial conditions. However, the notion of balance is intrinsically approximate: Emergence of imbalance from balanced initial conditions is generic, though often exponentially small; cf. Vanneste (2013). Temam and Wirosoetisno (2010) argue that even in the viscous case, there is no exact invariant balance manifold. A recent discussion can be found in Whitehead and Wingate (2014).

Balance in the equatorial band of latitudes is a more subtle concept, as the Coriolis parameter changes sign at the equator causing a singularity in straightforward small Rossby number expansions. There are suitable equatorial scalings and balance assumptions to formally circumvent the problem, usually at the price of losing some (linear) waves (McIntyre 2015; Verkley and van der Velde 2010; Theiss and Mohebalhojeh 2009). Recent work by Chan and Shepherd (2013) shows that it is possible to capture equatorial Rossby and Kelvin waves in a full hierarchy of equatorial balance models.

While the different scale regimes can have different physical mechanisms driving them, they are all described by the same general set of equations of motion. *Multi-scale asymptotics* can be used as a systematic means of deriving new sets of equations which only describe the flows at certain temporal and spatial scales, and their interactions with other regimes; see, for example, Majda and Klein (2003) or Klein (2010). These new sets of equations then describe the underlying dynamics of the respective flow regimes and, thus, provide a better understanding of the flow dynamics and the dominant balance conditions.

Recent studies elucidated the interactions between the planetary- and synoptic scale regimes (Dolaptchiev and Klein 2013) and between the planetary- and meso-scale regimes (Shaw and Shepherd 2009). The latter study also discusses how these multi-scale models can be used to systematically derive energy-consistent subgrid-scale parameterizations.

An unavoidable ingredient in comprehensive models of climate and global atmosphere and ocean dynamics is the parameterizations of unresolved degrees of freedom. The numerical aspects in the context of geostrophic turbulence are discussed in Danilov et al. (2019). Of particular interest to us is the use of stochastic modeling techniques for developing such parameterizations, in particular the connection of stochastic subgrid parameterizations with traditional deterministic multi-scale asymptotics. Theoretical work on the elimination of fast scales is often based on averaging techniques, which naturally apply to stochastic modeling as well. Mathematical introductions to this direction of model reduction are given by Givon et al. (2004) and Pavliotis and Stuart (2008) while Franzke et al. (2015) and Gottwald et al. (2017) provide introductions to stochastic climate modeling. The parameterization problem is also an issue of data collection and assimilation, which we do not discuss here.

Part of the motivation for pursuing stochastic modeling approaches derives from the observation that most numerical weather predictions have too little ensemble spread. This problem is amplified by the fact that the observed weather lies too often outside of the forecast ensemble. To increase the ensemble spread, stochastic parameterizations have been introduced. The European Centre for Medium-Range Weather Forecasts (ECMWF) is leading the introduction of stochastic parameterizations into numerical weather and seasonal climate prediction models. They currently use two different schemes operationally, the Stochastically Perturbed Parameterization Tendencies Scheme (SPPT) and the Stochastic Kinetic Energy Backscatter Scheme (SKEBS), described in Palmer et al. (2009).

SPPT is based on the notion that, especially with increasing numerical resolution, the equilibrium assumption no longer holds and the subgrid-scale state should be sampled rather than represented by the equilibrium mean. Consequently, SPPT multiplies the accumulated physical tendencies at each grid point and time step with a random pattern that has spatial and temporal correlations. SKEBS aims at representing model uncertainty arising from unresolved and unrepresented subgrid-scale processes by introducing random perturbations to the stream function. SKEBS is based on the rationale that numerical dissipation needs to be compensated by an injection of energy into the resolved scales. Current deterministic climate models, however, neglect this energy pathway. The use of stochastic parameterizations in climate models is less well established due to the required re-tuning of all deterministic parameterizations. The current use of stochastic parameterizations is rather *ad hoc* (Franzke et al. 2015) and does not consider energy and momentum consistency. For instance, SPPT has no rigorous mathematical justification; it is based on trial-and-error and empirical evidence that it increases the spread of ensemble forecasts. However, this does not necessarily make the forecasts more precise. SKEBS has some mathematical justification. Theory suggests that most of the “subgrid-scale” energy should be re-injected close to the truncation scale, i.e., at scales where numerical dissipation acts. However, in practice, the re-injected energy gets damped very quickly with no improvement in forecast skill. The usual remedy is to re-inject the energy into all scales at a rate empirically determined by trying to match a known energy spectrum. While there is growing evidence showing the benefits of stochastic

parameterizations, there is also a need for more rigorous mathematical approaches, particularly in light of the observation that errors introduced by inconsistent treatment of the interactions between resolved and unresolved scales can be very significant (Shaw and Shepherd 2009; Burkhardt and Becker 2006; Becker 2003).

This chapter, first, intends to give an overview on current progress and open issues in multi-scale modeling of geophysical flows. We aim at some broadness, but our focus is clearly biased by our own research interests in geometric and structure-preserving methods, dynamical systems techniques, and stochastic modeling.

Second, and no less important, a large part of the chapter is devoted to laying down foundational concepts, starting with the equations of motion, non-dimensionalization, the introduction of the classical limits in geophysical fluid dynamics, variational and Hamiltonian methods, dissipation, and stochastic model reduction. While much of this is standard textbook material, covered in great detail, for example, in the classical books by Gill (1982), Kamenkovich et al. (1986), and Pedlosky (1987), or in the more recent books by Salmon (1998), Vallis (2006), and Olbers et al. (2012), our intent is to introduce the foundations most relevant to the questions raised here in a concise and unified notation. In our presentation, we take the rotating Boussinesq equations as the single parent model from which all other models arise, aim at a clear statement of simplifying assumptions, and give some consideration to the full Coriolis term and to equatorial scalings.

The remainder of this chapter is structured as follows. Section 1.2 introduces the rotating Boussinesq equations as the governing equations of geophysical flow, discusses imbalance variables, non-dimensionalization, and several simplified models derived in different scaling limits: the hydrostatic approximation in form of the primitive equations, the quasi-geostrophic equations in several forms, and the shallow water equations. Section 1.3 discusses the variational principle which gives rise to the equations of motion as well as Poisson and Nambu formulations of the dynamics. Section 1.4 provides a high-level overview of the role of dissipation and its relation to turbulence and to dynamical systems methods for studying linear and nonlinear waves. Section 1.5 introduces systematic approaches for the stochastic modeling of fast motions. The chapter closes with a brief outline of questions we are addressing in our current research.

1.2 The Governing Equations

1.2.1 *Rotating Boussinesq Equations*

As the starting point for this exposition, we consider the rotating Boussinesq equations in locally Cartesian coordinates. These equations simplify the full equations for global atmospheric or oceanic dynamics in two fundamental ways.

First, the Boussinesq approximation is based on the observation that density fluctuations are so small that their impact on inertia is negligible, while their impact on

buoyancy remains significant. This assumption is typically very well satisfied in the ocean, so that most ocean general circulation models assume the Boussinesq approximation. For atmospheric flows, the Boussinesq approximation is more tenuous, but under the so-called anelastic approximation, the assumption of constant background pressure, and the use of vertical pressure coordinates, the equations governing large-scale atmospheric flows can also be written in the same form. One main consequence of the Boussinesq approximation is the absence of acoustic waves.

Second, the local metric on the earth is nearly Cartesian and, for the purpose of studying scale interactions, will be replaced by an exactly Cartesian metric. We still allow variations in the Coriolis parameter with latitude so that equatorial scalings can be explored. Special cases are the f -plane or tangent plane approximation where the Coriolis vector is assumed constant and the β -plane approximation where a linear approximation to the variation of the Coriolis parameter with latitude y is used.

We shall also restrict ourselves to the simplest possible equation of state where changes in density depend linearly on changes in temperature, neglect source terms, impose incompressibility, and assume that density (or, equivalently, potential temperature) is advected and subject to harmonic diffusion. With these provisions, our governing equations read

$$D_t \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\frac{1}{\rho_0} \nabla p - \frac{g\rho}{\rho_0} \mathbf{k} + \nu \Delta \mathbf{u}, \quad (1.1a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (1.1b)$$

$$D_t \rho = \kappa \Delta \rho, \quad (1.1c)$$

where $D_t = \partial_t + \mathbf{u} \cdot \nabla$ denotes the material derivative, \mathbf{u} is the three-dimensional fluid velocity field,

$$\boldsymbol{\Omega} = |\boldsymbol{\Omega}| \begin{pmatrix} 0 \\ \cos \vartheta \\ \sin \vartheta \end{pmatrix} \quad (1.2)$$

the angular velocity vector describing the rotation of the earth at latitude ϑ , ρ_0 a constant reference density, p the departure from hydrostatic pressure, g the constant of gravity, ρ the departure from the constant reference density, \mathbf{k} the unit vector in z (vertical)-direction, ν the coefficient of viscosity, and κ is proportional to the coefficient of thermal diffusion.

A derivation and a discussion of the underlying assumptions can be found in any classical textbook on geophysical fluid dynamics, e.g., Gill (1982), Pedlosky (1987), or Vallis (2006).

The dissipative terms on the right-hand side describe, at this point, molecular viscosity and diffusion processes. In typical large-scale circulation problems, they act at scales far beyond what can be resolved in any practical numerical simulation. Thus, they will be replaced, explicitly or implicitly via a stable numerical scheme by eddy diffusion and/or dissipative properties of the numerical scheme. Such terms may not

be harmonic or isotropic, an issue to which we will return in Section 1.4.1 below. In the remainder of this section, we shall only consider the inviscid Boussinesq equations independent of the issues relating to viscosity and viscous parameterizations.

We shall consider (1.1) on a domain Ω with a free upper boundary at $z = h(\mathbf{x}_h, t)$ and rigid bottom at $z = -b(\mathbf{x}_h)$; here and below, we use the subscript “h” to denote the two horizontal components of a vector in three dimensions. We shall not discuss lateral boundary conditions here. For highly idealized studies, for example, on the phenomenology of geostrophic turbulence, periodic lateral boundary conditions are frequently used. At the bottom boundary, we impose the impermeability condition $\mathbf{n} \cdot \mathbf{u} = 0$, where \mathbf{n} denotes the outward unit normal vector, which may be written

$$\mathbf{u}_h \cdot \nabla_h b + w = 0, \quad (1.3a)$$

plus viscous boundary conditions or bottom drag parameterizations if applicable. The free surface is described by the condition $\sigma(\mathbf{x}, t) \equiv h(\mathbf{x}_h, t) - z = 0$ and is subject to the kinematic boundary condition $D_t \sigma = 0$, i.e.,

$$\partial_t h + \mathbf{u}_h \cdot \nabla_h h = w. \quad (1.3b)$$

As a second condition at the free surface, we have the dynamic boundary condition

$$p = p_s, \quad (1.3c)$$

i.e., pressure equals a specified external pressure p_s at the surface. In the viscous case, the free surface boundary conditions are augmented by (wind) stress conditions. Integrating the incompressibility constraint (1.1b) in z and using the two kinematic boundary conditions (1.3a) and (1.3b), we obtain the *free surface equation*

$$\partial_t h + \nabla_h \cdot \int_{-b}^h \mathbf{u}_h \, dz = 0. \quad (1.4)$$

Propagation in time then uses the horizontal momentum equations, the free surface equation, and the advection of buoyancy (1.1c) as prognostic quantities while the vertical velocity w is reconstructed from the incompressibility constraint; see, e.g., Klingbeil and Burchard (2013).

The free surface equation can be approximated in various ways. Replacing the upper limit of integration in (1.4) by $z = 0$ while still keeping the time derivative $\partial_t h$ gives a linearization of the free surface equation which can be formulated in the spirit of Chorin’s projection method where the free surface update is obtained by solving an elliptic equation. This removes time-step restrictions due to fast surface waves (surface waves are damped when the time step becomes too large) at the expense of losing volume and tracer conservation—see, e.g., Griffies et al. (2001) for a discussion.

A more drastic approximation is a *rigid lid* upper boundary, where all equations are posed on a fixed domain bounded above at $z = 0$. Consequently, the dynamic boundary condition (1.3c) must be dropped. The time evolution of all fields can be computed by an incompressible solver. An approximation of the elevation of the free surface may then be diagnostically obtained by solving the equivalent hydrostatic pressure relation

$$p_s = g\rho_0 h \quad (1.5)$$

for h . The rigid lid approximation removes all surface gravity waves, which is often a reasonable approximation in the ocean.

We finally remark that the inviscid Boussinesq equations conserve the *total energy* or *Hamiltonian*

$$H = \int_{\Omega} \frac{1}{2} |\mathbf{u}|^2 + \frac{g}{\rho_0} \rho z \, dx \quad (1.6)$$

and materially conserve *potential vorticity*

$$q = (2\boldsymbol{\Omega} + \boldsymbol{\omega}) \cdot \nabla \rho, \quad (1.7)$$

where $\boldsymbol{\omega} = \nabla \times \mathbf{u}$ denotes the relative vorticity, i.e., q satisfies the advection equation

$$D_t q = 0. \quad (1.8)$$

Both conservation laws can be verified by direct computation. However, in Section 1.3.1 we will show that they emerge elegantly from symmetries in the underlying variational principle.

1.2.2 Imbalance Variables

Large-scale geophysical flow, at least in the sub-equatorial regime, is to a substantial part determined by potential vorticity alone (McIntyre and Norton 2000). It is therefore natural to use potential vorticity as one of the prognostic variables and to augment the set of prognostic variables by so-called *imbalance variables* suitably chosen such that they remain small so long as the flow is nearly balanced. In the context of the f -plane shallow water equations, these additional variables are divergence and ageostrophic vorticity; see the discussion in Section 1.2.6 for details. This equivalent reformulation of the equation of motion is central to the discussion of spontaneous emission of gravity waves by Ford et al. (2000); also see the discussion in von Storch et al. (2019). Mohebalhojeh and Dritschel (2001) report numerical advantages when simulating shallow water using this set of variables, and Dritschel et al. (2017) find that the differences between several variational balance models can

only be understood when looking at balance relations formulated in terms of such imbalance variables. Imbalance variables for the primitive equations are introduced, for example, by McIntyre and Norton (2000).

For the f -plane Boussinesq system in the so-called traditional approximation, $2\boldsymbol{\Omega} = (0, 0, f)$, with f a constant referred to as the Coriolis parameter, there are different possible choices. Dritschel and Viúdez (2003) use the horizontal components of

$$\mathbf{A} = \nabla \times \mathbf{u} - \frac{g}{\rho_0 f} \nabla \rho, \quad (1.9)$$

which leads to a nonlinear inversion problem for recovering the vector potential of \mathbf{A} . Alternatively, it is also possible to use ageostrophic vorticity and divergence (Vanneste 2013). However, as noted in Section 1.2.4 below, the traditional and the hydrostatic approximations are not independent. Thus, we shall derive (for purposes of exposition working at the linear level only) a version of the non-traditional f -plane Boussinesq system in imbalance variables. In this setting, $\boldsymbol{\Omega}$ is assumed constant but need not be parallel to the vertical direction \mathbf{k} .

To begin, we note that linear inertia-gravity waves require rotation and strong stratification, which can be expressed by assuming that density variations are small relative to a static vertical stratification profile. We write

$$\rho(\mathbf{x}, t) = \bar{\rho}(z) + \rho'(\mathbf{x}, t). \quad (1.10)$$

The *Brunt–Väisälä* or *buoyancy frequency* is given by

$$N^2 = -\frac{g}{\rho_0} \frac{\partial \bar{\rho}}{\partial z}, \quad (1.11)$$

which we assume to be independent of z in this section. Then the “thermodynamic equation” (1.1c) reads

$$D_t \rho - \frac{N^2 \rho_0}{g} w = 0 \quad (1.12)$$

and therefore, by incompressibility,

$$\partial_z D_t \rho + \frac{N^2 \rho_0}{g} \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} = 0. \quad (1.13)$$

Further, taking the full divergence of the inviscid Boussinesq momentum equation (1.1a), invoking incompressibility, differentiating in time, and using (1.13), we find

$$\nabla \cdot (2\boldsymbol{\Omega} \times \partial_t \mathbf{u}) + \frac{1}{\rho_0} \partial_t \Delta p - N^2 \nabla_{\mathbf{h}} \cdot \mathbf{u}_{\mathbf{h}} = \text{NL}, \quad (1.14)$$

where NL is used to denote any number of nonlinear terms, possibly different from one line to the next. Next, taking the horizontal divergence of (1.1a) yields

$$\partial_t \nabla_h \cdot \mathbf{u}_h + \nabla_h \cdot (2\boldsymbol{\Omega} \times \mathbf{u})_h + \frac{1}{\rho_0} \Delta_h p = \text{NL}. \quad (1.15)$$

This motivates taking the *horizontal divergence*

$$\delta = \nabla_h \cdot \mathbf{u}_h \quad (1.16)$$

and the *acceleration divergence* or *ageostrophic vorticity*

$$\gamma = -\nabla_h \cdot (2\boldsymbol{\Omega} \times \mathbf{u})_h - \frac{1}{\rho_0} \Delta_h p \quad (1.17)$$

as imbalance variables, so that $\partial_t \delta = \gamma + \text{NL}$. Then, using (1.14) to eliminate $\partial_t \Delta p$, we compute

$$\begin{aligned} \partial_t \Delta \gamma &= -\Delta \nabla_h \cdot (2\boldsymbol{\Omega} \times \partial_t \mathbf{u})_h + \Delta_h \nabla \cdot (2\boldsymbol{\Omega} \times \partial_t \mathbf{u}) - N^2 \Delta_h \delta + \text{NL} \\ &= -\partial_{zz} \nabla_h \cdot (2\boldsymbol{\Omega} \times \partial_t \mathbf{u})_h + \Delta_h (2\boldsymbol{\Omega}_h^\perp \cdot \partial_{tz} \mathbf{u}_h) - N^2 \Delta_h \delta + \text{NL}. \end{aligned} \quad (1.18)$$

Writing $(2\boldsymbol{\Omega} \times \mathbf{u})_h = f \mathbf{u}_h^\perp - w 2\boldsymbol{\Omega}_h^\perp$ with $\mathbf{u}_h^\perp = (-v, u)$ in the first term on the right, substituting the horizontal momentum equation for all instances of $\partial_t \mathbf{u}_h$, invoking incompressibility to replace all instances of $\partial_z w$ by $-\delta$, eliminating $\Delta_h p$ via the definition of $\gamma = \partial_t \delta + \text{NL}$, and collecting terms, we find

$$\begin{aligned} \partial_t \Delta \gamma &= -f^2 \partial_{zz} \delta - f \partial_z (2\boldsymbol{\Omega}_h \cdot \nabla_h \delta) + (2\boldsymbol{\Omega}_h^\perp \cdot \nabla_h)^2 \delta \\ &\quad + f \partial_z (2\boldsymbol{\Omega}_h \cdot \nabla_h^\perp \nabla_h^\perp \cdot \mathbf{u}_h - 2\boldsymbol{\Omega}_h \cdot \Delta_h \mathbf{u}_h) - (N^2 + |2\boldsymbol{\Omega}_h|^2) \Delta_h \delta + \text{NL}. \end{aligned} \quad (1.19)$$

Finally, applying the vector identity $\nabla_h \nabla_h + \nabla_h^\perp \nabla_h^\perp = I \Delta_h$ twice and collecting terms, we can write the system in the form of a nonlinearly perturbed wave equation, where

$$\partial_t \delta - \gamma = \text{NL}, \quad (1.20a)$$

$$\partial_t \Delta \gamma + (2\boldsymbol{\Omega} \cdot \nabla)^2 \delta + N^2 \Delta_h \delta = \text{NL}. \quad (1.20b)$$

Without the nonlinear terms, this system is equivalent to the equation considered by Gerkema and Shrira (2005a) to study linear waves on the non-traditional f -plane; for related work on the non-traditional β -plane, see Gerkema and Shrira (2005b), Kasahara and Gary (2010), and Stewart and Dellar (2012). The shallow water equations with full Coriolis parameter on the sphere are considered by Tort et al. (2014), and linear stability of nonlinear jet-type solutions is investigated by Tort et al. (2016).

We finally note that when the traditional approximation is made, (1.20) directly reduces to the system studied, e.g., by Vanneste (2013).

When the nonlinear terms in (1.20) are retained, these equations together with the advection equation for potential vorticity (1.7) form a closed system for the evolution of the f -plane Boussinesq equations because (\mathbf{u}, p, ρ) can be recovered from (q, δ, γ) by inverting nonlinear elliptic equations. These involve boundary conditions, and for vertically bounded domains with $w = 0$ at $z = 0, -H$ require solving $D_t \rho = 0$ at the top and bottom boundaries. The two-dimensional fields $b(\mathbf{x}_h, 0, t)$ and $b(\mathbf{x}_h, -H, t)$ are therefore additional degrees of freedom. Further complications arise for horizontally bounded domains.

1.2.3 Mid-latitude Scalings

Let us now consider possible scaling regimes for (1.1) on a tangent plane at mid-latitude ϑ_0 . In this regime, the flow is expected to be horizontally isotropic, but vertical scales will generally differ. We thus split (1.1a) into the horizontal and vertical component equations and non-dimensionalize by introducing typical horizontal velocity U , typical vertical velocity W , typical Coriolis parameter $f_0 = 2|\boldsymbol{\Omega}| \sin \vartheta_0$, typical horizontal length-scale L , typical vertical length-scale H (not to be confused with the Hamiltonian H), typical time-scale T , typical pressure scale P , and typical density perturbation scale Γ , inserting $\mathbf{u} = U\hat{\mathbf{u}}, w = W\hat{w}, 2\boldsymbol{\Omega} = f_0\hat{\boldsymbol{\Omega}}$, etc., into the Boussinesq equations (1.1), and finally dropping the hats, we obtain

$$\frac{U}{T} \partial_t \mathbf{u}_h + \frac{U^2}{L} \mathbf{u}_h \cdot \nabla_h \mathbf{u}_h + \frac{UW}{H} w \partial_z \mathbf{u}_h + f_0 U \boldsymbol{\Omega}_z \cdot \mathbf{u}_h^\perp - f_0 W \boldsymbol{\Omega}_h^\perp \cdot w = -\frac{P}{L\rho_0} \nabla_h p, \quad (1.21a)$$

$$\frac{W}{T} \partial_t w + \frac{UW}{L} \mathbf{u}_h \cdot \nabla_h w + \frac{W^2}{H} w \partial_z w + f_0 U \boldsymbol{\Omega}_h^\perp \cdot \mathbf{u}_h = -\frac{P}{H\rho_0} \partial_z p - \frac{g\Gamma}{\rho_0} \rho, \quad (1.21b)$$

$$\frac{U}{L} \nabla_h \cdot \mathbf{u}_h + \frac{W}{H} \partial_z w = 0, \quad (1.21c)$$

$$\frac{\Gamma}{T} \partial_t \rho + \frac{U\Gamma}{L} \mathbf{u}_h \cdot \nabla_h \rho + \frac{WN_0^2 \rho_0}{g} w \partial_z \rho = 0. \quad (1.21d)$$

Here and below, N_0 denotes the *typical* Brunt–Väisälä frequency

$$N_0^2 = -\frac{g}{\rho_0} \left[\frac{\partial \rho}{\partial z} \right], \quad (1.22)$$

where we write $[\partial \rho / \partial z]$ to denote the typical vertical density gradient. Note that it is not necessary to split off a static vertical profile as we did in Section 1.2.2.

We look at the problem on a horizontally advective time-scale, i.e., on a time-scale in which fluid parcels travel a horizontal distance of order one. This fixes

$$\frac{1}{T} \sim \frac{U}{L}. \quad (1.23)$$

Our goal is now to estimate the vertical velocity scale W . Introducing the *aspect ratio*

$$\alpha = \frac{H}{L}, \quad (1.24)$$

we may obtain a first simple estimate, $W \lesssim \alpha U$, directly from the incompressibility condition. However, the vertical velocity gradient does not necessarily participate in the dominant balance in the incompressibility relation. Indeed, we shall see that rotation as well as stratification may provide sharper scaling bounds on W .

Turning to the thermodynamic equation (1.21d), we obtain the scaling bound

$$W \lesssim \frac{g\Gamma}{\rho_0} \frac{U}{N_0^2 L} = \frac{g\Gamma}{\rho_0} \frac{L}{U} \alpha^2 \text{Fr}^2 \quad (1.25)$$

with *Froude number*

$$\text{Fr} = \frac{U}{N_0 H}. \quad (1.26)$$

Without loss of generality, we may take the scaling bound (1.25) to be sharp; if it is not, we can make it sharp by changing the definition of the Froude number.

We now make the assumption that the dominant balance in the horizontal momentum equation (1.21a) is between horizontal Coriolis force and horizontal pressure gradient and that the dominant balance in the vertical momentum equation (1.21b) is between buoyancy and vertical pressure gradient. This implies

$$f_0 L U \rho_0 \sim P \sim g \Gamma H. \quad (1.27)$$

Introducing the *Rossby number*

$$\text{Ro} = \frac{U}{f_0 L} \quad (1.28)$$

and writing $f \equiv \Omega_z$ for the non-dimensionalized rotation rate, we can rewrite the momentum equations as

$$\text{Ro} (\partial_t \mathbf{u}_h + \mathbf{u}_h \cdot \nabla_h \mathbf{u}_h) + \text{Fr}^2 w \partial_z \mathbf{u}_h + f \mathbf{u}_h^\perp - \alpha \frac{\text{Fr}^2}{\text{Ro}} \boldsymbol{\Omega}_h^\perp w = -\nabla_h p, \quad (1.29a)$$

$$\alpha^2 \text{Fr}^2 (\partial_t w + \mathbf{u}_h \cdot \nabla_h w) + \alpha^2 \frac{\text{Fr}^4}{\text{Ro}} w \partial_z w + \alpha \boldsymbol{\Omega}_h^\perp \cdot \mathbf{u}_h = -\partial_z p - \rho. \quad (1.29b)$$

We finally assume that we are on a 3D advective time-scale, i.e., that

$$\text{Ro} = \text{Fr}^2. \quad (1.30)$$

Then

$$\text{Ro} (\partial_t \mathbf{u}_h + \mathbf{u} \cdot \nabla \mathbf{u}_h) + f \mathbf{u}_h^\perp - \alpha \boldsymbol{\Omega}_h^\perp w = -\nabla_h p, \quad (1.31a)$$

$$\alpha^2 \text{Ro} (\partial_t w + \mathbf{u} \cdot \nabla w) + \alpha \boldsymbol{\Omega}_h^\perp \cdot \mathbf{u}_h = -\partial_z p - \rho. \quad (1.31b)$$

1.2.4 Hydrostatic Approximation

So far, we have not assumed smallness of any parameter beyond fixing the dominant balance in the momentum equations. The first substantial simplification of the equations of motion comes via the hydrostatic balance assumption

$$\alpha \ll 1 \quad (1.32)$$

while the Rossby number is at most $O(1)$. Thus, in 3D advective scaling, hydrostatic balance is purely a small aspect ratio assumption. If it is made, the contributions from the horizontal Coriolis force are small as well. We altogether obtain the hydrostatic *primitive equations*

$$\text{Ro} (\partial_t \mathbf{u}_h + \mathbf{u} \cdot \nabla \mathbf{u}_h) + f \mathbf{u}_h^\perp = -\nabla_h p, \quad (1.33a)$$

$$\partial_z p = -\rho, \quad (1.33b)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (1.33c)$$

$$D_t \rho = 0. \quad (1.33d)$$

In this context, we remark that it has long been known that the hydrostatic approximation requires the “traditional approximation” where the contributions from the horizontal Coriolis force are neglected to ensure that the Coriolis force remains energy-neutral; for example, see the discussion in White (2002) and Klein (2010).

While there have been early studies of the global circulation using balance models as discussed further below, contemporary OGCMs are based on the primitive equations. For the study of small-scale phenomena, the use of non-hydrostatic models is advancing (Fringer 2009). In the atmosphere, where non-hydrostatic effects are more pronounced, non-hydrostatic models are routinely used (Saito et al. 2007).

1.2.5 The Quasi-geostrophic Approximation on the β -plane

We now turn to the quasi-geostrophic approximation on a mid-latitude β -plane. It is a more severe approximation than hydrostaticity. We modify the assumptions made

in Section 1.2.3 in three respects. First, we look at the scaling

$$\text{Fr} \sim \text{Ro}, \quad (1.34)$$

thereby breaking the 3D advective scaling. Advection in the vertical is thus coming in at higher order than advection in the horizontal direction. Second, we assume that pressure variations are small relative to a static vertical stratification profile. As in Section 1.2.2, we split the density into a static stratification profile and a perturbation field, writing

$$\rho(\mathbf{x}, t) = \bar{\rho}(z) + \rho'(\mathbf{x}, t). \quad (1.35)$$

Correspondingly, we introduce the perturbation pressure ψ satisfying

$$\partial_z \psi = \partial_z p + \bar{\rho}. \quad (1.36)$$

Third, we make the β -plane approximation with the assumption that the change of the Coriolis parameter in y is $O(\text{Ro})$. In non-dimensionalized variables, this assumption reads

$$f = 1 + \text{Ro} \beta y, \quad (1.37)$$

where $\text{Ro} \beta$ is the meridional gradient of the Coriolis parameter. Inserting (1.35) into the inviscid thermodynamic equation (1.1c) and dropping the prime, we non-dimensionalize as follows:

$$\frac{U\Gamma}{L} (\partial_t \rho + \mathbf{u}_h \cdot \nabla_h \rho) + \frac{W\Gamma}{H} w \partial_z \rho + \frac{WN_0^2 \rho_0}{g} w \partial_z \bar{\rho} = 0. \quad (1.38)$$

Keeping the last term in the dominant balance leads, once again, to a scaling relation of the form (1.25) with identity in the first relation, so that, in non-dimensional variables,

$$\partial_t \rho + \mathbf{u}_h \cdot \nabla_h \rho + \text{Ro} w \partial_z \rho + w \partial_z \bar{\rho} = 0. \quad (1.39)$$

Similarly, the momentum equations in non-dimensional variables, under the same dominant balance assumptions as in Section 1.2.3 above, read

$$\text{Ro} (\partial_t \mathbf{u}_h + \mathbf{u}_h \cdot \nabla_h \mathbf{u}_h + \text{Ro} w \partial_z \mathbf{u}_h) + f \mathbf{u}_h^\perp - \alpha \text{Ro} \boldsymbol{\Omega}_h^\perp w = -\nabla_h \psi, \quad (1.40a)$$

$$\alpha^2 \text{Ro}^2 (\partial_t w + \mathbf{u}_h \cdot \nabla_h w + \text{Ro} w \partial_z w) + \alpha \boldsymbol{\Omega}_h^\perp \cdot \mathbf{u}_h = -\partial_z \psi - \rho, \quad (1.40b)$$

$$\nabla_h \cdot \mathbf{u}_h + \text{Ro} \partial_z w = 0. \quad (1.40c)$$

The quasi-geostrophic equations are obtained by the leading order of this system in the formal limit $\text{Ro} \rightarrow 0$ and $\alpha \rightarrow 0$. The leading-order vertical momentum equation is, once again, the hydrostatic balance relation

$$\partial_z \psi = -\rho. \quad (1.41)$$

To determine the leading-order balance in the horizontal momentum equation (1.40a), we need to separate the divergence from the curl component. For the divergence, the dominant contribution is geostrophic balance, i.e.,

$$\mathbf{u}_h = \nabla_h^\perp \psi. \quad (1.42)$$

(Note that there is no f in this relation as the deviation from constant Coriolis parameter comes in at $O(\text{Ro})$ only.) The leading order of the curl of (1.40a) could be obtained by direct computation, but it is easier to work from the expression for the Boussinesq potential vorticity which, in dimensional variables, is given by (1.7). In non-dimensional variables, the potential vorticity reads

$$\begin{aligned} q &= \left(f_0 \boldsymbol{\Omega}_h + \frac{U}{H} \partial_z \mathbf{u}_h^\perp - \frac{W}{L} \nabla_h^\perp w \right) \cdot \frac{\Gamma}{L} \nabla_h \rho \\ &\quad + \left(f_0 f + \frac{U}{L} \nabla_h^\perp \cdot \mathbf{u}_h \right) \left(\frac{\Gamma}{H} \partial_z \rho + \frac{N_0^2 \rho_0}{g} \partial_z \bar{\rho} \right) \\ &= \frac{U\Gamma}{HL} \left(\left(\frac{\alpha}{\text{Ro}} \boldsymbol{\Omega}_h + \partial_z \mathbf{u}_h^\perp - \alpha^2 \text{Ro} \nabla_h^\perp w \right) \cdot \nabla_h \rho \right. \\ &\quad \left. + \left(\frac{1}{\text{Ro}} + \beta y + \nabla_h^\perp \cdot \mathbf{u}_h \right) \left(\partial_z \rho + \frac{1}{\text{Ro}} \partial_z \bar{\rho} \right) \right) \end{aligned} \quad (1.43)$$

and satisfies the advection equation

$$(\partial_t + \mathbf{u}_h \cdot \nabla_h + \text{Ro} w \partial_z) q = 0. \quad (1.44)$$

Since $\bar{\rho}$ does not depend on time or on the horizontal position, the leading-order terms appear at $O(\text{Ro}^{-1})$ and read

$$D_t^h (\partial_z \rho + \nabla_h^\perp \cdot \mathbf{u}_h \partial_z \bar{\rho} + \beta y \partial_z \bar{\rho}) + w \partial_{zz} \bar{\rho} = 0, \quad (1.45)$$

where $D_t^h = \partial_t + \mathbf{u}_h \cdot \nabla_h$ denotes the horizontal material derivative. Eliminating w with the leading-order terms of (1.39), dividing through by $\partial_z \bar{\rho}$ and simplifying, we obtain the quasi-geostrophic potential vorticity equation,

$$D_t^h \left(\partial_z \frac{\rho}{\partial_z \bar{\rho}} + \nabla_h^\perp \cdot \mathbf{u}_h + \beta y \right) = 0. \quad (1.46)$$

Using hydrostatic balance, the leading order of (1.40b) and the curl of geostrophic balance (1.42), the quasi-geostrophic equations can be written in the well-known closed form

$$D_t^h \left(\Delta_h \psi - \partial_z \frac{\partial_z \psi}{\partial_z \bar{\rho}} + \beta y \right) = 0, \quad (1.47a)$$

$$\mathbf{u}_h = \nabla_h^\perp \psi. \quad (1.47b)$$

The advected quantity in (1.47a) is the quasi-geostrophic potential vorticity. To recover the stream function ψ from the potential vorticity, we need to solve a second-order equation. It is elliptic provided that $\partial_z \bar{\rho} < 0$, i.e., the fluid is stably stratified. Moreover, we need boundary conditions. At the lateral boundaries, the no-flux condition $\mathbf{n} \cdot \mathbf{u} = 0$ constrains only the tangential derivative of ψ . A consistent lateral boundary condition is $\psi = 0$ on a simply-connected domain; for the multiply-connected case, see McWilliams (1977). In more idealized situations, channel geometries or, on the f -plane, periodic boundary conditions are frequently used.

At the top and bottom boundaries, we use Neumann boundary conditions. Due to hydrostatic balance, $\partial_z \psi = -\rho$, where ρ satisfies the leading order of (1.39),

$$D_t^h \rho + w \partial_z \bar{\rho} = 0. \quad (1.48)$$

At the top boundary, it is common to assume rigid lid conditions, i.e., $w = 0$. For a correction to a rigid lid condition, see, e.g., Olbers et al. (2012). At the bottom boundary, the impermeability condition reads $\mathbf{u}_h \cdot \nabla_h b + w = 0$, with $b(\mathbf{x}_h)$ denoting the equilibrium depth.

To re-dimensionalize the quasi-geostrophic equations, we recall the definition of the z -dependent Brunt–Väisälä frequency (1.11). Further, it is convenient to express ψ in units of a horizontal stream function, not in pressure units. Then equation (1.47) takes the form

$$\partial_t q + \nabla_h^\perp \psi \cdot \nabla q = 0, \quad (1.49a)$$

$$q = f + \Delta_h \psi + f_0^2 \partial_z \frac{\partial_z \psi}{N^2(z)}, \quad (1.49b)$$

where q is the *quasi-geostrophic potential vorticity*. Linearized about the trivial solution, these equations allow propagation of internal waves in the horizontal plane. Their phase velocities c_n are related to the eigenvalues $\lambda_n = c_n^{-2}$ of the vertical structure operator, the last term in (1.49b). Of particular importance is the speed c_1 of the fastest wave. The associated horizontal length-scale is called the (*first*) *internal or baroclinic Rossby radius of deformation*,

$$L_d = \frac{c_1}{f_0} = \frac{N_0 H}{\pi f_0}; \quad (1.50)$$

the second equality holds as stated only in the case of uniform stratification where $N(z) = N_0$. It is the scale at which buoyancy forces and Coriolis forces are equally important; it is also approximately the scale of strongest conversion of potential into

kinetic energy via baroclinic instability; see, e.g., Vallis (2006) for details. In the ocean, L_d varies from less than 10 km at high latitudes up to 200 km in the tropics.

As the quasi-geostrophic equations are typically used for proof-of-concept studies, two simplifications are particularly useful. First, if the equations are posed in a flat horizontal layer and density and velocity are independent of z , we obtain the two-dimensional or barotropic quasi-geostrophic equation

$$\partial_t q + \nabla^\perp \psi \cdot \nabla q = 0, \quad (1.51a)$$

$$q = f + \Delta \psi, \quad (1.51b)$$

where we drop the subscript “h” on the operators as all fields are fully two-dimensional. This form of the barotropic quasi-geostrophic equation implicitly carries a rigid lid assumption. It is possible to derive the equation without this assumption, in which case the expression for potential vorticity acquires an extra term; see (1.69) and the surrounding discussion.

Second, we can derive a simple model for a baroclinic rotating flow by assuming that the fluid moves in two uniform layers with constant depths H_1 and H_2 , respectively, with layer 1 assumed on top of layer 2. We suppose that the potential vorticities q_i and stream functions ψ_i , where $i \in \{1, 2\}$, are taken at the layer centers. The density, on the other hand, is taken at the layer boundaries. Using finite differences, $\partial_z \psi \approx 2(\psi_1 - \psi_2)/(H_1 + H_2)$ at the layer interface. At the top and bottom interfaces, boundary condition (1.48) with $w = 0$ applies and the hydrostatic relation implies that $\partial_z \psi$ is advected. Altogether, using finite differences across each layer for the outer z -derivative in (1.49b), we obtain the quasi-geostrophic two-layer equations

$$(\partial_t + \nabla^\perp \psi_1 \cdot \nabla) \left(f + \Delta \psi_1 + \frac{2 f_0^2}{N_0^2 H_1 (H_1 + H_2)} (\psi_2 - \psi_1) \right), \quad (1.52a)$$

$$(\partial_t + \nabla^\perp \psi_2 \cdot \nabla) \left(f + \Delta \psi_2 + \frac{2 f_0^2}{N_0^2 H_2 (H_1 + H_2)} (\psi_1 - \psi_2) \right). \quad (1.52b)$$

When the two layers have equal depth $H_1 = H_2 = H/2$, we can write the two-layer quasi-geostrophic system in the symmetric form

$$\partial_t q_i + \nabla^\perp \psi_i \cdot \nabla q_i = 0, \quad (1.53a)$$

$$q_i = f + \Delta \psi_i + (-1)^i k_d^2 (\psi_1 - \psi_2)/2, \quad (1.53b)$$

with $k_d = L_d^{-1}$, where the (single) internal Rossby radius is now given by

$$L_d = \frac{N_0 H}{\sqrt{8} f_0}. \quad (1.54)$$

A detailed exposition of more general multi-level and multi-layer models can be found in Pedlosky (1987) or Kamenkovich et al. (1986). For a variational perspective on quasi-geostrophic theory, see Holm and Zeitlin (1998) and Bokhove et al. (1998).

1.2.6 Rotating Shallow Water Equations

We shall finally introduce the rotating shallow water equations which describe the horizontal motion in a thin layer of an incompressible fluid with a free surface and constant density. The shallow water equations are often used as a simple test bed in situations where baroclinic effects are negligible or to be excluded.

Assuming that $v = 0$ and, for simplicity, that density is constant, the Boussinesq equations (1.1) reduce to the three-dimensional Euler equations for a homogeneous ideal fluid with rotation and gravitational forces acting in the vertical. Following the notation used above, they read

$$D_t \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\frac{1}{\rho_0} \nabla p - g\mathbf{k}, \quad (1.55a)$$

$$\nabla \cdot \mathbf{u} = 0. \quad (1.55b)$$

We now make the following scaling assumptions. First, as before, we impose an advective time-scale (1.23). Second, we assume that the aspect ratio α , defined in (1.24), is small. Without stratification, we only have the divergence condition (1.55b) to constrain the vertical velocity, so that $W \lesssim \alpha U$. Third, we suppose that

$$\text{Ro} \lesssim \text{Bu} = \frac{gH}{f_0^2 L^2}, \quad (1.56)$$

where the dimensionless parameter Bu is known as the Burger number, which can also be interpreted as the square ratio of the Rossby radius of deformation to the horizontal length-scale L . We remark that condition (1.56) is consistent with both the semi-geostrophic and the quasi-geostrophic scaling regimes of shallow water theory; see the discussion in Section 1.2.7 below.

Under these assumptions, the formal limit $\alpha \rightarrow 0$ imposes, again simultaneously, the hydrostatic and the traditional approximation, so that, using a rescaled pressure but otherwise dimensional variables,

$$D_t \mathbf{u}_h + f \mathbf{u}_h^\perp = -\nabla_h p, \quad (1.57a)$$

$$0 = \partial_z p - g, \quad (1.57b)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (1.57c)$$

The hydrostatic equation (1.57b) implies that the pressure is fully determined by the hydrostatic pressure. To be definite, let us suppose that $z = 0$ describes the equilib-

rium free surface, $h(\mathbf{x}_h, t)$ the departure of the free surface from equilibrium, and $b(\mathbf{x}_h)$ the equilibrium depth of the layer. If the pressure at the free surface is zero, integration of (1.57b) yields

$$p(\mathbf{x}_h, t) = g(h(\mathbf{x}_h, t) - z). \quad (1.58)$$

In particular, the horizontal pressure gradient is entirely independent of z . Thus, *assuming* that the velocity field is initially independent of z , it will remain so for all times, and the material advection operator D_t in (1.57a) reduces to horizontal material advection D_t^h . Finally, we can eliminate w from the divergence condition (1.57c) by vertical integration under zero normal flow boundary conditions, yielding a continuity equation for the total layer depth. Dropping the h-subscript, as all quantities are now fully two-dimensional, the resulting equations, known as the *rotating shallow water equations*, read

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + f \mathbf{u}^\perp + g \nabla h = 0, \quad (1.59a)$$

$$\partial_t h + \nabla \cdot ((h + b)\mathbf{u}) = 0. \quad (1.59b)$$

These equations can also be derived from a variational principle; see, e.g., Salmon (1998). The generalization to stratified rotating shallow water equations is based on assuming that the fluid consists of layers with constant density that are separated by interfaces. As for the above single-layer rotating shallow water equations, one assumes that the amplitude of the deformation of each interface to be much less than the layer depths. For a variational derivation of the multi-layer rotating shallow water equations, see, e.g., Salmon (1982) and Stewart and Dellar (2010).

The rotating shallow water equations conserve the energy

$$H = \frac{1}{2} \int_{\Omega} (h + b) |\mathbf{u}|^2 + g h^2 \, d\mathbf{x}, \quad (1.60)$$

and materially conserve the potential vorticity

$$q = \frac{f + \zeta}{h + b} = \frac{\zeta_a}{h + b} \quad (1.61)$$

where

$$\zeta = \nabla^\perp \cdot \mathbf{u} \quad (1.62)$$

is the *relative vorticity* and $\zeta_a = f + \zeta$ the *absolute vorticity*. Both conservation laws can be shown to arise as Noetherian conservation laws as outlined in Section 1.3.1.

To study emission and propagation of gravity waves, it is useful to reformulate the rotating shallow water equations in terms of imbalance variables as we have done for the rotating Boussinesq equations in Section 1.2.2. Following Salmon (2007), we use vorticity ζ , divergence $\delta = \nabla \cdot \mathbf{u}$, and height field h . The reason is that geostrophic motion is divergence free, so that divergence δ and *ageostrophic vorticity*

$\gamma = f\xi - g\Delta h$ can be used as dynamic indicators for imbalance. Introducing a *momentum stream function* χ and a *momentum potential* θ so that

$$(h + b)\mathbf{u} = \nabla^\perp \chi + \nabla \theta \quad (1.63)$$

and a generalized Bernoulli function

$$\Phi = \frac{1}{2} |\mathbf{u}|^2 + gh, \quad (1.64)$$

we can write the rotating shallow water equations in the form

$$\partial_t \xi = \nabla^\perp q \cdot \nabla \chi - \nabla \cdot (q \nabla \theta), \quad (1.65a)$$

$$\partial_t \delta = \nabla^\perp q \cdot \nabla \theta + \nabla \cdot (q \nabla \chi) - \Delta \Phi, \quad (1.65b)$$

$$\partial_t h = -\Delta \theta. \quad (1.65c)$$

We remark that Mohebalhojeh and Dritschel (2001) prefer a very similar formulation using δ and γ as imbalance variables. In particular, they derive high-order balance conditions, termed δ - γ balance, by requiring any number of time derivatives of δ and γ to vanish. Note that when periodic boundary conditions are used in idealized prototype studies, system (1.65) needs to be augmented with two ordinary differential equations for the mean velocity field; see, e.g., the discussion in Dritschel et al. (2017).

1.2.7 Geostrophic Scalings

The Boussinesq equations with a free surface upper boundary also support surface gravity waves. Surface waves are retained in the shallow water approximation, and textbook linear wave theory shows that their maximal phase velocity is bounded by $c_e = \sqrt{gH}$. This speed defines, as does (1.50) for the internal modes, the *external* or *barotropic Rossby radius of deformation*

$$L_e = \frac{c_e}{f_0} = \frac{\sqrt{gH}}{f_0} \quad (1.66)$$

as the horizontal scale at which free surface effects and rotation are of equal importance. In the deep ocean, for example, $L_e \approx 2000$ km, thus it is often appropriate to consider a rigid lid which imposes $L_e = \infty$ and focus on the first internal Rossby radius L_d . In the atmosphere, the two radii are much closer with $L_e \approx 2000$ km and $L_d \approx 800$ km.

Geostrophic scalings are introduced to capture the regime when rotational forces are dominant. This regime is universally characterized by a small Rossby number (1.28), but the limit can be approached in different ways distinguished by the scaling of the *Burger number*

$$\text{Bu} = \frac{L_{\text{Ro}}^2}{L^2} = \frac{\text{Ro}^2}{\text{Fr}^2}, \quad (1.67)$$

where, depending on the context, L_{Ro} denotes the internal or external Rossby radius of deformation. In the internal wave picture, the Froude number is given by (1.26). For surface gravity waves in the shallow water framework, (1.67) can be seen as implicitly defining the shallow water Froude number which, however, is not needed as an independent scaling parameter.

Qualitatively, there are three different regimes. When the Burger number is small, the flow is rotation dominated. The *semi-geostrophic* limit, discussed further below, is representative of this regime. At $\text{Bu} = O(1)$, buoyancy and rotation are both important. The *quasi-geostrophic* limit is representative of this regime. For large Burger numbers, the effects of stratification are dominant; we will not consider this situation further. See, e.g., Babin et al. (2002) for a detailed exposition of the different geostrophic scaling regimes.

We already took a detailed look at quasi-geostrophy in Section 1.2.5. There, we took the rotating Boussinesq equations as the starting point and scaled with $\text{Bu} = O(1)$ where the Burger number was linked to the internal Rossby radius. We specialized single-layer and double-layer models in a second step. However, it is also possible to derive a single-layer quasi-geostrophic equation from the rotating shallow water equations. In this case, buoyancy comes from the free surface elevation, so the Burger number must be based on the external Rossby radius. In order to maintain geostrophic balance at leading order, we must assume that variations in the surface elevation are small, more precisely, are $O(\text{Ro})$. It is then possible to derive a quasi-geostrophic potential vorticity by linearizing the shallow water potential vorticity (1.61) as follows. Taking a constant layer depth $b = H$ with $h \ll H$ and $\zeta \ll f$, we can write

$$q = \frac{f + \zeta}{H} \frac{1}{1 + h/H} \approx \frac{f + \zeta}{H} \left(1 - \frac{h}{H}\right) \approx \frac{1}{H} \left(f - f \frac{h}{H} + \zeta\right). \quad (1.68)$$

Dropping the constant prefactor, recalling that $\zeta = \Delta\psi$, and using leading-order geostrophic balance in the form $f\psi \approx gh$, we obtain a new quasi-geostrophic potential vorticity

$$q = f - L_e^{-2} \psi + \Delta\psi. \quad (1.69)$$

This expression contains an extra *stretching term* $L_e^{-2} \psi$ not present in (1.51b) which is a remnant of free surface effects. It is required on scales larger than the external Rossby radius, even though the use of the quasi-geostrophic approximation on such scales becomes questionable as changes in the Coriolis parameter are no longer small.

Let us now turn our attention to the situation when the Burger number is small. In this case, the most important distinguished limit has $\text{Bu} = \text{Ro}$. It is referred to as the *semi-geostrophic* limit or the *frontal geostrophic* regime in the context of frontal geostrophic adjustment (Zeitlin et al. 2003). The study of this limit goes

back to the geostrophic momentum approximation (Eliassen 1948). The resulting semi-geostrophic equations were rewritten by Hoskins and solved via an ingenious change of coordinates (Hoskins 1975; Cullen and Purser 1984). They continue to attract interest due to their connection to optimal transport theory and the resulting possibility to make mathematical sense of generalized frontal-type solutions (Benamou and Brenier 1998; Cullen 2008) and for the turbulence emerging from them (Ragone and Badin 2016). The geostrophic momentum approximation and Hoskins' transformation inspired Salmon (1983, 1985) to make corresponding approximations directly to Hamilton's principle so as to preserve geometrical structure and automatically preserve conservation laws. Salmon's approach is generalized in Oliver (2006); corresponding results for stratified flow where the primitive equations serve as the parent model are due to Salmon (1996) and Oliver and Vasylykevych (2016). Numerical studies suggest that Salmon's so-called L_1 model is particularly robust and accurate (Allen et al. 2002; Dritschel et al. 2017). A direct numerical comparison of Hoskins' semi-geostrophic equations with its generalized solution structure and the L_1 family of models which possess global classical solutions has not yet been done.

Both semi-geostrophic and quasi-geostrophic models cannot support inertia-gravity waves. Rossby waves are the only possible linear wave solutions; see, e.g., Vallis (2006) for quasi-geostrophic Rossby wave theory. Thus, even though quasi-geostrophic models formally allow larger Burger numbers, their derivation imposes a smallness assumption on buoyancy perturbations, so that it is *a priori* not clear whether they are more accurate than any of the semi-geostrophic models for nearly balanced flow even in the $Bu = O(1)$ regime. Despite some preliminary attempts aimed at exploring frontal scale turbulence (Badin 2014), to our knowledge a systematic numerical study has not yet been done.

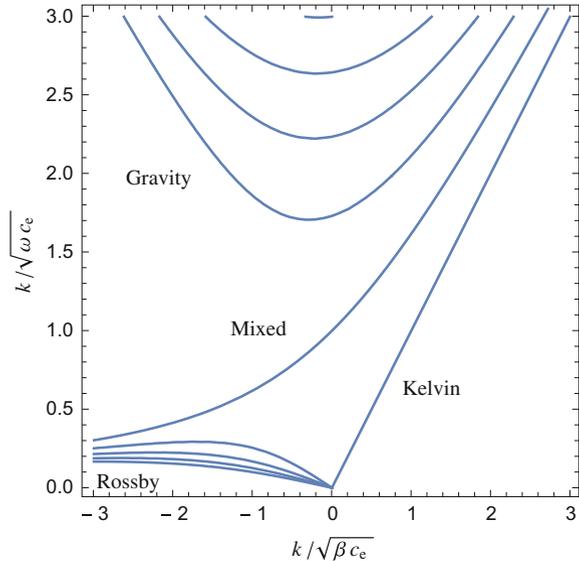
Mathematically rigorous justifications of the quasi-geostrophic splitting into fast and slow modes have started with the work of Embid and Majda (1996) and Babin et al. (1996, 1997). These ideas are extended and reviewed, for example, by Babin et al. (2002), Majda (2003), Saint-Raymond (2010), and Cheng and Mahalov (2013). The book by Majda (2003), in particular, contains an elementary exposition of averaging over fast waves which can be seen as a deterministic precursor of the stochastic averaging in Section 1.5.3 below, and Dutrifoy et al. (2009) consider a model for equatorial balance.

A full mathematical justification for any of the semi-geostrophic limits remains, to the best of our knowledge, open.

1.2.8 Equatorial Scalings

Near the equator, the mid-latitude scalings discussed above break down as the rotation vector $\boldsymbol{\Omega}$ aligns with the horizontal. In the equatorial β -plane approximation, $f_0 = 0$ and the vertical component of the Coriolis force takes the form $f = \beta y$. Correspondingly, the mid-latitude notion of geostrophic balance breaks down. Thus, near the

Fig. 1.1 Dispersion curves of linear equatorial waves of the rotating shallow water equations. Here k is the wavenumber and ω the frequency with β the meridional gradient of the Coriolis parameter and $c_e = \sqrt{gH}$ the gravity wave speed for mean fluid thickness H



equator the fluid ceases to be approximately constrained to quasi-two-dimensional motion. Moreover, at least in the atmosphere, different physics—moist processes and deep convection—become dominant features of the observed dynamics. Whether some notion of balance in the absence of moist processes persists up to and across the equator is currently not well understood.

One instance where these difficulties manifest themselves is the weakening of scale separation between different types of linear waves as compared to the mid-latitudes. For example, linearizing the equatorial rotating shallow water equations about a steady state with constant height field yields the dispersion curves for eigenmodes plotted in Figure 1.1. In addition to the slow Rossby and Kelvin waves also Yanai waves, which are mixed Rossby–gravity waves, are present. Here, only the Rossby waves reflect the non-trivial Coriolis effect of the β -plane approximation and are geostrophically balanced. However, Kelvin waves are important factors in low-frequency variations in the tropics.

As in the mid-latitudes, it is natural to seek a scaling which is able to filter the faster waves in some asymptotic limit. The mid-latitude definition of the Rossby number, where $\text{Ro} = U/(fL)$, becomes singular at the equator and cannot be used as a scaling parameter. Noting that the characteristic length-scales may be strongly anisotropic near the equator, we may proceed as follows. Let c denote the characteristic wave speed. For example, in the shallow water approximation, $c = c_e = \sqrt{gH}$ while for stratified quasi-geostrophic or Boussinesq flow, $c = c_1 = N_0H/\pi$ as introduced in the previous sections. The meridional length-scale L_y at which rotation is comparable to wave propagation satisfies $fL_y = c$. Moreover, at this distance from the equator, the vertical Coriolis parameter satisfies $f = \beta L_y$. Solving for the meridional length-scale, we obtain

$$L_y = \sqrt{\frac{c}{\beta}}, \quad (1.70)$$

which can be regarded as an *equatorial Rossby radius of deformation*. In combination with a typical zonal scale L_x , we may define an *equatorial Rossby number* as

$$\text{Ro}_{\text{eq}} = \frac{U}{fL_x} = \frac{U}{\sqrt{c\beta} L_x}. \quad (1.71)$$

Its ratio with the Froude number $\text{Fr} = U/c$ leads to the *equatorial Burger number* as the horizontal aspect ratio

$$\text{Bu} = \frac{L_y^2}{L_x^2} = \frac{\text{Ro}_{\text{eq}}^2}{\text{Fr}^2}. \quad (1.72)$$

It is typically small; $\varepsilon = \sqrt{\text{Bu}}$ has been used as the main expansion parameter, for example, by Majda (2003) and Chan and Shepherd (2013).

Assuming, in addition, a low vertical aspect ratio in the sense that $H^2 \ll L_x L_y$, one obtains hydrostatic balance as for the mid-latitude scaling in Section 1.2.4 (Chan and Shepherd 2013). In order to obtain a hierarchy of balanced models from an ε -expansion, they assume $\text{Fr} = 1$ and apply the slaving method of Warn et al. (1995). Balanced models in the tropics which do not assume the traditional approximation were derived by Julien et al. (2006).

1.3 Variational Principles and Hamiltonian Mechanics

The inviscid Boussinesq equations and all of the simplified models discussed in Section 1.2 are infinite-dimensional Hamiltonian systems. In this section, we will review several formally equivalent points of view: Hamilton's variational principle in which the equations of motion arise as stationary points of an action functional, the Poisson formulation of Hamiltonian fluid mechanics, and, closely related, the Nambu formulation.

The derivation of each of the models from Section 1.2 can be made systematic through the use of Hamilton's variational principle. This approach has multiple advantages: First of all, it allows to formulate a geometrical setting of the dynamics; from the study of the Lagrangian density, one can study the continuous symmetries of the system and, by Noether's theorem, derive associated conservation laws; further, one can apply different dynamical approximations directly to the Lagrangian density. When the approximations respect the continuous symmetries, the approximated systems will possess analogous conserved quantities. Fluid Lagrangians, in particular, have a special symmetry, the particle relabeling symmetry, which states that the exchange of fluid labels does not affect the distribution of mass. The associ-

ated conserved quantity to this symmetry is the potential vorticity of the fluid. This result shows that potential vorticity does not just appear from a skillful manipulation of the equations of motion, but is instead a signature of a more fundamental property of the system.

While Hamilton's principle has been used to derive numerical methods for partial differential equations (Stern et al. 2015), structure-preserving numerical approximations for fluid equations more readily arise by discretizing Poisson or Nambu brackets. While we will not go into the numerical aspects here, we explain the Poisson and Nambu formulations using the rotating shallow water equations as an example.

More detailed expositions in the context of geophysical fluid dynamics can be found in the textbooks by Salmon (1998) and Badin and Crisciani (2018); for the mathematical foundations, see, for example, Marsden and Ratiu (2013) and Arnold and Khesin (1999).

1.3.1 Variational Principles

The inviscid form of the Boussinesq equations (1.1) can be derived from a variational principle as follows. For simplicity, we consider the case of a rigid lid upper boundary. Let \mathfrak{g} denote the Lie algebra of vector fields on Ω satisfying the incompressibility condition (1.1b) on the domain Ω with impermeability conditions on all boundaries. Let η denote the flow of a time-dependent vector field $\mathbf{u} \in \mathfrak{g}$, i.e.,

$$\dot{\eta}(\mathbf{a}, t) = \mathbf{u}(\eta(\mathbf{a}, t), t) \quad \text{with} \quad \eta(\mathbf{a}, 0) = \mathbf{a}. \quad (1.73)$$

Here and in the following, the letter \mathbf{a} is used for Lagrangian label coordinates, while $\mathbf{x} = \eta(\mathbf{a}, t)$ denotes the corresponding Eulerian position at time t . As \mathbf{u} is divergence free, η is volume preserving. In the following, we shall write $\dot{\eta} = \mathbf{u} \circ \eta$ for short. Correspondingly, the advection of density equation, (1.1c) with $\kappa = 0$, is equivalent to

$$\rho \circ \eta = \rho_{\text{in}}, \quad (1.74)$$

where ρ_{in} is the given initial distribution of the density.

Throughout the chapter, we use the letters L and ℓ (with appropriate subscripts as necessary) to distinguish Lagrangians expressed in Lagrangian and Eulerian quantities, respectively. With this notation in place, the Boussinesq Lagrangian reads

$$L(\eta, \dot{\eta}; \rho_{\text{in}}) = \int_{\Omega} \mathbf{R} \circ \eta \cdot \dot{\eta} + \frac{1}{2} |\dot{\eta}|^2 - \frac{g}{\rho_0} \rho_{\text{in}} \eta_3 \, d\mathbf{a}, \quad (1.75)$$

where $\nabla \times \mathbf{R} = 2\boldsymbol{\Omega}$. We note that L can be expressed in terms of purely Eulerian quantities as

$$L(\boldsymbol{\eta}, \dot{\boldsymbol{\eta}}; \rho_{\text{in}}) = \int_{\Omega} \mathbf{R} \cdot \mathbf{u} + \frac{1}{2} |\mathbf{u}|^2 - \frac{g}{\rho_0} \rho z \, d\mathbf{x} \equiv \ell(\mathbf{u}, \rho). \quad (1.76)$$

The first term in the Lagrangian is the Coriolis term. It only contributes to the symplectic form, but does not feature in the energy. The second and third terms are the difference of kinetic and potential energies, as for simple mechanical systems.

We observe that L is invariant under compositions of the flow map with arbitrary volume- and domain-preserving maps. This is known as the *particle relabeling symmetry*. For such Lagrangians, the Euler–Poincaré theorem for continua (Holm et al. 1998 or Holm et al. 2009, Theorem 17.8) asserts that the following are equivalent.

(i) $\boldsymbol{\eta}$ satisfies the *variational principle*

$$\delta \int_{t_1}^{t_2} L(\boldsymbol{\eta}, \dot{\boldsymbol{\eta}}; \rho_{\text{in}}) \, dt = 0 \quad (1.77)$$

with respect to variations of the flow map $\delta\boldsymbol{\eta} = \mathbf{w} \circ \boldsymbol{\eta}$ where \mathbf{w} is a curve in \mathfrak{g} vanishing at the temporal endpoints.

(ii) \mathbf{u} and ρ satisfy the *reduced variational principle*

$$\delta \int_{t_1}^{t_2} \ell(\mathbf{u}, \rho) \, dt = 0, \quad (1.78)$$

where the variations $\delta\mathbf{u}$ and $\delta\rho$ are subject to the Lin constraints

$$\delta\mathbf{u} = \dot{\mathbf{w}} + \nabla\mathbf{w}\mathbf{u} - \nabla\mathbf{u}\mathbf{w} = \dot{\mathbf{w}} + [\mathbf{u}, \mathbf{w}], \quad (1.79a)$$

$$\delta\rho + \mathbf{w} \cdot \nabla\rho = 0, \quad (1.79b)$$

with \mathbf{w} as in (i).

(iii) \mathbf{m} and ρ satisfy the *Euler–Poincaré equation*

$$\int_{\Omega} (\partial_t + \mathcal{L}_{\mathbf{u}})\mathbf{m} \cdot \mathbf{w} + \frac{\delta\ell}{\delta\rho} \mathcal{L}_{\mathbf{w}}\rho \, d\mathbf{x} = 0 \quad (1.80)$$

for every $\mathbf{w} \in \mathfrak{g}$, where \mathcal{L} denotes the Lie derivative and \mathbf{m} is the momentum one-form

$$\mathbf{m} = \frac{\delta\ell}{\delta\mathbf{u}}. \quad (1.81)$$

In the language of vector fields in a region of \mathbb{R}^3 , the Euler–Poincaré equation (1.80) reads

$$\int_{\Omega} \left(\partial_t \mathbf{m} + (\nabla \times \mathbf{m}) \times \mathbf{u} + \nabla(\mathbf{m} \cdot \mathbf{u}) + \frac{\delta\ell}{\delta\rho} \nabla\rho \right) \cdot \mathbf{w} \, d\mathbf{x} = 0 \quad (1.82)$$

for every $\mathbf{w} \in \mathfrak{g}$. Due to the Hodge decomposition, the term in parentheses must be a gradient, i.e.,

$$\partial_t \mathbf{m} + (\nabla \times \mathbf{m}) \times \mathbf{u} + \frac{\delta \ell}{\delta \rho} \nabla \rho = \nabla \phi. \quad (1.83)$$

Noting that

$$\frac{\delta \ell}{\delta \mathbf{u}} = \mathbf{R} + \mathbf{u} \quad \text{and} \quad \frac{\delta \ell}{\delta \rho} = -\frac{gz}{\rho_0}, \quad (1.84)$$

redefining the potential

$$\phi = -\frac{p}{\rho_0} - \frac{g}{\rho_0} z \rho - \frac{1}{2} |\mathbf{u}|^2, \quad (1.85)$$

and using the vector identity $(\nabla \times \mathbf{u}) \times \mathbf{u} = \mathbf{u} \cdot \nabla \mathbf{u} - \frac{1}{2} \nabla |\mathbf{u}|^2$, we can write the Euler–Poincaré equation for L as

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\frac{1}{\rho_0} \nabla p - \frac{g\rho}{\rho_0} \mathbf{k}. \quad (1.86)$$

Thus, we recovered the inviscid form of the momentum equation (1.1a).

We remark that the traditional approach to variational derivation of the primitive equations treats the geopotential as a Lagrange multiplier responsible for enforcing the incompressibility constraint. Here, we build the constraint into the definition of the configuration space. The gradient of the geopotential then appears naturally due to the fact that the L^2 pairing with divergence-free vector fields determines a vector field only up to a gradient. Both approaches, of course, lead to identical equations of motion.

As the Boussinesq Lagrangian (1.75) is invariant under time translation, the model possesses a conserved energy of the form

$$H = \int_{\Omega} \frac{\delta \ell}{\delta \mathbf{u}} \cdot \mathbf{u} \, dx - \ell(\mathbf{u}, \rho) = \int_{\Omega} \frac{1}{2} |\mathbf{u}|^2 + \frac{g}{\rho_0} \rho z \, dx. \quad (1.87)$$

The symmetry of the balance model Lagrangian under particle relabeling leads to a material conservation law for the potential vorticity of the balance model. It can be derived geometrically as follows. First note that the abstract Euler–Poincaré equation (1.80) can be written

$$(\partial_t + \mathcal{L}_{\mathbf{u}}) \mathbf{m} + \frac{\delta \ell}{\delta \rho} d\rho = d\phi, \quad (1.88)$$

where d denotes the exterior derivative. Taking the exterior (wedge) product between the exterior derivative of (1.88) and $d\rho$, we find that

$$\begin{aligned}
0 &= d((\partial_t + \mathcal{L}_u)\mathbf{m} + \frac{\delta \ell}{\delta \rho} d\rho - d\phi) \wedge d\rho \\
&= (\partial_t + \mathcal{L}_u)(d\mathbf{m} \wedge d\rho) - d\mathbf{m} \wedge d(\partial_t + \mathcal{L}_u)\rho \\
&= (\partial_t + \mathcal{L}_u)(d\mathbf{m} \wedge d\rho),
\end{aligned} \tag{1.89}$$

where we used the commutativity of Lie and exterior derivatives in the second equality and the advection of ρ in the third equality. In three dimensions, we can identify this conservation law with material advection of the scalar quantity

$$q = *(d\mathbf{m} \wedge d\rho), \tag{1.90}$$

where $*$ denotes the Hodge dual operator. Indeed, writing $\mu = dx_1 \wedge dx_2 \wedge dz$ to denote the canonical volume form, we have $d\mathbf{m} \wedge d\rho = q\mu$, so that

$$0 = (\partial_t + \mathcal{L}_u)(q\mu) = \mu (\partial_t + \mathcal{L}_u)q + q \mathcal{L}_u\mu. \tag{1.91}$$

Since the flow is volume preserving and μ is non-degenerate, this proves that $\partial_t q + \mathcal{L}_u q = 0$, i.e., q is conserved on fluid particles.

Going back to the language of vector calculus, using expression (1.84) for the momentum and writing $\boldsymbol{\omega} = \nabla \times \mathbf{u}$ for the relative vorticity, we recover the familiar expression for the *Ertel potential vorticity*,

$$q = (\nabla \times \mathbf{m}) \cdot \nabla \rho = (2\boldsymbol{\Omega} + \boldsymbol{\omega}) \cdot \nabla \rho. \tag{1.92}$$

We remark that the derivation above corresponds to taking the inner product of the curl of the Euler–Poincaré equations (1.83) with $\nabla \rho$ and manipulating correspondingly; the advantage of the abstract approach is that commuting exterior and Lie derivative in traditional notation is not linked to any intrinsic operation, thus requires tedious verification.

1.3.2 Variational Model Reduction

The variational principle which underlies the equations of motion can be used to derive simplified models. All approximations are done at the level of the Lagrangian; the simplified equations of motion then arise from the approximated Lagrangian in a second step. When the approximations respect the symmetries of the Lagrangian, the associated conservation laws, in our setting the conservation of energy and of potential vorticity, will be preserved as well.

The use of the Hamilton principle to derive balance models was pioneered by Salmon (1985, 1996). Salmon’s method was generalized by Oliver (2006) where the approximate balance manifold is interpreted as a Dirac constraint induced by a truncated change of coordinates. Structure preservation, however, does not imply

well-posedness of the initial value problem, as is evident in the numerical study of Dritschel et al. (2017) which indicates a strong preference for Salmon’s so-called L_1 -model. Çalik et al. (2013) prove global well-posedness for a large class of variational balance models, including the L_1 -model. Gottwald and Oliver (2014) give a justification of the general method to arbitrary order of approximation in a finite-dimensional model context, and Oliver and Vasylykevych (2013, 2016) show that the variational approach is flexible enough to cover spatially varying Coriolis functions and stratified flow, respectively.

The use of language from geometry and of the associated variational principles is particularly advantageous when working in general curvilinear coordinates (Tort and Dubos 2014) which is crucial when working with global models but not as pertinent to the more theoretical questions we raise here and will not be discussed further.

In a different line of research, Koide and Kodama (2012) and references cited therein demonstrate that a stochastic version of the Hamilton principle can give rise to various dissipative partial differential equations, including the Navier–Stokes equations and higher-order corrections to harmonic diffusion. A different approach is taken by Holm (2015) who derives fully stochastic fluid equations from a variational principle and discusses their conservation law structure.

1.3.3 Poisson Formulation

All of the models above can also be cast in a Poisson or a Nambu formulation. We illustrate this using the rotating shallow water equations as an example.

The rotating shallow water equations (here, for simplicity, with $b = \text{const}$ and h denoting the total layer depth) can be cast in a non-canonical Poisson formulation (see, e.g., Shepherd 1990). Suppose F is an arbitrary functional of \mathbf{u} and h . Then

$$\dot{F} = \{F, H\} \quad (1.93)$$

with shallow water Hamiltonian H given by (1.60) and Poisson bracket

$$\{F, H\} = \int_{\Omega} q F_u^\perp \cdot H_u - F_u \cdot \nabla H_h + H_u \cdot \nabla F_h \, dx \quad (1.94)$$

where we use subscripts to express functional derivatives. A Poisson bracket is skew-symmetric, satisfies the Leibniz rule

$$\{f, gh\} = \{f, g\}h + \{f, h\}g \quad (1.95)$$

and the Jacobi identity

$$\{\{f, g\}, h\} + \{\{g, h\}, f\} + \{\{h, f\}, g\} = 0. \quad (1.96)$$

The shallow water equations (1.59) are recovered by substituting point evaluations of h and $h\mathbf{u}$ for the functional F . The Poisson formulation can also be written in the form

$$\dot{F} = \int_{\Omega} F_{\xi}^T \mathbb{J} H_{\xi} \, d\mathbf{x}, \quad (1.97)$$

where $\xi = (\mathbf{u}, h)$ and \mathbb{J} is the non-canonical Poisson operator

$$\mathbb{J} = - \begin{pmatrix} qJ \nabla \\ \nabla^T 0 \end{pmatrix}. \quad (1.98)$$

Here J denotes the canonical 2×2 symplectic matrix $J\mathbf{w} \equiv \mathbf{w}^{\perp}$ for any $\mathbf{w} \in \mathbb{R}^2$ and we think of ∇ as a column operator.

From (1.98), it is possible to find the Casimir invariants of the system, i.e., functionals C satisfying

$$\{F, C\} = 0 \quad (1.99)$$

for every functional F . Expressing the bracket in terms of the Poisson operator defined via (1.97) and (1.98), we see that the Casimirs are precisely the functionals which belong to the kernel of \mathbb{J} , i.e.,

$$\mathbb{J} C_{\xi} = 0. \quad (1.100)$$

When \mathbb{J} is invertible, the Casimirs are just the constant functionals with respect to ξ . For the non-canonical Poisson brackets of fluid dynamics, however, \mathbb{J} typically has a non-trivial kernel, i.e., non-trivial Casimirs. Here, for the rotating shallow water equations, it can be shown that there is a class of Casimirs of the form

$$C = \int_{\Omega} h \gamma(q) \, d\mathbf{x}, \quad (1.101)$$

where γ is an arbitrary function of potential vorticity, not to be confused with the ageostrophic vorticity. Indeed, by the chain rule,

$$C_u = h \gamma'(q) q_u = -\nabla^{\perp} \gamma'(q), \quad (1.102a)$$

$$C_h = \gamma(q) - q \gamma'(q), \quad (1.102b)$$

so that, using (1.98),

$$\mathbb{J} C_{\xi} = \begin{pmatrix} -q \nabla \gamma'(q) - \nabla \gamma(q) + \nabla(q \gamma'(q)) \\ \nabla \cdot \nabla^{\perp} \gamma'(q) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (1.103)$$

Hence, the Casimir condition (1.101) is satisfied. The conservation law for mass corresponds to $\gamma = 1$, while $\gamma(q) = q$ yields

$$\frac{d}{dt} \int_{\Omega} h q \, dx = \frac{d}{dt} \int_{\Omega} \zeta_a \, dx = 0, \quad (1.104)$$

where the first equality is due to (1.61). By Stokes' theorem, (1.104) corresponds to the usual conservation of circulation, which implies conservation of potential vorticity (1.61). More generally, the choice $\gamma(q) = q^n$ corresponds to the conservation of the n th-order enstrophy.

We remark that the Poisson formulation above can be cast into a mixed Eulerian–Lagrangian form that can be seen as a precursor to numerical particle or particle–mesh schemes (Bokhove and Oliver 2006). Further, the rotating shallow water equations in ζ - δ - q variables, see (1.65), can also be written in non-canonical Hamiltonian form as follows:

$$\frac{\partial}{\partial t} \begin{pmatrix} \zeta \\ \delta \\ h \end{pmatrix} = \mathbb{J} \begin{pmatrix} H_{\zeta} \\ H_{\delta} \\ H_h \end{pmatrix}, \quad (1.105)$$

where

$$\mathbb{J} = \begin{pmatrix} -J_q & L_q & 0 \\ -L_q & -J_q & -\Delta \\ 0 & \Delta & 0 \end{pmatrix} \quad (1.106)$$

is a non-canonical Poisson operator whose component operators are defined by $J_q f \equiv J(q, f)$ and $L_q f \equiv \nabla \cdot (q \nabla f)$ for a generic function f ; for details, see Blender and Badin (2017).

1.3.4 Nambu Formulation

Even if the Hamiltonian formulation of the equations of motion for fluid flows is useful to address questions regarding the geometry of the system, including the study of continuous symmetries and associated conservation laws through Noether's theorem, a systematic method for the derivation of the non-canonical Poisson brackets is still generally lacking and the derivation often relies on guesswork. An alternative formulation of dynamics was proposed by Nambu (1973), who suggested an extension of Hamiltonian dynamics which is based on Liouville's theorem and, differently from classical Hamiltonian mechanics, makes use of several conserved quantities, i.e., the Casimirs of the system, which can be considered as additional Hamiltonians and define manifolds whose intersection determines the trajectory in state space. In analogy to Poisson brackets in Hamiltonian mechanics, the resulting dynamics is determined by the Nambu bracket. A Nambu bracket is an n -linear map acting on smooth functions on a manifold that is completely antisymmetric, satisfies the Leibniz rule

$$\{f_1, \dots, f_{n-1}, gh\} = \{f_1, \dots, f_{n-1}, g\}h + \{f_1, \dots, f_{n-1}, h\}g \quad (1.107)$$

and the Jacobi identity

$$\begin{aligned} & \{\{f_1, \dots, f_{n-1}, g_1\}, g_2, \dots, g_n\} + \{g_1, \{f_1, \dots, f_{n-1}, g_2\}, g_3, \dots, g_n\} + \dots \\ & + \{g_1, \dots, g_{n-1}, \{f_1, \dots, f_{n-1}, g_n\}\} = \{f_1, \dots, f_{n-1}, \{g_1, \dots, g_n\}\}. \end{aligned} \quad (1.108)$$

A general theory of Nambu–Poisson structures was outlined by Takhtajan (1994).

For the rotating shallow water equations, it is convenient to pass from (\mathbf{u}, h) to (ζ, δ, h) . Using the chain rule for functional derivatives,

$$F_{\mathbf{u}} = -\nabla^\perp F_\zeta - \nabla F_\delta \quad (1.109)$$

while F_h remains unchanged. With (1.109), the Poisson bracket for the rotating shallow water equations (1.94) takes the form of the sum of two Poisson brackets and a Nambu bracket (Salmon 2005, 2007), i.e.,

$$\dot{F} = \{F, H\}_{\delta\delta} + \{F, H\}_{\zeta\zeta} + \{F, H\}_{\zeta\delta h} \quad (1.110)$$

with

$$\{F, H\}_{\delta\delta} = \int_{\Omega} q \nabla^\perp F_\delta \cdot \nabla H_\delta \, \mathbf{d}\mathbf{x}, \quad (1.111a)$$

$$\{F, H\}_{\zeta\zeta} = \int_{\Omega} q \nabla^\perp F_\zeta \cdot \nabla H_\zeta \, \mathbf{d}\mathbf{x}, \quad (1.111b)$$

and

$$\{F, H\}_{\zeta\delta h} = \int_{\Omega} q (\nabla F_\delta \cdot \nabla H_\zeta - \nabla H_\delta \cdot \nabla F_\zeta) + \nabla F_\delta \cdot \nabla H_h - \nabla H_\delta \cdot \nabla F_h \, \mathbf{d}\mathbf{x}. \quad (1.111c)$$

The subscripts indicate the variables entering the functional derivatives in the brackets. Each of these brackets is antisymmetric and has Casimir functionals given by the kernel of its associated Poisson operator, i.e., satisfy (1.101). Once again, of particular interest are the n moments

$$Z_n = \frac{1}{2+n} \int_{\Omega} h q^{n+2} \, \mathbf{d}\mathbf{x} = \frac{1}{2+n} \int_{\Omega} \frac{\zeta_a^{n+2}}{h^{n+1}} \, \mathbf{d}\mathbf{x}, \quad (1.112)$$

where $n = 0$ yields the usual enstrophy. Using (1.112), the Poisson brackets in (1.110) can be rewritten as

$$\begin{aligned} \{F, H\}_{\delta\delta} &= \{F, H, Z_n\}_{\delta\delta\zeta} \\ &= \frac{1}{3+2n} \int_{\Omega} \frac{1}{q^n} \left[J(F_{\delta}, H_{\delta})(Z_n)_{\zeta} + \text{cyc}(F, H, Z_n) \right] \mathbf{d}\mathbf{x}, \end{aligned} \quad (1.113a)$$

$$\begin{aligned} \{F, H\}_{\zeta\zeta} &= \{F, H, Z_n\}_{\zeta\zeta\zeta} \\ &= \frac{1}{3+2n} \int_{\Omega} \frac{1}{q^n} \left[J(F_{\zeta}, H_{\zeta})(Z_n)_{\zeta} + \text{cyc}(F, H, Z_n) \right] \mathbf{d}\mathbf{x}, \end{aligned} \quad (1.113b)$$

and

$$\begin{aligned} \{F, H\}_{\zeta\delta h} &= \{F, H, Z_n\}_{\zeta\delta h} \\ &= -\frac{1}{1+n} \int_{\Omega} \frac{1}{q^n} \left(\frac{\partial_x F_{\delta} \partial_x H_{\zeta} - \partial_x H_{\delta} \partial_x F_{\zeta}}{\partial_x q} \partial_x (Z_n)_h + \text{cyc}(F, H, Z_n) \right. \\ &\quad \left. + \frac{\partial_y F_{\delta} \partial_y H_{\zeta} - \partial_y H_{\delta} \partial_y F_{\zeta}}{\partial_y q} \partial_y (Z_n)_h + \text{cyc}(F, H, Z_n) \right) \mathbf{d}\mathbf{x}, \end{aligned} \quad (1.113c)$$

where ‘‘cyc’’ denotes expressions derived from the previous term by all cyclic permutations of the indicated symbols.

The sum of the three brackets comprises the complete Nambu bracket for the shallow water equations. This formulation has the advantage that it is easily discretized. The resulting numerical schemes have excellent conservation properties (Salmon 2005). Note that in general, each Poisson bracket corresponds to an infinite number of distinct Nambu brackets, depending on the set of Casimirs which are used (Chatterjee 1996). This non-uniqueness of the Nambu formulation implies the existence of different conservative numerical schemes for the same set of equations, allowing substantial flexibility for the problem under consideration, for example, for the elimination of the singularities emerging from the terms $(\partial_x q)^{-1}$ and $(\partial_y q)^{-1}$ in the third bracket through a rewriting of the brackets before discretization.

For the Nambu brackets for other geophysical fluid equations, including Rayleigh–Bénard convection, see Blender and Badin (2015).

1.4 Dissipation, Turbulence, and Nonlinear Waves

1.4.1 Viscosity and Dissipation

So far, we have looked at a variety of models for geophysical flow without dissipation. However, consistent modeling of viscous dissipation and boundary friction is essential for ensuring proper distribution of energy across scales. Including frictional forces is a non-trivial problem, in particular when looking at effective models for large-scale motion and for numerical simulation where physical dissipation ranges are typically far smaller than what can be numerically resolved.

At the microscopic level, the molecular kinematic viscosity for geophysical fluid flow is modeled as a Newtonian viscosity. Likewise, buoyancy is diffusive with a harmonic diffusion operator, as indicated in our initial formulation of the Boussinesq equations (1.1). While the viscosity coefficients are extremely small, their presence is important on a mathematical level due to the regularizing or smoothing effect. Moreover, small bulk viscosities and frictional layers (Ekman layers), though thin, influence the large-scale flow even in the limit of vanishing viscosity, which gives good reasons to take viscous effects into account even for laminar flow. Examples are the fast rotation and weak vertical diffusion limit system with drag term discussed in Grenier and Masmoudi (1997) and Chemin et al. (2006), and the uniformly enhanced Lagrangian drift from an oscillatory layer (Julien and Knobloch 2007). An early overview on the linear stability of stratified flow with viscosity can be found in Emanuel (1979, 1984). Notably, various stratified flows exhibit intermittency, that is, coexistence of laminar and turbulent regions, also in the geophysical context; see, e.g., Brethouwer et al. (2012) and Ansgore and Mellado (2016).

Crucially, enstrophy or energy cascades in geostrophic or fully developed three-dimensional turbulence require an energy sink at the small scales. This is true even though, in geostrophic turbulence, the dominant transfer of energy is toward larger scales; for a detailed discussion, see, e.g., Danilov et al. (2019). Thus, energetically consistent models require dissipation at small scales, usually referred to as “turbulent viscosity” or “eddy viscosity.” In addition, dissipation is also implicitly or explicitly required to ensure the stability of numerical schemes.

The simplest *ad hoc* approach adds dissipation to the horizontal and vertical velocities with different viscosity coefficients (Pedlosky 1987). This results in the following form of the viscous rotating Boussinesq equations

$$D_t \mathbf{u}_h + (2\boldsymbol{\Omega} \times \mathbf{u})_h = -\frac{1}{\rho_0} \nabla_h p + \nu_h \Delta_h \mathbf{u}_h + \nu_v \partial_{zz} \mathbf{u}_h, \quad (1.114a)$$

$$D_t w + \boldsymbol{\Omega}_h^\perp \cdot \mathbf{u}_h = -\frac{1}{\rho_0} \partial_z p - \frac{g\rho}{\rho_0} + \nu_h \Delta_h w + \nu_v \partial_{zz} w \quad (1.114b)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (1.114c)$$

$$D_t \rho = \kappa \Delta \rho, \quad (1.114d)$$

where ν_h and ν_v are coefficients of horizontal and vertical eddy viscosity, respectively.

In single-layer simplified models such as the barotropic quasi-geostrophic equation (1.51), dissipation comes in the form of bottom drag as the main energy sink and viscosity (or hyperviscosity) as the main enstrophy sink. Under the β -plane approximation $f = f_0 + \beta y$, the dynamics can be written in terms of the relative vorticity $\zeta = \Delta \psi$, so that

$$\partial_t \zeta + \nabla^\perp \psi \cdot \nabla \zeta + \beta \partial_x \psi = F - \lambda \zeta + \nu \Delta \zeta, \quad (1.115)$$

where the terms on the right-hand side are forcing F , linear (Rayleigh) damping to model bottom friction with parameter λ describing an inverse time-scale for the

vorticity decay due to bottom drag, and Newtonian viscosity with parameter ν . In a numerical simulation of a forced system, viscosity is typically replaced by hyperviscosity to remove enstrophy without depleting energy across most resolved scales. To model geostrophic turbulence, both types of dissipation are necessary. A more detailed discussion of this topic is provided in Danilov et al. (2019).

The relevance of the form of viscosity for the mathematical theory is highlighted by the fact that only very recently the well-posedness of the primitive equations with only horizontal viscosity has been established by Cao et al. (2016); also see Charve and Ngo (2011). For full viscosity in all directions much more is known, in particular stability of (nearly) geostrophically balanced dynamics was shown by Temam and Wirosoetisno (2010) and Charve (2006, 2018). Inclusion of at least some kind of dissipation is crucial for these results, but is not necessarily physically consistent nor suitable as a model for energy dissipation on these scales, which is the motivation for the study in Olbers and Eden (2013).

As mentioned, there are no universally accepted criteria for the correct form of dissipation, in particular in connection with numerical schemes. However, a noteworthy prominent approach motivated also by the problem of unfeasible small-scale resolution in practice is “Large Eddy Simulation” (LES) as discussed, e.g., in Sagaut (2006).

In the (next) simplest form, the subgrid closure and parametrization problem of viscosity yield an effective “filtered” Navier–Stokes model for the velocity vector \mathbf{u} in which the turbulent viscosity term reads

$$-\nabla \cdot (v_e(\mathbf{u}) S(\mathbf{u})), \quad (1.116)$$

where $S(\mathbf{u}) = \frac{1}{2}(\nabla \mathbf{u} + (\nabla \mathbf{u})^T)$ is the rate-of-strain tensor and $v_e(\mathbf{u})$ the eddy viscosity. For instance, Smagorinsky (1963) considers $v_e(\mathbf{u}) = C |S(\mathbf{u})|$ with constant C . Variants that are more faithful in preserving the turbulent kinetic energy spectrum can be found in Schaefer-Rolffs and Becker (2013), Schaefer-Rolffs et al. (2015), and Trias et al. (2015).

On a different level, we note that including viscous terms necessitates additional boundary conditions. This in turn yields viscous so-called Ekman boundary layers, which feed back onto the large-scale motion. We also point out that balancing eddy buoyancy fluxes and viscous effects may necessitate additional scaling anisotropy (Grooms et al. 2011).

Concluding this discussion, we emphasize that the presence of dissipation requires to include driving mechanism in order to maintain non-trivial flow. The precise form of the source term is yet another non-trivial modeling issue. Simple idealized configurations are discussed in Danilov et al. (2019); for a discussion in a more realistic context, see, for example, Olbers and Eden (2013).

1.4.2 *Nonlinear Waves and Dynamical Systems Methods*

Wave phenomena at different scales organize the structure of geophysical flows and play a vital role for the transport of energy. From a bifurcation theory viewpoint, linear waves raise expectations for nonlinear wave phenomena. The only nonlinear terms in the standard models as introduced above stem from the material derivative. Geostrophic balance partly turns these into a Poisson bracket nonlinearity, which vanishes for monochromatic plane waves. For this reason, the barotropic quasi-geostrophic equation (1.51) possesses linear Rossby waves with unconstrained amplitude as exact nonlinear solutions and, more generally, special plane waves can sometimes be exact solutions of nonlinear fluid equations (Majda 2003; Julien and Knobloch 2007). However, the barotropic quasi-geostrophic equations are special in this regard. For its parent models, waves are not of plane form but nonlinearly selected, even in the unstratified setting of the rotating shallow water equations.

In a different manner, baroclinic instability relates to bifurcations from Rossby waves via amplitude equations. A discussion of this for a two-layer model can be found in Pedlosky (1987). On the level of the equatorial shallow water wave equations, the amplitude equation approach to nonlinear phenomena has been exploited by Boyd in the 1980s to identify various modulated Rossby and Kelvin waves; cf. Boyd (1980) and the more recent Boyd (2002).

However, it appears that nonlinear waves in geophysical flows received much less attention after the period of seminal progress in the 1980s—with the exception of numerical studies, see below. In the past decade, the subject has regained momentum in particular from a more mathematical viewpoint (Zeitlin et al. 2003; Bouchut et al. 2005; Constantin 2013; Hsu 2014); see Khouider et al. (2013) for a recent review. A major motivation for the study of nonlinear waves is their role in balanced flow, energy transport, and wave breaking in the form of shocks for inviscid models (Zeitlin et al. 2003).

The vast majority of research considers idealized planar or cylindrical geometry and either no viscosity or simple molecular dissipation. The effect of viscosity and Ekman layer formation has been mathematically studied in Dalibard and Saint-Raymond (2010) concerning existence and stability of steady states in a certain scaling limit of fast rotation and small layer thickness. Goh and Wayne (2018) recently studied the role of two-dimensional Oseen vortices in a closely related limit. Rossby–Haurwitz waves are explicit nonlinear spherical waves which received increasing attention recently (Thuburn and Li 2000; Callaghan and Forbes 2006; Ibragimov 2011; Schubert et al. 2009; Smith and Dritschel 2006; Boyd and Zhou 2008). An “intermediate” model accounting for geometric terms from the Mercator projection has been proposed in Bates and Grimshaw (2014) and provides a possible connection between planar and spherical nonlinear wave phenomena.

Another aspect of nonlinear waves is the emergence of shocks in frontogenesis and the adjustment problem. With attention to energy transfer in the geostrophic adjustment problem, this has been studied in Blumen and Wu (1995) and more recently by Reznik (2015), where additional references can be found. In a simplified one-

dimensional setting of the shallow water equations, an essentially explicit approach to fronts has been given in Plougonven and Zeitlin (2005); also see von Storch et al. (2019).

While dynamical systems methods have been originally developed for finite-dimensional problems, dissipative or more generally parabolic partial differential equations often have a finite-dimensional character, which allows for application of tools from generic dynamical systems theory. For instance, compactness and viscosity allow to infer the existence of an invariant and attracting inertial manifold as proven for the spherical Navier–Stokes equations in Temam and Wang (1993).

For domains with one large or unbounded direction, the spatial dynamical systems perspective pushed nonlinear wave theory, in particular regarding stability, for prototypical parabolic equations in the past decades; see, e.g., Kapitula and Promislow (2013), Sandstede (2002), Meyries et al. (2014), and references therein. In the much more prominent inviscid and dispersive case, the presence of conserved quantities gives the mathematical theory a different character, especially concerning stability. We refer to the recent study of Balmforth et al. (2013).

In general, the nonlinear response of a nonlinear parabolic system to linear instabilities can be cast in terms of reduced equations which filter different aspects. On extended domains, the aforementioned modulation equations describe the slow dynamics over large scales near onset; rigorous error estimates have been derived in several contexts (e.g., Schneider 1994; Doelman et al. 2009). For parabolic systems, the classical center manifold reduction provides a conjugacy of the full dynamical system to a decoupled product of trivial linear flow and a lower-dimensional nonlinear part (unique up to exponential error) that is more amenable to analysis, in particular via normal forms (Haragus and Iooss 2011). More selective filtering of bifurcating solutions can be cast in terms of Lyapunov–Schmidt reduction, a standard tool to prove existence of non-trivial bifurcating solutions in amplitude equations or truncated normal forms on a center manifold (Vanderbauwhede 2012). Indeed, the full flow of a reduced equation may be an invalid approximation (over infinite time horizons), but selected solutions may still have exact counterparts in the full system. Notably, bifurcations in the stratified Boussinesq equations have been analyzed by Hsia et al. (2007). Viscous regularization can also act as a filter to reduce the complexity of bifurcations at onset, such as resonance phenomena and allow to determine nonlinear stability for objects other than single shocks (Beck et al. 2010).

The bifurcation theory viewpoint can be exploited numerically by using continuation algorithms to compute branches of nonlinear solutions, their stability and bifurcations and thus, for these solutions, offers an alternative to direct numerical simulation. Dedicated software packages include AUTO,¹ LOCA,² and pde2path.³

In the context of geophysical flows, bifurcations occur naturally in compartment models or coupled ocean–atmosphere systems, e.g., explored numerically in Dijkstra et al. (2014). A review of bifurcation analyses can be found in Simonnet et al. (2009)

¹<http://indy.cs.concordia.ca/auto/>.

²<http://www.cs.sandia.gov/loca/>.

³<http://www.staff.uni-oldenburg.de/hannes.uecker/pde2path/>.

and more broadly for nonlinear waves in fluid turbulence in Kawahara et al. (2012). Co-existing branches of equatorial nonlinear waves have been found numerically by Boyd (2002), Boyd and Zhou (2008). An approach toward nonlinear waves in layered models is to simplify the nonlinear interactions by only retaining resonant triads. A recent study of this type can be found in Bates and Grimshaw (2014). The concepts of bifurcations and nonlinear waves are also insightful for understanding intermittency in wall-bounded flows; see, e.g., Brethouwer et al. (2012) for a setting related to geophysics.

1.5 Stochastic Model Reduction

Recent advances in systematic stochastic climate modeling (Majda et al. 2008; Franzke et al. 2005, 2015; Franzke and Majda 2006; Monahan and Culina 2011) provided new insights into the structural form of the terms accounting for the interaction between resolved and unresolved processes. The approach is based on the adiabatic elimination of fast variables (Kurtz 1973; Papanicolaou 1976; Gardiner 2009) and demonstrates the necessity of nonlinear damping and state-dependent noise; previously, only linear damping and additive noise have been considered. State-dependent noise is known to be responsible for noise-induced drifts (Gardiner 2009), which has the potential to ameliorate some of the known biases. Roberts (2008) and Wang and Roberts (2013) propose a normal form transform to systematically separate slow and fast variables in multi-scale systems. In many situations, it is not *a priori* clear what the slow and fast variables are, so the applicability of this approach for geophysical fluid dynamics needs to be evaluated.

The Mori–Zwanzig formalism (e.g., Zwanzig 2001; Chorin et al. 2000; Wouters and Lucarini 2013; Gottwald et al. 2017) predicts the emergence of memory terms in reduced-order models. Memory terms are rarely considered in current parametrization schemes. Thus, there is an urgent need to develop a systematic framework for the interaction between resolved and unresolved processes and their representation in numerical climate models.

In the context of climate science, stochastic models were first proposed by Hasselmann (1976). A major advance came with the development of the stochastic mode reduction strategy, which is described in detail in Majda et al. (2001, 2008) and is applied, for example, in Majda et al. (2001, 2002, 2003), Franzke et al. (2005), Franzke and Majda (2006), and Franzke et al. (2015). Stochastic mode reduction starts from the equations used in a climate model, with an external forcing, a linear operator and a quadratic nonlinear operator. Splitting the state vector into slow and fast components, assuming scale separation, and replacing the quadratic self-interaction of the fast modes by a stochastic process lead to a stochastic differential equation for the slow variables alone by using the stochastic mode reduction procedure (see next subsection). By doing so, structurally new terms arise such as a deterministic cubic term which acts predominantly as nonlinear damping and both additive and multiplicative noise terms. The multiplicative noise and the cubic term

stem from the nonlinear interaction between the resolved and unresolved modes. Rigorous justification is possible only in the limit of time-scale separation, though in practice the reduced-order models perform well even in the case of moderate or no time-scale separation (Majda et al. 2008; Dolaptchiev et al. 2013; Stinis 2006; Franzke et al. 2005; Franzke and Majda 2006). How to do this in a systematic fashion is the topic of the next subsection.

1.5.1 Basic Setup

To illustrate the basic idea of the stochastic mode reduction strategy, we consider the abstract dynamical system

$$\frac{dz}{dt} = F + Lz + B(z, z), \quad (1.117)$$

where z is the state vector, F denotes the forcing, L a linear and B a quadratic nonlinear operator. For convenience, we assume that F is constant in time; for time-dependent forcing, see Franzke (2013). The linear operator L contains, in particular, advection, the effect of topography, and linear damping. The operator B conserves energy and satisfies the Liouville property, i.e., the dynamical system with only B on its right-hand side is measure preserving (for details, see Franzke et al. 2005). We note that important simplified models for geophysical flow such as the barotropic vorticity equation or the quasi-geostrophic equations can be studied in this framework; see Franzke et al. (2005) for the former and Franzke and Majda (2006) for the latter.

The state vector $z = (\mathbf{x}, \mathbf{y})$ is now split into slow modes \mathbf{x} and fast modes \mathbf{y} . This decomposition is typically done using Empirical Orthogonal Function (EOF) analysis (Franzke et al. 2005; Franzke and Majda 2006) or Principal Interaction Patterns (PIP) (Kwasniok 2004; Crommelin and Majda 2004). These patterns now constitute a complete orthonormal basis and can be used as basis functions in the same way as Fourier modes are used for spectral models (Holton and Hakim 2012). Furthermore, because the leading EOFs typically also decay the slowest (Franzke et al. 2005; Franzke and Majda 2006), it is sensible to use the leading EOFs as the resolved modes. Now we can rewrite (1.117) in terms of slow and fast modes:

$$\frac{d\mathbf{x}}{dt} = F^x + L^{xx}\mathbf{x} + L^{xy}\mathbf{y} + B^{xxx}(\mathbf{x}, \mathbf{x}) + B^{xxy}(\mathbf{x}, \mathbf{y}) + B^{xyy}(\mathbf{y}, \mathbf{y}), \quad (1.118a)$$

$$\frac{d\mathbf{y}}{dt} = F^y + L^{yx}\mathbf{x} + L^{yy}\mathbf{y} + B^{yxx}(\mathbf{x}, \mathbf{x}) + B^{yyx}(\mathbf{x}, \mathbf{y}) + B^{yyy}(\mathbf{y}, \mathbf{y}). \quad (1.118b)$$

We use here the following notation: The first superscript denotes the variable (left-hand side time derivative) the corresponding right-hand side term acts on (e.g., F^x acts on \mathbf{x}), the second and third superscripts denote the variables whose actions and interactions induce a tendency.

In order to carry out the stochastic mode reduction strategy, we introduce a small parameter ε which quantifies the time-scale separation between the slow modes \mathbf{x} and fast modes \mathbf{y} . The parameter ε can also be interpreted as the ratio of the autocorrelation time-scale between the slow and the fast modes. We posit

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}^x + L^{xx}\mathbf{x} + B^{xxx}(\mathbf{x}, \mathbf{x}) + \frac{1}{\varepsilon} (L^{xy}\mathbf{y} + B^{xxy}(\mathbf{x}, \mathbf{y}) + B^{xyy}(\mathbf{y}, \mathbf{y})), \quad (1.119a)$$

$$\frac{d\mathbf{y}}{dt} = \frac{1}{\varepsilon} (\mathbf{F}^y + L^{yx}\mathbf{x} + L^{yy}\mathbf{y} + B^{yxx}(\mathbf{x}, \mathbf{x}) + B^{yxy}(\mathbf{x}, \mathbf{y})) + \frac{1}{\varepsilon^2} B^{yyy}(\mathbf{y}, \mathbf{y}). \quad (1.119b)$$

While the introduction of the time-scale parameter ε in front of some of the tendency terms is somewhat arbitrary and currently mainly based on physical intuition, our current research aims at putting it on a more systematic footing by using multi-scale asymptotics for geophysical flows (Klein 2010; Shaw and Shepherd 2009; Dolaptchiev and Klein 2013).

So far, the equations are fully deterministic. We now make the following *stochastic modeling assumption*: We assume that deterministic nonlinear terms involving \mathbf{x} are mixing with sufficiently fast decay of correlation so that the nonlinear term involving only the fast modes $B^{yyy}(\mathbf{y}, \mathbf{y})$ can be represented by a stochastic term (Majda et al. 1999, 2001), i.e., we approximate

$$\frac{1}{\varepsilon^2} B^{yyy}(\mathbf{y}, \mathbf{y}) dt \approx \text{Stochastic Process}. \quad (1.120)$$

The intuition is that this term is effectively delta-correlated on the slow time-scale. It is illustrative to assume that the stochastic process has the form of an Ornstein–Uhlenbeck (OU) process

$$\frac{1}{\varepsilon^2} B^{yyy}(\mathbf{y}, \mathbf{y}) dt \approx -\frac{\Gamma}{\varepsilon^2} \mathbf{y} dt + \frac{\sigma}{\varepsilon} d\mathbf{W}, \quad (1.121)$$

where Γ and σ are positive-definite matrices and \mathbf{W} is a vector-valued Wiener process. However, the stochastic process does not need to be explicitly specified as shown by Franzke et al. (2005). Inserting (1.121) into the deterministic model (1.119), we obtain a system of stochastic differential equations

$$d\mathbf{x} = \left(\mathbf{F}^x + L^{xx}\mathbf{x} + B^{xxx}(\mathbf{x}, \mathbf{x}) + \frac{1}{\varepsilon} (L^{xy}\mathbf{y} + B^{xxy}(\mathbf{x}, \mathbf{y}) + B^{xyy}(\mathbf{y}, \mathbf{y})) \right) dt, \quad (1.122a)$$

$$d\mathbf{y} = \left(\frac{1}{\varepsilon} (\mathbf{F}^y + L^{yx}\mathbf{x} + L^{yy}\mathbf{y} + B^{yxx}(\mathbf{x}, \mathbf{x}) + B^{yxy}(\mathbf{x}, \mathbf{y})) - \frac{\Gamma}{\varepsilon^2} \mathbf{y} \right) dt + \frac{\sigma}{\varepsilon} d\mathbf{W}. \quad (1.122b)$$

We interpret the SDE in (1.122b) in the sense of Itô. In general, there are two main ways of evaluating the stochastic integrals featuring in the integral form of an SDE, Itô and Stratonovich. From the mathematical perspective, both interpretations are different but interchangeable as an Itô integral can always be rewritten in terms of a Stratonovich integral and vice versa. In modeling, however, the SDE usually arises from a more complicated process or dynamical system in a certain asymptotic limit. In this case, an Itô differential equation represents the situation where the limit dynamics has vanishing autocorrelation while a Stratonovich interpretation represents the situation when the limit dynamics retains finite autocorrelation; see, e.g., the discussion in Moon and Wettlaufer (2014). In our situation, the modeling assumption already entails the idea that terms which cause finite autocorrelations are kept in the deterministic part of (1.122b) so that the noise term only represents the asymptotically uncorrelated parts. Thus, the noise term in (1.122b) should be interpreted in the sense of Itô. We remark that from the technical perspective, an Itô process has the Martingale property which greatly simplifies working with stochastic time integrals. However, functions of Itô processes cannot be differentiated using the normal chain rule as an extra term arises. The stochastic chain rule is called Itô's Lemma and gives rise to the parabolic term in the Fokker–Planck and Kolmogorov backward equations which we will encounter in the next section; for a background on these concepts, see, e.g., Gardiner (2009).

1.5.2 *Slow Dynamics via the Kolmogorov Backward Equation*

In the following, we seek an effective equation for the slow modes of the system (1.22) in the limit $\varepsilon \rightarrow 0$. The method uses adiabatic elimination of fast variables (Kurtz 1973; Papanicolaou 1976; Pavliotis and Stuart 2008) and has been pioneered in the present setting by Majda et al. (1999, 2001). The strategy is the following. Every Itô stochastic differential equation, in particular our system (1.22), has an associated Kolmogorov backward equation (KBE). (For deterministic systems, ODE or PDE, this equation also exists and is usually referred to as the Liouville equation.) It is a parabolic PDE that describes the backward-in-time evolution of the probability distribution function of the system for a given hit probability. While it is rarely possible to solve KBEs numerically in more than a small number of dimensions, the KBE can be useful as an intermediate step in the derivation of a model that is again practically computable. Here, we will subject the KBE to classical multiple scales asymptotics with respect to the small parameter ε . The derivation is successful if the leading non-trivial contribution to the asymptotic series takes the form of a Kolmogorov backward equation in the slow variable \mathbf{x} only, which can be rewritten and simulated as an effective Itô stochastic differential equation.

We start by recalling that for an Itô SDE in the abstract form

$$d\mathbf{x}(t) = \mathbf{a}(\mathbf{x}, t) dt + B(\mathbf{x}, t) d\mathbf{W}(t), \quad (1.123)$$

the corresponding Kolmogorov backward equation reads

$$\frac{\partial p(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^N a_i(\mathbf{x}, t) \frac{\partial p(\mathbf{x}, t)}{\partial x_i} - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N D_{ij}(\mathbf{x}, t) \frac{\partial^2 p(\mathbf{x}, t)}{\partial x_i \partial x_j} \quad (1.124)$$

with

$$D_{ij}(\mathbf{x}, t) = \sum_{k=1}^M b_{ik}(\mathbf{x}, t) b_{jk}(\mathbf{x}, t). \quad (1.125)$$

A derivation and general introduction can be found, for example, in the book by Risken (1996). Applying this notion to our system in the specific form (1.22), we obtain

$$-\frac{\partial \rho^\varepsilon}{\partial t} = \frac{1}{\varepsilon^2} \mathcal{L}_1 \rho^\varepsilon + \frac{1}{\varepsilon} \mathcal{L}_2 \rho^\varepsilon + \mathcal{L}_3 \rho^\varepsilon, \quad (1.126)$$

where the operators \mathcal{L}_i are given by

$$\mathcal{L}_1 = \sum_j \left(-\gamma_j y_j \frac{\partial}{\partial y_j} + \frac{1}{2} \sigma_j^2 \frac{\partial^2}{\partial y_j^2} \right), \quad (1.127a)$$

$$\begin{aligned} \mathcal{L}_2 = & \sum_{j,k} \left(L_{jk}^{xy} y_k + \frac{1}{2} \sum_l (2 B_{jkl}^{xxy} x_k y_l + B_{jkl}^{xyy} y_k y_l) \right) \frac{\partial}{\partial x_j} \\ & + \sum_{j,k} \left(L_{jk}^{yx} x_k + L_{jk}^{yy} y_k + \frac{1}{2} \sum_l (B_{jkl}^{xyy} x_k x_l + 2 B_{jkl}^{yyx} y_k x_l) \right) \frac{\partial}{\partial y_j}, \end{aligned} \quad (1.127b)$$

$$\mathcal{L}_3 = \sum_j \left(F_j \sum_k L_{jk}^{xx} x_k + \frac{1}{2} \sum_{kl} B_{jkl}^{xxx} x_k x_l \right) \frac{\partial}{\partial x_k}. \quad (1.127c)$$

The adiabatic elimination of fast variables now proceeds by performing an asymptotic expansion of the KBE (1.126) in powers of ε . We shall sketch only the main steps and refer the reader to the original work by Majda et al. (2001) for full details. To begin with, we expand ρ^ε as a formal power series, writing

$$\rho^\varepsilon = \rho_0 + \varepsilon \rho_1 + \varepsilon^2 \rho_2 + \dots \quad (1.128)$$

Inserting this expansion into (1.126) and selecting terms of equal order in ε , we obtain a sequence of equations, the first three of which read

$$\mathcal{L}_1 \rho_0 = 0, \quad (1.129a)$$

$$\mathcal{L}_1 \rho_1 = -\mathcal{L}_2 \rho_0, \quad (1.129b)$$

$$\mathcal{L}_1 \rho_2 = -\frac{\partial \rho_0}{\partial t} - \mathcal{L}_3 \rho_0 - \mathcal{L}_2 \rho_1. \quad (1.129c)$$

The structure of these equations tells us that we need to find suitable solvability conditions. Equation (1.129a) implies that ρ_0 belongs to the null space of \mathcal{L}_1 , i.e., $\mathbb{P} \rho_0 = \rho_0$, where \mathbb{P} is the expectation with respect to the invariant measure of the OU process and \mathcal{L}_1 is the generator of the OU process. Hence, (1.129a) shows that ρ_0 is independent of the fast variables \mathbf{y} , thus represents an invariant measure for the fast dynamics in the full SDE. It is easy to see that (1.129a) and (1.129b) satisfy the solvability conditions. Now we want to sketch the derivation of dynamic equation for ρ_0 which can be derived from the solvability condition for equation (1.129c).

Taking the expectation of (1.129b), we obtain the following solvability condition

$$\mathbb{P} \mathcal{L}_2 \rho_0 = \mathbb{P} \mathcal{L}_2 \mathbb{P} \rho_0 = 0. \quad (1.130)$$

If this equation were not be satisfied, the fast modes would induce fast effects on the slow modes. This leads to the solution of (1.129b):

$$\rho_1 = -\mathcal{L}_1^{-1} \mathcal{L}_2 \mathbb{P} \rho_0. \quad (1.131)$$

By inserting this expression into (1.129c) and taking the expectation, we get the solvability condition for ρ_2 :

$$-\frac{\partial \rho_0}{\partial t} = \mathbb{P} \mathcal{L}_3 \mathbb{P} \rho_0 - \mathbb{P} \mathcal{L}_2 \mathcal{L}_1^{-1} \mathcal{L}_2 \mathbb{P} \rho_0. \quad (1.132)$$

In the limit that $\varepsilon \rightarrow 0$, ρ^ε converges to ρ_0 , the solution of equation (1.132), see Theorem 4.4 in Majda et al. (2001) and also Kurtz (1973).

The Kolmogorov backward equation (1.132) can now be transformed back into an effective stochastic differential equation which is computable. The resulting Itô SDE has the form

$$d\mathbf{x} = (\mathbf{F} + L\mathbf{x} + B(\mathbf{x}, \mathbf{x}) + M(\mathbf{x}, \mathbf{x}, \mathbf{x})) dt + \sigma_A d\mathbf{W}_A + \sigma_M(\mathbf{x}) d\mathbf{W}_M, \quad (1.133)$$

where full expressions for the different terms can be found in Majda et al. (2001). Equation (1.133) depends only on the slow variables \mathbf{x} and approximates the statistics of the slow variables of the full SDE (1.22). We note that its right-hand side contains structurally new terms such as a cubic nonlinearity which generally acts as a damping term but still allows some unstable nonlinear directions (Majda et al. 2009), as well as additive and multiplicative noise terms denoted by subscripts A and M , respectively. The multiplicative noise term arises from the nonlinear interaction between resolved and unresolved modes while the additive noise is the results of nonlinear interactions between unresolved modes and the linear interaction between resolved and unresolved modes, respectively.

1.5.3 Direct Averaging

It is instructive to illustrate the stochastic mode reduction procedure by working directly on the equations without the use of the KBE for a special case which allows the direct analytic derivation of the effective equations. For this purpose, we use the following nonlinear triad stochastic interaction equations (Majda et al. 1999) by using the method of averaging (Kurtz 1973; Papanicolaou 1976; Monahan and Culina 2011):

$$dx_1(t) = \frac{b_1}{\varepsilon} x_2(t) y(t) dt \quad (1.134a)$$

$$dx_2(t) = \frac{b_2}{\varepsilon} x_1(t) y(t) dt \quad (1.134b)$$

$$dy(t) = \frac{b_3}{\varepsilon} x_1(t) x_2(t) dt - \frac{\gamma}{\varepsilon^2} y(t) dt + \frac{\sigma}{\varepsilon} dW(t) \quad (1.134c)$$

Now we formally solve the equation for y :

$$y(t) = e^{-\frac{\gamma t}{\varepsilon^2}} y + \frac{b_3}{\varepsilon} \int_0^t e^{-\frac{\gamma(t-s)}{\varepsilon^2}} x_1(s) x_2(s) ds + h(t) \quad (1.135)$$

where

$$h(t) = \frac{\sigma}{\varepsilon} \int_0^t e^{-\frac{\gamma(t-s)}{\varepsilon^2}} dW(s) \quad (1.136)$$

We note that (1.135) contains a time integral which can be interpreted as a memory kernel (Gottwald et al. 2017). The existence of a memory kernel is already predicted in the Mori–Zwanzig formalism (Mori 1965; Zwanzig 1973, 2001; Wouters and Lucarini 2013; Chorin et al. 2000).

However, in the case of time-scale separation so that $\varepsilon \rightarrow 0$, the above term becomes Markovian (Majda et al. 2001) by performing integration by parts:

$$y(t) \rightarrow \varepsilon \frac{b_3}{\gamma} x_1(t) x_2(t) + h(t) \quad (1.137)$$

and

$$h(t) \rightarrow \varepsilon \frac{\sigma}{\gamma} dW(t) \quad (1.138)$$

If we now insert equations (1.137) and (1.138) into (1.134a) and (1.134b), we obtain

$$dx_1(t) = \frac{b_1 b_3}{\gamma} x_1(t) x_2^2(t) dt + \frac{b_1 \sigma}{\gamma} x_2(t) \circ dW(t), \quad (1.139a)$$

$$dx_2(t) = \frac{b_2 b_3}{\gamma} x_1^2(t) x_2(t) dt + \frac{b_2 \sigma}{\gamma} x_1(t) \circ dW(t), \quad (1.139b)$$

which must to be interpreted in the Stratonovich sense as indicated by the \circ in front of the Wiener process. The Stratonovich interpretation arises because the integration performed in (1.136) introduces non-negligible autocorrelation. However, we can transform the Stratonovich SDE into Itô form, picking up an additional noise-induced drift term:

$$dx_1(t) = \frac{b_1 b_2 \sigma^2}{2\gamma^2} x_1(t) dt + \frac{b_1 b_3}{\gamma} x_1(t) x_2^2(t) dt + \frac{b_1 \sigma}{\gamma} x_2(t) dW(t), \quad (1.140)$$

$$dx_2(t) = \frac{b_1 b_2 \sigma^2}{2\gamma^2} x_2(t) dt + \frac{b_2 b_3}{\gamma} x_1^2(t) x_2(t) dt + \frac{b_2 \sigma}{\gamma} x_1(t) dW(t). \quad (1.141)$$

This reduced model has cubic nonlinearity and multiplicative noise. Future research will explore the effects of the memory kernel since in the climate system we do not have any obvious time-scale separations.

1.6 Outlook

In this chapter, we aimed at providing a concise overview of some of the basic models and concepts of geophysical fluid dynamics, introducing consistent notation, scaling and non-dimensionalization, and the hierarchical relations between the different models. We also briefly introduced current research themes such as variational balance models in Section 1.3.1 and stochastic mode reduction in Section 1.5. In our future research, we plan to follow these four broad research directions: (i) Specify the limits of validity of asymptotic regimes and the interaction of different scale regimes, (ii) examine how dynamic properties such as stability, bifurcating nonlinear waves, and nonlinear interactions out of linear waves persist across model hierarchies, (iii) include effective dissipation in scaling analysis and the role of eddy dissipation, (iv) study model hierarchies to develop parameterizations for processes not represented at a coarse level. Here we will use the separation of balanced motion from the meso-scales, where our aim is the develop new energy and momentum consistent stochastic parameterizations of the unresolved processes.

Acknowledgements We thank Sergey Danilov for helpful advice on several aspects of this paper, and Federica Gugole and Gözde Özden for careful proof reading of parts of the manuscript. CF was supported by the German Research Foundation (DFG) through the Cluster of Excellence CliSAP (EXC177) at the University of Hamburg and DFG grant FR3515/3-1. GB was partially supported by DFG grants FOR1740 and BA5068/8-1.

References

- Allen, J.S., Holm, D.D., Newberger, P.A.: Toward an extended-geostrophic Euler–Poincaré model for mesoscale oceanographic flow. In: Norbury, J., Roulstone, I. (eds.) *Large-Scale Atmosphere–Ocean Dynamics*, vol. 1, pp. 101–125. Cambridge University Press (2002)
- Ansonge, C., Mellado, J.P.: Analyses of external and global intermittency in the logarithmic layer of Ekman flow. *J. Fluid Mech.* **805**, 611–635 (2016)
- Arnold, V.I., Khesin, B.A.: *Topological Methods in Hydrodynamics*. Springer (1999)
- Babin, A., Mahalov, A., Nicolaenko, B.: Global splitting, integrability and regularity of 3D Euler and Navier–Stokes equations for uniformly rotating fluids. *Eur. J. Mech. B Fluids* **15**(3), 291–300 (1996)
- Babin, A., Mahalov, A., Nicolaenko, B.: Global splitting and regularity of rotating shallow-water equations. *Eur. J. Mech. B Fluids* **16**(5), 725–754 (1997)
- Babin, A., Mahalov, A., Nicolaenko, B.: Fast singular oscillating limits of stably-stratified 3D Euler and Navier–Stokes equations and ageostrophic wave fronts. In: Norbury, J., Roulstone, I. (eds.) *Large-Scale Atmosphere–Ocean Dynamics*, vol. 1, pp. 126–201. Cambridge University Press (2002)
- Badin, G.: On the role of non-uniform stratification and short-wave instabilities in three-layer quasi-geostrophic turbulence. *Phys. Fluids* **26**(9), 096603 (2014)
- Badin, G., Crisciani, F.: *Variational Formulation of Fluid and Geophysical Fluid Dynamics—Mechanics, Symmetries and Conservation Laws*. Springer (2018)
- Balmforth, N.J., Morrison, P.J., Thiffeault, J.-L.: Pattern formation in Hamiltonian systems with continuous spectra; a normal-form single-wave model (2013). [arXiv:1303.0065](https://arxiv.org/abs/1303.0065)
- Bates, M.L., Grimshaw, R.H.J.: An extended equatorial plane: linear spectrum and resonant triads. *Geophys. Astrophys. Fluid Dyn.* **108**(1), 1–19 (2014)
- Beck, M., Sandstede, B., Zumbrun, K.: Nonlinear stability of time-periodic viscous shocks. *Arch. Ration. Mech. Anal.* **196**(3), 1011–1076 (2010)
- Becker, E.: Frictional heating in global climate models. *Mon. Weather Rev.* **131**, 508–520 (2003)
- Benamou, J.D., Brenier, Y.: Weak existence for the semigeostrophic equations formulated as a coupled Monge–Ampère/transport problem. *SIAM J. Appl. Math.* **58**(5), 1450–1461 (1998)
- Blender, R., Badin, G.: Hydrodynamic Nambu brackets derived by geometric constraints. *J. Phys. A: Math. Theor.* **48**(10), 105501 (2015)
- Blender, R., Badin, G.: Construction of Hamiltonian and Nambu forms for the shallow water equations. *Fluids* **2**(2), 24 (2017)
- Blumen, W., Wu, R.: Geostrophic adjustment: Frontogenesis and energy conversion. *J. Phys. Oceanogr.* **25**(3), 428–438 (1995)
- Bokhove, O., Oliver, M.: Parcel Eulerian–Lagrangian fluid dynamics of rotating geophysical flows. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **462**, 2563–2573 (2006)
- Bokhove, O., Vanneste, J., Warn, J.: A variational formulation for barotropic quasi-geostrophic flows. *Geophys. Astrophys. Fluid Dyn.* **88**(1–4), 67–79 (1998)
- Bouchut, F., Le Sommer, J., Zeitlin, V.: Breaking of balanced and unbalanced equatorial waves. *Chaos* **15**(1), 013503 (2005)
- Boyd, J.P.: Equatorial solitary waves. Part I: Rossby solitons. *J. Phys. Oceanogr.* **10**(11), 1699–1717 (1980)
- Boyd, J.P.: Equatorial solitary waves. Part V: initial value experiments, coexisting branches, and tilted-pair instability. *J. Phys. Oceanogr.* **32**(9), 2589–2602 (2002)
- Boyd, J.P., Zhou, C.: Kelvin waves in the nonlinear shallow water equations on the sphere: nonlinear travelling waves and the corner wave bifurcation. *J. Fluid Mech.* **617**, 187–205 (2008)
- Brethouwer, G., Duguet, Y., Schlatter, P.: Turbulent–laminar coexistence in wall flows with Coriolis, buoyancy or Lorentz forces. *J. Fluid Mech.* **704**, 137–172 (2012)
- Burkhardt, U., Becker, E.: A consistent diffusion–dissipation parameterization in the ECHAM climate model. *Mon. Weather Rev.* **134**, 1194–1204 (2006)

- Çalik, M., Oliver, M., Vasilykevych, S.: Global well-posedness for the generalized large-scale semi-geostrophic equations. *Arch. Ration. Mech. Anal.* **207**(3), 969–990 (2013)
- Callaghan, T.G., Forbes, L.K.: Computing large-amplitude progressive Rossby waves on a sphere. *J. Comput. Phys.* **217**(2), 845–865 (2006)
- Cao, C., Li, J., Titi, E.S.: Global well-posedness of the three-dimensional primitive equations with only horizontal viscosity and diffusion. *Commun. Pure Appl. Math.* **69**(8), 1492–1531 (2016)
- Chan, I.H., Shepherd, T.G.: Balance model for equatorial long waves. *J. Fluid Mech.* **725**, 55–90 (2013)
- Charve, F.: Asymptotics and vortex patches for the quasigeostrophic approximation. *J. Math. Pures Appl.* **85**(4), 493–539 (2006)
- Charve, F.: Asymptotics and lower bound for the lifespan of solutions to the primitive equations. *Acta Appl. Math.* 1–37 (2018)
- Charve, F., Ngo, V.-S.: Global existence for the primitive equations with small anisotropic viscosity. *Rev. Mat. Iberoam.* **27**(1), 1–38 (2011)
- Chatterjee, R.: Dynamical symmetries and Nambu mechanics. *Lett. Math. Phys.* **36**, 117–126 (1996)
- Chemin, J., Desjardins, B., Gallagher, I., Grenier, E.: *Mathematical Geophysics: An Introduction to Rotating Fluids and the Navier–Stokes Equations*. Clarendon Press (2006)
- Cheng, B., Mahalov, A.: Time-averages of fast oscillatory systems. *Discrete Contin. Dyn. Syst. Ser. S* **6**(5), 1151–1162 (2013)
- Chorin, A.J., Hald, O.H., Kupferman, R.: Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Natl. Acad. Sci.* **97**(7), 2968–2973 (2000)
- Constantin, A.: Some three-dimensional nonlinear equatorial flows. *J. Phys. Oceanogr.* **43**(1), 165–175 (2013)
- Crommelin, D., Majda, A.: Strategies for model reduction: comparing different optimal bases. *J. Atmos. Sci.* **61**(17), 2206–2217 (2004)
- Cullen, M.J.P.: A comparison of numerical solutions to the Eady frontogenesis problem. *Q. J. R. Meteorol. Soc.* **134**(637), 2143–2155 (2008)
- Cullen, M.J.P., Purser, R.J.: An extended Lagrangian theory of semi-geostrophic frontogenesis. *J. Atmos. Sci.* **41**(9), 1477–1497 (1984)
- Dalibard, A.-L., Saint-Raymond, L.: Mathematical study of the β -plane model for rotating fluids in a thin layer. *J. Math. Pures Appl.* (9), **94**(2), 131–169 (2010)
- Danilov, S., Juricke, S., Kutsenko, A., Oliver, M.: Toward consistent subgrid momentum closures in ocean models. This volume, Chapter 5 (2019)
- Dijkstra, H.A., Wubs, F.W., Cliffe, A.K., Doedel, E., Dragomirescu, I.F., Eckhardt, B., Gelfgat, A.Y., Hazel, A.L., Lucarini, V., Salinger, A.G., et al.: Numerical bifurcation methods and their application to fluid dynamics: analysis beyond simulation. *Commun. Comput. Phys.* **15**(1), 1–45 (2014)
- Doelman, A., Sandstede, B., Scheel, A., Schneider, G.: The dynamics of modulated wave trains. *Mem. Am. Math. Soc.* **199**(934), viii+105 (2009)
- Dolaptchiev, S.I., Klein, R.: A multiscale model for the planetary and synoptic motions in the atmosphere. *J. Atmos. Sci.* **70**, 2963–2981 (2013)
- Dolaptchiev, S.I., Timofeyev, I., Achatz, U.: Subgrid-scale closure for the inviscid Burgers-Hopf equation. *Commun. Math. Sci.* **11**, 757–777 (2013)
- Dritschel, D.G., Gottwald, G.A., Oliver, M.: Comparison of variational balance models for the rotating shallow water equations. *J. Fluid Mech.* **822**, 689–716 (2017)
- Dritschel, D.G., Viúdez, A.: A balanced approach to modelling rotating stably stratified geophysical flows. *J. Fluid Mech.* **488**, 123–150 (2003)
- Dutrifoy, A., Majda, A.J., Schochet, S.: A simple justification of the singular limit for equatorial shallow-water dynamics. *Commun. Pure Appl. Math.* **62**(3), 322–333 (2009)
- Eliassen, A.: The quasi-static equations of motion with pressure as independent variable. *Geofys. Publ.* **17**, 1–44 (1948)
- Emanuel, K.A.: Inertial instability and mesoscale convective systems. Part I: linear theory of inertial instability in rotating viscous fluids. *J. Atmos. Sci.* **36**(12), 2425–2449 (1979)

- Emanuel, K.A.: Comments on “inertial instability and mesoscale convective systems. Part I: linear theory of inertial instability in rotating viscous fluids”. *J. Atmos. Sci.* **42**(7), 747–752 (1984)
- Embid, P.F., Majda, A.J.: Averaging over fast gravity waves for geophysical flows with arbitrary potential vorticity. *Commun. Partial Differ. Equ.* **21**(3–4), 619–658 (1996)
- Ford, R., McIntyre, M.E., Norton, W.A.: Balance and the slow quasimanifold: some explicit results. *J. Atmos. Sci.* **57**(9), 1236–1254 (2000)
- Franzke, C., Majda, A.J.: Low-order stochastic mode reduction for a prototype atmospheric GCM. *J. Atmos. Sci.* **63**, 457–479 (2006)
- Franzke, C., Majda, A.J., Vanden-Eijnden, E.: Low-order stochastic mode reduction for a realistic barotropic model climate. *J. Atmos. Sci.* **62**, 1722–1745 (2005)
- Franzke, C., O’Kane, T., Berner, J., Williams, P., Lucarini, V.: Stochastic climate theory and modelling. *WIREs Clim. Change* **6**, 63–78 (2015)
- Franzke, C.L.E.: Predictions of critical transitions with non-stationary reduced order models. *Phys. D* **262**, 35–47 (2013)
- Fringer, O.B.: Towards nonhydrostatic ocean modeling with large-eddy simulation. In: Glickson, D. (ed.) *Oceanography in 2025*, pp. 81–83. National Academies Press (2009)
- Gardiner, C.W.: *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer (2009)
- Gerkema, T., Shrira, V.I.: Near-inertial waves in the ocean: beyond the ‘traditional approximation’. *J. Fluid Mech.* **529**, 195–219 (2005a)
- Gerkema, T., Shrira, V.I.: Near-inertial waves on the “nontraditional” β -plane. *J. Geophys. Res. Oceans* **110**(C1) (2005b)
- Gill, A.E.: *Atmosphere-Ocean Dynamics*. Academic Press (1982)
- Givon, D., Kupferman, R., Stuart, A.: Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17**(6), R55 (2004)
- Goh, R., Wayne, C.E.: Vortices in stably-stratified rapidly rotating Boussinesq convection (2018). [arXiv:1802.05369](https://arxiv.org/abs/1802.05369)
- Gottwald, G., Crommelin, D., Franzke, C.: Stochastic climate theory. In: Franzke, C., O’Kane, T. (eds.) *Nonlinear and Stochastic Climate Dynamics*, pp. 209–240. Cambridge University Press (2017)
- Gottwald, G.A., Oliver, M.: Slow dynamics via degenerate variational asymptotics. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **470**(2170), 20140460 (2014)
- Grenier, E., Masmoudi, N.: Ekman layers of rotating fluids, the case of well prepared initial data. *Commun. Partial Differ. Equ.* **22**(5–6), 953–975 (1997)
- Griffies, S.M., Pacanowski, R.C., Schmidt, M., Balaji, V.: Tracer conservation with an explicit free surface method for z -coordinate ocean models. *Mon. Weather Rev.* **129**(5), 1081–1098 (2001)
- Grooms, I., Julien, K., Fox-Kemper, B.: On the interactions between planetary geostrophy and mesoscale eddies. *Dyn. Atmos. Oceans* **51**(3), 109–136 (2011)
- Haragus, M., Iooss, G.: *Local Bifurcations, Center Manifolds, and Normal Forms in Infinite-dimensional Dynamical Systems*. Springer (2011)
- Hasselmann, K.: Stochastic climate models. Part I. Theory. *Tellus* **28**(6), 473–485 (1976)
- Holm, D.D.: Variational principles for stochastic fluid dynamics. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **471**(2176), 20140963, 19 (2015)
- Holm, D.D., Marsden, J.E., Ratiu, T.S.: The Euler-Poincaré equations and semidirect products with applications to continuum theories. *Adv. Math.* **137**(1), 1–81 (1998)
- Holm, D.D., Schmah, T., Stoica, C.: *Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions*. Oxford University Press (2009)
- Holm, D.D., Zeitlin, V.: Hamilton’s principle for quasigeostrophic motion. *Phys. Fluids* **10**(4), 800–806 (1998)
- Holton, J.R., Hakim, G.J.: *An Introduction To Dynamic Meteorology*. Academic Press (2012)
- Hoskins, B.J.: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.* **32**(2), 233–242 (1975)

- Hsia, C.-H., Ma, T., Wang, S.: Stratified rotating Boussinesq equations in geophysical fluid dynamics: dynamic bifurcation and periodic solutions. *J. Math. Phys.* **48**(6), 065602, 20 (2007)
- Hsu, H.-C.: An exact solution for nonlinear internal equatorial waves in the f -plane approximation. *J. Math. Fluid Mech.* **16**(3), 463–471 (2014)
- Ibragimov, R.N.: Nonlinear viscous fluid patterns in a thin rotating spherical domain and applications. *Phys. Fluids* **23**(12), 123102 (2011)
- Julien, K., Knobloch, E.: Reduced models for fluid flows with strong constraints. *J. Math. Phys.* **48**(6) (2007)
- Julien, K., Knobloch, E., Milliff, R., Werne, J.: Generalized quasi-geostrophy for spatially anisotropic rotationally constrained flows. *J. Fluid Mech.* **555**, 233–274 (2006)
- Kamenkovich, V.M., Koshlyakov, M.N., Monin, A.S.: *Synoptic Eddies in the Ocean*. D. Reidel Publishing Company (1986)
- Kapitula, T., Promislow, K.: *Spectral and Dynamical Stability of Nonlinear Waves*. Springer (2013)
- Kasahara, A., Gary, J.M.: Studies of inertio-gravity waves on midlatitude beta-plane without the traditional approximation. *Q. J. R. Meteorol. Soc.* **136**(647), 517–536 (2010)
- Kawahara, G., Uhlmann, M., van Veen, L.: The significance of simple invariant solutions in turbulent flows. *Ann. Rev. Fluid Mech.* **44**(1), 203–225 (2012)
- Khouider, B., Majda, A.J., Stechmann, S.N.: Climate science in the tropics: waves, vortices and PDEs. *Nonlinearity* **26**(1), R1 (2013)
- Klein, R.: Scale-dependent models for atmospheric flows. *Ann. Rev. Fluid Mech.* **42**, 249–274 (2010)
- Klingbeil, K., Burchard, H.: Implementation of a direct nonhydrostatic pressure gradient discretisation into a layered ocean model. *Ocean Model.* **65**, 64–77 (2013)
- Koide, T., Kodama, T.: Navier–Stokes, Gross–Pitaevskii and generalized diffusion equations using the stochastic variational method. *J. Phys. A* **45**(25), 255204, 18 (2012)
- Kurtz, T.G.: A limit theorem for perturbed operator semigroups with applications to random evolutions. *J. Funct. Anal.* **12**(1), 55–67 (1973)
- Kwasniok, F.: Empirical low-order models of barotropic flow. *J. Atmos. Sci.* **61**(2), 235–245 (2004)
- Majda, A.: *Introduction to PDEs and Waves for the Atmosphere and Ocean*. American Mathematical Society (2003)
- Majda, A., Franzke, C., Crommelin, D.: Normal forms for reduced stochastic climate models. *Proc. Natl. Acad. Sci. USA* **106**, 3649–3653 (2009)
- Majda, A., Timofeyev, I., Vanden-Eijnden, E.: A priori tests of a stochastic mode reduction strategy. *Phys. D* **170**, 206–252 (2002)
- Majda, A.J., Franzke, C., Khouider, B.: An applied mathematics perspective on stochastic modelling for climate. *Philos. Trans. R. Soc. A* **366**, 2429–2455 (2008)
- Majda, A.J., Klein, R.: Systematic multiscale models for the tropics. *J. Atmos. Sci.* **60**(2), 393–408 (2003)
- Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA* **96**(26), 14687–14691 (1999)
- Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: A mathematical framework for stochastic climate models. *Commun. Pure Appl. Math.* **54**(8), 891–974 (2001)
- Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.* **60**(14), 1705–1722 (2003)
- Marsden, J.E., Ratiu, T.: *Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems*. Springer (2013)
- McIntyre, M.: Dynamical meteorology—balanced flow. In: Pyle, J., Zhang, F. (eds.) *Encyclopedia of Atmospheric Sciences*, pp. 298–303, 2nd edn. Academic Press, Oxford (2015)
- McIntyre, M.E., Norton, W.A.: Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.* **57**(9), 1214–1235 (2000)
- McWilliams, J.C.: A note on a consistent quasigeostrophic model in a multiply connected domain. *Dyn. Atmos. Oceans* **1**(5), 427–441 (1977)

- Meyries, M., Rademacher, J., Siero, E.: Quasilinear parabolic reaction-diffusion systems: User's guide to well-posedness, spectra and stability of travelling waves. *SIAM J. Appl. Dyn. Sys.* **13**, 249–275 (2014)
- Mohebalhojeh, A.R., Dritschel, D.G.: Hierarchies of balance conditions for the f -plane shallow-water equations. *J. Atmos. Sci.* **58**(16), 2411–2426 (2001)
- Monahan, A.H., Culina, J.: Stochastic averaging of idealized climate models. *J. Clim.* **24**(12), 3068–3088 (2011)
- Moon, W., Wettlaufer, J.S.: On the interpretation of Stratonovich calculus. *New J. Phys.* **16**(5), 055017 (2014)
- Mori, H.: Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.* **33**(3), 423–455 (1965)
- Nambu, Y.: Generalized Hamiltonian dynamics. *Phys. Rev. D* **7**(8), 2405 (1973)
- Olbers, D., Eden, C.: A global model for the diapycnal diffusivity induced by internal gravity waves. *J. Phys. Oceanogr.* **43**(8), 1759–1779 (2013)
- Olbers, D., Willebrand, J., Eden, C.: *Ocean Dynamics*. Springer (2012)
- Oliver, M.: Variational asymptotics for rotating shallow water near geostrophy: a transformational approach. *J. Fluid Mech.* **551**, 197–234 (2006)
- Oliver, M., Vasilyevych, S.: Generalized LSG models with spatially varying Coriolis parameter. *Geophys. Astrophys. Fluid Dyn.* **107**, 259–276 (2013)
- Oliver, M., Vasilyevych, S.: Generalized large-scale semigeostrophic approximations for the f -plane primitive equations. *J. Phys. A: Math. Theor.* **49**, 184001 (2016)
- Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., Weisheimer, A.: Stochastic parametrization and model uncertainty. Technical report, ECMWF (2009)
- Papanicolaou, G.C.: Some probabilistic problems and methods in singular perturbations. *Rocky Mt. J. Math.* **6**(4), 653–674 (1976). Summer Research Conference on Singular Perturbations: Theory and Applications. Northern Arizona University, Flagstaff, Arizona (1975)
- Pavliotis, G.A., Stuart, A.: *Multiscale Methods: Averaging and Homogenization*. Springer (2008)
- Pedlosky, J.: *Geophysical Fluid Dynamics*, 2nd edn. Springer (1987)
- Plougonven, R., Zeitlin, V.: Lagrangian approach to geostrophic adjustment of frontal anomalies in a stratified fluid. *Geophys. Astrophys. Fluid Dyn.* **99**(2), 101–135 (2005)
- Ragone, F., Badin, G.: A study of surface semi-geostrophic turbulence: freely decaying dynamics. *J. Fluid Mech.* **792**, 740–774 (2016)
- Reznik, G.M.: Wave adjustment: general concept and examples. *J. Fluid Mech.* **779**, 514–543 (2015)
- Risken, H.: *The Fokker–Planck Equation*. Springer (1996)
- Roberts, A.J.: Normal form transforms separate slow and fast modes in stochastic dynamical systems. *Phys. D* **387**(1), 12–38 (2008)
- Sagaut, P.: *Large Eddy Simulation for Incompressible Flows: An Introduction*. Springer (2006)
- Saint-Raymond, L.: Lecture notes: Mathematical study of singular perturbation problems. Applications to large-scale oceanography. *Journées Eq. Deriv. Part.* 1–49 (2010)
- Saito, K., Ishida, J., Aranami, K., Hara, T., Segawa, T., Narita, M., Honda, Y.: Nonhydrostatic atmospheric models and operational development at JMA. *J. Meteorol. Soc. Jpn.* **85B**, 271–304 (2007)
- Salmon, R.: The shape of the main thermocline. *J. Phys. Oceanogr.* **12**, 1458–1479 (1982)
- Salmon, R.: Practical use of Hamilton's principle. *J. Fluid Mech.* **132**, 431–444 (1983)
- Salmon, R.: New equations for nearly geostrophic flow. *J. Fluid Mech.* **153**, 461–477 (1985)
- Salmon, R.: Large-scale semigeostrophic equations for use in ocean circulation models. *J. Fluid Mech.* **318**, 85–105 (1996)
- Salmon, R.: *Lectures on Geophysical Fluid Dynamics*. Oxford University Press (1998)
- Salmon, R.: A general method for conserving quantities related to potential vorticity in numerical models. *Nonlinearity* **18**(5), R1 (2005)
- Salmon, R.: A general method for conserving energy and potential enstrophy in shallow water models. *J. Atmos. Sci.* **64**, 515–531 (2007)

- Sandstede, B.: Stability of travelling waves. In: *Handbook of Dynamical Systems*, vol. 2, pp. 983–1055. North-Holland, Amsterdam (2002)
- Schaefer-Rolffs, U., Becker, E.: Horizontal momentum diffusion in GCMs using the dynamic Smagorinsky model. *Mon. Weather Rev.* **141**(3), 887–899 (2013)
- Schaefer-Rolffs, U., Knöpfel, R., Becker, E.: A scale invariance criterion for LES parametrizations. *Meteorol. Z.* **24**(1), 3–13 (2015)
- Schneider, G.: Error estimates for the Ginzburg-Landau approximation. *Z. Angew. Math. Phys.* **45**(3), 433–457 (1994)
- Schubert, W.H., Taft, R.K., Silvers, L.G.: Shallow water quasi-geostrophic theory on the sphere. *J. Adv. Model. Earth Syst.* **1**(2), 2 (2009)
- Shaw, T.A., Shepherd, T.G.: A theoretical framework for energy and momentum consistency in subgrid-scale parameterization for climate models. *J. Atmos. Sci.* **66**, 3095–3114 (2009)
- Shepherd, T.G.: Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. *Adv. Geophys.* **32**, 287–338 (1990)
- Simonnet, E., Dijkstra, H.A., Ghil, M.: Bifurcation analysis of ocean, atmosphere, and climate models. In: Ciarlet, P. (ed.) *Handbook of Numerical Analysis*, vol. 14, pp. 187–229. Elsevier (2009)
- Smagorinsky, J.: General circulation experiments with the primitive equations. *Mon. Weather Rev.* **91**(3), 99–164 (1963)
- Smith, R.K., Dritschel, D.G.: Revisiting the Rossby-Haurwitz wave test case with contour advection. *J. Comput. Phys.* **217**(2), 473–484 (2006)
- Stern, A., Tong, Y., Desbrun, M., Marsden, J.E.: Geometric computational electrodynamics with variational integrators and discrete differential forms. In: Chang, D.E., Holm, D.D., Patrick, G., Ratiu, T. (eds.) *Geometry, Mechanics, and Dynamics: The Legacy of Jerry Marsden*, pp. 437–475. Springer (2015)
- Stewart, A.L., Dellar, P.J.: Multilayer shallow water equations with complete Coriolis force. Part 1. Derivation on a non-traditional beta-plane. *J. Fluid Mech.* **651**, 387 (2010)
- Stewart, A.L., Dellar, P.J.: Multilayer shallow water equations with complete Coriolis force. Part 2. Linear plane waves. *J. Fluid Mech.* **690**, 16–50 (2012)
- Stinis, P.: A comparative study of two stochastic mode reduction methods. *Phys. D* **213**(2), 197–213 (2006)
- Takhtajan, L.: On foundation of the generalized Nambu mechanics. *Commun. Math. Phys.* **160**(2), 295–315 (1994)
- Temam, R., Wang, S.H.: Inertial forms of Navier-Stokes equations on the sphere. *J. Funct. Anal.* **117**(1), 215–242 (1993)
- Temam, R., Wirosoetisno, D.: Slow manifolds and invariant sets of the primitive equations. *J. Atmos. Sci.* **68**(3), 675–682 (2010)
- Theiss, J., Mohebalhojeh, A.R.: The equatorial counterpart of the quasi-geostrophic model. *J. Fluid Mech.* **637**, 327–356 (2009)
- Thuburn, J., Li, Y.: Numerical simulations of Rossby-Haurwitz waves. *Tellus A* **52**(2), 181–189 (2000)
- Tort, M., Dubos, T.: Usual approximations to the equations of atmospheric motion: a variational perspective. *J. Atmos. Sci.* **71**(7), 2452–2466 (2014)
- Tort, M., Dubos, T., Bouchut, F., Zeitlin, V.: Consistent shallow-water equations on the rotating sphere with complete Coriolis force and topography. *J. Fluid Mech.* **748**, 789–821 (2014)
- Tort, M., Ribstein, B., Zeitlin, V.: Symmetric and asymmetric inertial instability of zonal jets on the f -plane with complete Coriolis force. *J. Fluid Mech.* **788**, 274–302 (2016)
- Trias, F.X., Folch, D., Gorobets, A., Oliva, A.: Building proper invariants for eddy-viscosity subgrid-scale models. *Phys. Fluids* **27**(6) (2015)
- Vallis, G.K.: *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. Cambridge University Press (2006)
- Vanderbauwhede, A.: Lyapunov–Schmidt method for dynamical systems. In: *Mathematics of Complexity and Dynamical Systems*, vol. 1–3, pp. 937–952. Springer, New York (2012)

- Vanneste, J.: Balance and spontaneous wave generation in geophysical flows. *Ann. Rev. Fluid Mech.* **45**(1), 147–172 (2013)
- Verkley, W., van der Velde, I.: Balanced dynamics in the tropics. *Q. J. R. Meteorol. Soc.* **136**(646), 41–49 (2010)
- von Storch, J.-S., Badin, G., Oliver, M.: The interior energy pathway: inertial gravity wave emission by oceanic flows. This volume, Chapter 2 (2019)
- Wang, W., Roberts, A.J.: Slow manifold and averaging for slow-fast stochastic differential system. *J. Math. Anal. Appl.* **398**(2), 822–839 (2013)
- Warn, T., Bokhove, O., Shepherd, T., Vallis, G.: Rossby number expansions, slaving principles, and balance dynamics. *Q. J. R. Meteorol. Soc.* **121**(523), 723–739 (1995)
- White, A.A.: A view of the equations of meteorological dynamics and various approximations. In: Norbury, J., Roulstone, I. (eds.) *Large-Scale Atmosphere–ocean Dynamics*, vol. 1, pp. 1–100. Cambridge University Press (2002)
- Whitehead, J.P., Wingate, B.A.: The influence of fast waves and fluctuations on the evolution of the dynamics on the slow manifold. *J. Fluid Mech.* **757**, 155–178 (2014)
- Wouters, J., Lucarini, V.: Multi-level dynamical systems: connecting the Ruelle response theory and the Mori-Zwanzig approach. *J. Stat. Phys.* **151**(5), 850–860 (2013)
- Zeitlin, V., Medvedev, S.B., Plougonven, R.: Frontal geostrophic adjustment, slow manifold and nonlinear wave phenomena in one-dimensional rotating shallow water. Part 1. Theory. *J. Fluid Mech.* **481**, 269–290 (2003)
- Zwanzig, R.: Nonlinear generalized Langevin equations. *J. Stat. Phys.* **9**(3), 215–220 (1973)
- Zwanzig, R.: *Nonequilibrium Statistical Mechanics*. Oxford University Press (2001)

Chapter 2

The Interior Energy Pathway: Inertia-Gravity Wave Emission by Oceanic Flows



Jin-Song von Storch, Gualtiero Badin and Marcel Oliver

Abstract We review the possible role of spontaneous emission and subsequent capture of internal gravity waves (IGWs) for dissipation in oceanic flows under conditions characteristic for the ocean circulation. Dissipation is necessary for the transfer of energy from the essentially balanced large-scale ocean circulation and mesoscale eddy fields down to smaller scales where instabilities and subsequent small-scale turbulence complete the route to dissipation. Spontaneous wave emission by flows is a viable route to dissipation. For quasi-balanced flows, characterized by a small Rossby number, the amplitudes of emitted waves are expected to be small. However, once being emitted into a three-dimensional eddying flow field, waves can undergo refraction and may be “captured.” During wave capture, the wavenumber grows exponentially, ultimately leading to breakup and dissipation. For flows with not too small Rossby number, e.g., for flows in the vicinity of strong fronts, dissipation occurs in a more complex manner. It can occur via spontaneous wave emission and subsequent wave capture, with the amplitudes of waves emitted in frontal systems being expected to be larger than amplitudes of waves emitted by quasi-balanced flows. It can also occur through turbulence and filamentation emerging from frontogenesis. So far, quantitative importance of this energy pathway—crucial for determining correct eddy viscosities in general circulation models—is not known. Toward an answer to this question, we discuss IGWs diagnostics, review spontaneous emission of both quasi-balanced and less-balanced frontal flows, and discuss recent numerical results based on a high-resolution ocean general circulation model.

J.-S. von Storch

Max-Planck-Institut für Meteorologie (MPI-M), Hamburg, Germany

e-mail: jin-song.von.storch@mpimet.mpg.de

G. Badin

Center for Earth System Research and Sustainability (CEN), Universität Hamburg,
Hamburg, Germany

e-mail: gualtiero.badin@uni-hamburg.de

M. Oliver (✉)

Jacobs University, Bremen, Germany

e-mail: m.oliver@jacobs-university.de

© Springer Nature Switzerland AG 2019

C. Eden and A. Iske (eds.), *Energy Transfers in Atmosphere and Ocean*,

Mathematics of Planet Earth 1, https://doi.org/10.1007/978-3-030-05704-6_2

2.1 Introduction

The general circulation of the ocean is forced at large scales by winds and buoyancy fluxes at the sea surface. The directly forced large-scale flows are thought to transfer their energy mainly via baroclinic instabilities to mesoscale eddies that are essentially balanced and account for almost three-fourths of kinetic energy of all oceanic motions when leaving out surface waves (von Storch et al. 2012).¹ While the generation of these eddies via baroclinic and barotropic instabilities of the directly forced large-scale flows is well understood, little is known about the dissipation of these eddies. Within the paradigm of geostrophic turbulence (Charney 1971), mesoscale eddies tend to transfer energy upscale (inverse energy cascade). To ensure an equilibrium state, additional processes must transfer energy from mesoscales to smaller scales so that small-scale turbulence can complete the down-scale cascade of energy to dissipation.

The problem of identifying these energy transfer processes as the ocean's routes to dissipation was brought into focus by Müller et al. (2005). Their paper suggests three possible routes: the inertia-gravity wave route, the instability route, and the boundary route. They argue that none of these routes has received strong theoretical or observational confirmation. However, without a solid understanding of the ocean's route to dissipation, ocean general circulation models have to rely on, to a significant degree, arbitrarily chosen eddy viscosity coefficients. So far, most of the energetic considerations have concentrated on waves generated by external factors, such as the winds and the bottom topography. For example, the high-resolution numerical simulations by Nikurashin et al. (2013) emphasize the importance of gravity waves generated by balanced flow over rough bottom topography, inducing a forward energy cascade en route to dissipation.

Recent numerical studies showed strong evidence that eddying flows are able to spontaneously emit internal gravity waves (see the special issue on 'Spontaneous Imbalance' of the *Journal of the Atmospheric Sciences* from May 2009). Waves, once emitted by an eddying flow, can be refracted by the flow and become captured in certain flow configurations. Wave capture is characterized by an exponential increase in wavenumber and an exponential decrease in the intrinsic group velocity. Furthermore, the wave amplitude grows also exponentially, thereby draining energy from the background flow. These changes make captured waves prone to dissipation and provide a mechanism for an interior route to dissipation directly from quasi-balanced eddying flows themselves, without involving ocean's boundaries and other unbalanced flows. However, it is still open whether this mechanism is significant for dissipation in the real ocean.

¹Note that a global estimate of the ocean kinetic energy, which requires three-dimensional information about flow velocity, is not available from observation and has to be derived using eddy-resolving numerical simulations.

The process of spontaneous imbalance has been researched extensively following the pioneering work of Lighthill (1952) and has received interest in the past decade. For details, we refer the reader to the excellent reviews by Vanneste (2013) for the theory and by Plougonven and Zhang (2014) for observations in the atmosphere. A broad overview can also be gained from the special issue on ‘Spontaneous Imbalance’ of the *Journal of the Atmospheric Sciences* from May 2009.

In this review, we will therefore focus on the role of spontaneous imbalance for the energy pathway to dissipation and on the diagnostics for wave detection.

The Chapter is organized as follows. In Section 2.2, we review the process of spontaneous imbalance. As discussed, the emitted waves have only small amplitudes, but we speculate they might still play a non-negligible role for energy dissipation through their steepening following their capture by the mean flow. In Section 2.3, we thus review the mechanism of wave capture which is crucial if wave emission by quasi-balanced flows with small Rossby number is to play a noticeable role for the total dissipation of energy. In Section 2.4, we review the interaction between the spontaneously emitted IGWs and fronts, not necessarily in the small Rossby number regime, focusing especially on the role of waves in frontal geostrophic adjustment. In Section 2.5, we switch to the diagnostics for the detection of the waves, with particular emphasis especially given to the method of “optimal potential vorticity balance.” In Section 2.6, we discuss the role of high-resolution general circulation model for investigating spontaneous wave emission. We conclude with a summary and discussion in Section 2.7.

2.2 Rotating Shallow Water Equations and Spontaneous Emission

The study of wave emission by vortical flows, also known as spontaneous imbalance or spontaneous wave emission, reaches back to the celebrated paper on aerodynamic sound generation by Lighthill (1952). Ford (1994) and Ford et al. (2000) showed that the same mechanism applies to the generation of IGWs in nearly balanced rotating flow. In particular, Ford (1994) pointed out that the rotating shallow water (RSW) equations play a particular role for studying spontaneous imbalance: in the absence of rotation, these equations are equivalent to the equations for a two-dimensional adiabatic gas. Furthermore, the RSW equations are the simplest equations in which both vortical motions and inertia-gravity waves can exist. In the following, we review the basic properties of the shallow water equations as a simple model for barotropic flow, then proceed to explain the mechanism of spontaneous wave generation.

2.2.1 Shallow Water on the f -Plane

The shallow water equations on the f -plane can be written as

$$\partial_t(h\mathbf{u}) + \nabla \cdot (h\mathbf{u} \otimes \mathbf{u}) + f(h\mathbf{u})^\perp = -\frac{1}{2}g \nabla h^2, \quad (2.1a)$$

$$\partial_t h + \nabla \cdot (h\mathbf{u}) = 0, \quad (2.1b)$$

where $h = h(\mathbf{x}, t)$ denotes the layer depth, $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$ the two-dimensional velocity field, \mathbf{x} the vector of horizontal Cartesian coordinates, $\nabla = (\partial_x, \partial_y)$, f the Coriolis parameter, $\mathbf{u}^\perp = (-u_2, u_1)$, and g is the constant of gravitational acceleration. The system conserves the *total energy*

$$H = \frac{1}{2} \int (h |\mathbf{u}|^2 + g h^2) \, d\mathbf{x}, \quad (2.2)$$

in time and the *potential vorticity*

$$q = \frac{\zeta + f}{h}, \quad (2.3)$$

where $\zeta = \nabla^\perp \cdot \mathbf{u}$ denotes the *relative vorticity* along particle trajectories. The material conservation of potential vorticity implies conservation of the *potential enstrophy*

$$Z = \frac{1}{2} \int h q^2 \, d\mathbf{x}. \quad (2.4)$$

We remark that the shallow water equations can be written in a non-canonical Hamiltonian form, see Franzke et al. (2019).

2.2.2 Spontaneous Emission

Taking divergence and curl of the shallow water momentum equation (2.1a), we obtain, respectively,

$$\partial_t \nabla \cdot (h\mathbf{u}) + \nabla \nabla : (h\mathbf{u} \otimes \mathbf{u}) - f \nabla^\perp \cdot (h\mathbf{u}) = -\frac{1}{2}g \Delta h^2, \quad (2.5a)$$

$$\partial_t \nabla^\perp \cdot (h\mathbf{u}) + \nabla^\perp \nabla : (h\mathbf{u} \otimes \mathbf{u}) + f \nabla \cdot (h\mathbf{u}) = 0, \quad (2.5b)$$

where we use Δ to denote the Laplacian and $\nabla \nabla$ to denote the Hessian operator. Further, “ \otimes ” denotes the tensor product of two vectors and the colon denotes contraction over both indices of two 2-tensors. Taking the time derivative of (2.5a), eliminating the Coriolis term using (2.5b), and replacing $\nabla \cdot (h\mathbf{u})$ by $-\partial_t h$ using the continuity equation (2.1b), we find

$$-\partial_t^3 h + \partial_t \nabla^2 : (\mathbf{h}\mathbf{u} \otimes \mathbf{u}) + f \nabla^\perp \nabla : (\mathbf{h}\mathbf{u} \otimes \mathbf{u}) - f^2 \partial_t h = -\frac{1}{2} g \partial_t \Delta h^2. \quad (2.6)$$

Consider a vortical flow described by (2.6) confined to a bounded region of the spatial domain. We denote the constant layer depth far from the center of the vortical motions by h_0 and assume that h_0 is constant. Then the right-hand side of (2.6) can be rewritten as

$$-\frac{1}{2} g \partial_t \Delta h^2 = -\frac{1}{2} g \partial_t \Delta (h - h_0)^2 - g h_0 \partial_t \Delta h. \quad (2.7)$$

Writing $c_0 = (gh_0)^{1/2}$ to denote the gravity wave phase speed far from the center of vortical motions and rearranging terms, we obtain

$$\left(\frac{\partial^2}{\partial t^2} + f^2 - c_0^2 \Delta \right) \partial_t h = \nabla \nabla : \mathbb{T} \quad (2.8)$$

with

$$\mathbb{T} = \partial_t (\mathbf{h}\mathbf{u} \otimes \mathbf{u}) - f h \frac{\mathbf{u}^\perp \otimes \mathbf{u} + \mathbf{u} \otimes \mathbf{u}^\perp}{2} + \mathbb{I} \frac{g}{2} \partial_t (h - h_0)^2, \quad (2.9)$$

where \mathbb{I} denotes the 2×2 -identity matrix.

Equation (2.8) can be seen as the shallow-water equivalent of the Lighthill equation (Ford 1994; Ford et al. 2000). The left-hand side of (2.8) is a linear inertia-gravity wave operator. When the Froude number $\text{Fr} = U/c_0$, with U denoting the typical velocity in the region of the vortical flow, is small, the right-hand side of (2.8) depends, to good approximation, on the vortical flow alone and can hence be considered as a given source of inertia-gravity waves. In other words, the source term is known as soon as the vortical flow is known. Ford et al. (2000) study the case when potential vorticity anomalies are confined to a finite-size region surrounded by a region in which only linear waves propagate. Using matched asymptotic expansions with the Froude number as small parameter, they link an unsteady vortex pattern in the inner region to emitted propagating waves in the outer region, thereby giving an explicit description of the mechanism of spontaneous emission.

2.2.3 Beyond Shallow Water

Going beyond prototype studies, there is now an extensive body of work using high-resolution numerical simulations on more general flows including surface-intensified fronts (Gall et al. 1988; Danioux et al. 2012), wave emission by baroclinically unstable flows (O'Sullivan and Dunkerton 1995; Zhang 2004; Plougonven and Snyder 2005, 2007; Viúdez 2006; Viúdez and Dritschel 2006; Nadiga 2014), potential vorticity anomalies (Lott et al. 2010, 2012), and dipolar flows (Viúdez 2007; Snyder et al. 2007, 2009; Wang et al. 2009; Wang and Zhang 2010). Of particular interest for the present discussion is the study of O'Sullivan and Dunkerton (1995), illustrated

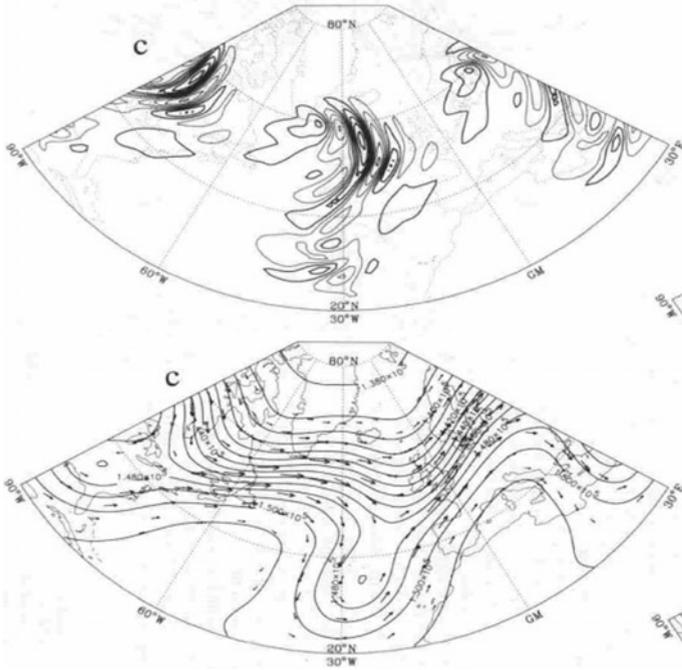


Fig. 2.1 Horizontal velocity divergence (top) and geopotential (bottom) at 130 mb on day 11 of the T126 model simulation reported by O’Sullivan and Dunkerton (1995). Vectors show the horizontal wind. The presence of IGWs is apparent in the velocity divergence field. The quasi-balanced background flow emitting the waves is described by the geopotential contours and wind vectors in the bottom figure. From O’Sullivan and Dunkerton (1995)

in Figure 2.1. The atmospheric flow considered there is comparable to the situation in the Southern Ocean, even though the oceanic baroclinic eddies have much smaller spatial scales than those in the atmosphere. Both the distorted jet considered by O’Sullivan and Dunkerton and the eddying circumpolar currents in the Southern Ocean represent essentially balanced flows.

Most of the studies cited above focus on spontaneous imbalance as a wave generation mechanism and do not directly consider the fate of waves after being emitted. According to Lighthill (1952) and Ford (1994), the waves resulting from spontaneous emission have wavelengths much larger than the spatial scales of the wave source. The waves identified in the references mentioned (e.g., in O’Sullivan and Dunkerton 1995) have, on the contrary, extremely short wavelengths (i.e., wavelengths close to the grid size of the model). Thus, the working hypothesis is that these waves do not represent waves just emitted by the flows, but waves, after having been emitted, are refracted by the flows in which they are embedded, a possibility pointed out by Plougonven and Snyder (2005). This is further discussed in Section 2.3 below.

Theoretical arguments show that spontaneous emission of IGWs is a generic phenomenon (Vanneste and Yavneh 2004, 2007; Vanneste 2008, 2013). However, the strength of IGW emission depends on the magnitude of Rossby number characterizing the flow. IGW emission is strong for flows with large Rossby number. This happens at least (i) in the equatorial ocean where the Coriolis parameter is small and (ii) in the vicinity of strong fronts with large and spatially confined flow velocities that lead to large Rossby numbers.

In this review, we focus on issues pertaining to the generic situation of extratropical flows and refer the reader for the specifics of equatorial dynamics to Le Sommer et al. (2004) and Medvedev and Zeitlin (2005). Regarding spontaneous generation of near-internal gravity waves in the Kuroshio front, Nagai et al. (2015) found that the energy emitted by the waves is larger than the energy produced by lee waves created by the interaction of balanced flows with the bottom topography (Nikurashin and Ferrari 2011). Most of the energy radiated from the frontal region, however, is being reabsorbed, suggesting that wave capture plays an important role in frontogenesis. The situation in frontal regions will be addressed in Section 2.4 further below.

IGW emission is weak when the Rossby number is small. In this case, there is a clear separation between the timescales of the flow and the timescales of the emitted IGWs. The slow, quasi-balanced components can be accurately described up to terms that are exponentially small in the Rossby number. Following the theoretical arguments, waves spontaneously emitted by the flows must be exponentially weak. Even though most atmospheric and oceanic flows are in this regime, wave emission is not necessarily negligible due to subsequent refraction and wave capture affecting all IGWs, however weak.

There is renewed interest in understanding refraction and capture by the background flow in connection with spontaneous wave emission in vortical flows. Plougonven and Snyder (2005) demonstrate the importance of wave capture by showing that some characteristics of the IGWs emitted by tropospheric jets are determined by the process of wave capture rather than by the generation mechanism itself. Wave-trapping in the ocean has been studied by Kunze (1985), Klein and Treguier (1995), Xing and Davies (2004) and, in data analysis, by van Haren (2003) with special emphasis on near-inertial waves. Trapping of sub-inertial waves in anticyclonic regions of the flow has also been studied theoretically for axisymmetric vortices (Kunze and Boss 1998; Llewellyn Smith 1999).

2.3 Ray Equations and Wave Capture

Theoretical studies of the interaction between emitted waves and a background flow are often done using ray equations formulated for a wave packet located at $\mathbf{x} = \mathbf{x}(t)$ traveling with intrinsic group velocity $\mathbf{c}_{g,i}$ in a background flow $\mathbf{u} = \mathbf{u}(\mathbf{x}, t)$,

$$\frac{d_g \mathbf{x}}{dt} \equiv \mathbf{c}_g = \mathbf{c}_{g,i} + \mathbf{u}, \quad (2.10a)$$

$$\frac{d_g \mathbf{k}}{dt} = -(\nabla \mathbf{u}) \cdot \mathbf{k}, \quad (2.10b)$$

where \mathbf{c}_g is the effective group velocity, d_g/dt denotes the time derivative in a frame traveling with the wave packet, $\nabla \mathbf{u}$ is the gradient tensor of the flow \mathbf{u} .

To obtain equation (2.10b), the intrinsic dispersion relation is assumed to be independent of \mathbf{x} and t and the WKB approximation is used. Under these assumptions, the path of the wave train is completely determined by equation (2.10a) and the temporal evolution of the wavenumber vector $\mathbf{k} = (k, l, m)$ along the path is described by (2.10b); the refraction process is, in particular, independent of stratification. The derivation of (2.10) in the spirit of a WKB approximation, and including the next order terms is sketched in Vanneste (2013), who proceeds to explain the mechanism of spontaneous emission. Broutman et al. (2004) review ray tracing in geophysical fluid dynamics. Olbers et al. (2012) give an extensive treatment of wave propagation in a variable environment.²

When the intrinsic group velocity is inversely related to the modulus of the wavenumber vector, as is the case for internal IGWs, a wave packet propagating in a flow field is refracted under suitable circumstances such that the ray equation describing the evolution of the wavenumber takes the same form as the equation describing the evolution of the gradient of a passive tracer which is known to grow exponentially in time. Correspondingly, the relative group velocity vanishes and the wave amplitude increases; the wave packet becomes “captured” or “frozen” into the background flow. We note that in wave capture it is the *horizontal* (and vertical) wavelength which goes to zero. This is different from the formation of a critical layer where vertical shear causes only the *vertical* wavelength to go to zero while leaving the horizontal wavenumber unchanged.

To obtain analytical solutions, Jones (1969) considered a three-dimensional flow with constant gradient tensor $\nabla \mathbf{u}$. More recently, Bühler and McIntyre (2005) and Plougonven and Snyder (2005) studied a horizontally divergent flow $\mathbf{u} = (u, v, 0)$ with a gradient tensor of the form

$$\nabla \mathbf{u} = \begin{pmatrix} u_x & v_x & 0 \\ u_y & v_y & 0 \\ u_z & v_z & 0 \end{pmatrix}. \quad (2.11)$$

For captured waves, the intrinsic group velocity is essentially zero, so that the group velocity \mathbf{c}_g is essentially determined by \mathbf{u} . The direction of \mathbf{c}_g generally differs from the direction of the asymptotic wave vector \mathbf{k} . This is most easily seen in the special case considered by Plougonven and Snyder (2005) where $\mathbf{u} = (u, v, w)$ has the gradient tensor

²By writing the dispersion relation in the form $\omega(\mathbf{k}, t) = \omega_i(\mathbf{x}, t) + \mathbf{u}(\mathbf{x}, t) \cdot \mathbf{k} = \Omega(\mathbf{k}; \mathbf{x}, t)$ and assuming that the intrinsic dispersion relation and hence the intrinsic frequency ω_i is independent of \mathbf{x} , their equation (6.46) reduces to our equation (2.10b).

$$\nabla \mathbf{u} = \begin{pmatrix} -\alpha & 0 & 0 \\ 0 & \alpha & 0 \\ \beta & \gamma & 0 \end{pmatrix} \quad (2.12)$$

so that the flow is hyperbolic in the horizontal plane and the solution to (2.10b) reads

$$\mathbf{k}_h(t) = (k_0 e^{\alpha t}, l_0 e^{-\alpha t}). \quad (2.13)$$

Thus, the wave vector asymptotically aligns with one of the axes (i.e., the heteroclinic orbits) at the hyperbolic points while the captured waves move along the hyperbolic flow. The angle between the ray and the horizontal wavenumber vector changes as the captured wave packet is advected by the hyperbolic flow. Such changes make it more difficult to identify the relation between the captured waves and the flow which refracts the waves.

Realistic flow fields are often locally axial, with the horizontal component \mathbf{u}_h of \mathbf{u} being approximately aligned with the isobars. Choosing a frame in which x is tangential and y is perpendicular to the direction of the flow, a horizontal purely axial flow is described by $\mathbf{u}_h = (u(x, y), 0)$. This flow can have a shear u_y transverse to the flow axis and a stretching rate u_x along the flow axis. An incompressible Boussinesq flow must then have a vertical velocity w with $u_x + w_z = 0$.

Generally, u_x and u_y have complicated spatial structures. In the core of axial flows, however, all flows reveal the common feature that the shear u_y vanishes. With $u_y = 0$, $u_x = a \neq 0$ a constant, and $w_z = -u_x = -a$, the gradient tensor takes the simple form

$$\nabla \mathbf{u} = \begin{pmatrix} u_x & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & w_z \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -a \end{pmatrix}, \quad (2.14)$$

so that the solution to (2.10b) takes the simple form

$$\mathbf{k}(t) = (k_0 e^{-at}, l_0, m_0 e^{at}). \quad (2.15)$$

When the flow is horizontally compressing, i.e., when $a = u_x < 0$, the asymptotic wave vector aligns with the direction of the flow, and also with the direction of propagation of the wave packet, provided that the intrinsic group velocity is either aligned with \mathbf{u} or small compared to \mathbf{u} . As $|\mathbf{k}|$ grows exponentially, the vertical wavenumber decreases exponentially and the vertical component of the intrinsic group velocity grows exponentially. For an ocean limited by its lower and upper boundary, a wave train with a high vertical group velocity reaches the boundary in a short time. Its reflection at the boundary can easily lead to the emergence of vertically standing structures.

The refraction through axial flows can be significant for two reasons. First, axial flows represent an essential ingredient of a realistic atmospheric and oceanic flow field. Second, the properties of waves predicted by (2.15) with a negative a so that the magnitude of the component of \mathbf{k} along the flow is large (or the corresponding

wavelength is small) and \mathbf{k} is aligned with the flow (in other words, aligned with the wave crests are perpendicular to the flow) are qualitatively consistent with what is seen in the atmosphere (Figure 2.1) and in the ocean (Figure 2.9).

2.4 Interactions Between IGWs and Density Fronts

2.4.1 Wave Capture in Frontal Strain

The emission of IGWs is particularly strong during frontogenesis. In this classical problem, the Rossby number $\text{Ro} = U/fL$ and the bulk Richardson number

$$\text{Ri} = \frac{N_0^2}{(\Delta U/\Delta z)^2} = \frac{N_0^2 f^2}{S_0^4}, \quad (2.16)$$

where N_0^2 and S_0^2 are, respectively, the vertical and horizontal buoyancy gradients, are both $O(1)$, implying a dominant role of ageostrophic shear. They also change with time and can thus affect wave emission and subsequent recapture. Surface density fronts, or frontal regions created in the interior by mesoscale dynamics, are characterized by large values of strain, so the mechanism of wave capture, described in the previous section, is expected to be particularly active. To study wave capture in this context, the configurations discussed in there are insufficient as they are used under the assumption that both the time dependence of the background flow and the role of ageostrophic shear have a negligible influence on the dynamics of IGWs. The wave capture theory by Bühler and McIntyre (2005) also assumes that the intrinsic frequency and the polarization relation of the waves are independent of the geostrophic flow, while in reality they are both functions of the Richardson number. Notice, however, that slow changes of N , u , and $\nabla \mathbf{u}$ could be incorporated using WKB theory.

Thomas (2012) suggests a more elaborate flow configuration featuring a time-varying background flow and ageostrophic shear which is set up as follows. Starting from the inviscid adiabatic rotating Boussinesq equations

$$D_t \mathbf{u} + f \mathbf{e}_z \times \mathbf{u} = -\frac{1}{\rho_0} \nabla p + b \mathbf{e}_z, \quad (2.17a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (2.17b)$$

$$D_t b = 0, \quad (2.17c)$$

where b denotes the buoyancy and \mathbf{e}_z the unit vector in the vertical direction, we split $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$, $b = \bar{b} + b'$, and $p = \bar{p} + p'$ into mean flow and perturbation components. Now assume that the mean flow has constant horizontal strain α and linear vertical shear with time-varying coefficients $\hat{\mathbf{u}}(t) = (\hat{u}(t), \hat{v}(t), 0)$, so that

$$\bar{\mathbf{u}}(\mathbf{x}, t) = \begin{pmatrix} \alpha x + \hat{u}(t) z \\ -\alpha y + \hat{v}(t) z \\ 0 \end{pmatrix}, \quad (2.18a)$$

$$\bar{b}(\mathbf{x}, t) = N^2(t) z - S^2(t) y, \quad (2.18b)$$

and

$$\bar{p}(\mathbf{x}, t) = -\rho_0 \left(\frac{1}{2} \alpha^2 (x^2 + y^2) + f \alpha x y - \frac{1}{2} N^2(t) z^2 + S^2(t) y z \right). \quad (2.18c)$$

This mean flow yields the tensor

$$\nabla \bar{\mathbf{u}} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & -\alpha & 0 \\ \hat{u}(t) & \hat{v}(t) & 0 \end{pmatrix}. \quad (2.19)$$

Inserting the mean flow ansatz (2.18) with vanishing perturbation quantities back into (2.17), we obtain a system of ordinary differential equations for the time-dependent coefficients, namely

$$\frac{d}{dt} \begin{pmatrix} \hat{u} \\ \hat{v} \\ S^2 \\ N^2 \end{pmatrix} = \begin{pmatrix} f \hat{v} - \alpha \hat{u} \\ -f \hat{u} + \alpha \hat{v} + S^2 \\ \alpha S^2 \\ \hat{v} S^2 \end{pmatrix} \quad (2.20)$$

with initial conditions $u(0) = S_0^2/f$, $\hat{v}(0) = 2\alpha S_0^2/f^2$, $N^2(0) = N_0^2$, and $S^2(0) = S_0^2$, which correspond to an initial vertical and horizontal buoyancy gradient and to corresponding geostrophic velocities in the presence of strain α . System (2.20) is easily solved for the vertical and horizontal buoyancy gradients,

$$S^2(t) = S_0^2 e^{\alpha t}, \quad (2.21a)$$

$$N^2(t) = N_0^2 + \frac{S_0^4}{f^2} (e^{2\alpha t} - 1), \quad (2.21b)$$

and the mean flow

$$\bar{\mathbf{u}}(\mathbf{x}, t) = \begin{pmatrix} \alpha x + f^{-1} S^2(t) z \\ -\alpha y + 2\alpha f^{-2} S^2(t) z \\ 0 \end{pmatrix}. \quad (2.21c)$$

The amplification in time of the buoyancy gradients is a characteristic feature of frontogenesis.

Assuming that the perturbation is invariant in the x -direction, the equations for the perturbation fields read

$$\bar{D}_t \mathbf{u}' + \mathbf{u}' \cdot \nabla \bar{u} + \mathbf{u}' \cdot \nabla \mathbf{u}' + f \mathbf{e}_z \times \mathbf{u}' = -\frac{1}{\rho_0} \nabla p' + b' \mathbf{e}_z, \quad (2.22a)$$

$$\nabla \cdot \mathbf{u}' = 0, \quad (2.22b)$$

$$\bar{D}_t b' + \mathbf{u}' \cdot \nabla \bar{b} + \mathbf{u}' \cdot \nabla b' = 0, \quad (2.22c)$$

where $\bar{D}_t = \partial_t + \bar{v} \partial_y$. Inserting the plane wave ansatz $\mathbf{u}'(\mathbf{x}, t) = e^{i\mathbf{k} \cdot \mathbf{x}}$ with $\mathbf{k} = (0, l, m)$ into (2.22), we find that the wavenumber vector changes in time according to

$$\mathbf{k}(t) = (0, l_0 e^{\alpha t}, m_0 - S_0^2 f^{-2} (e^{2\alpha t} - 1)). \quad (2.23)$$

The polarization angle (the angle the streamlines make with the horizontal direction)

$$\theta = -\tan^{-1}\left(\frac{l}{m}\right) \quad (2.24)$$

changes with time in the presence of frontogenetic strain. According to (2.23), it asymptotes to

$$\theta_\rho = \tan^{-1}\left(\frac{S^2}{N^2}\right) \quad (2.25)$$

for all initial angles except zero. Further, the perturbation kinetic energy density

$$e = \frac{1}{2} |\mathbf{u}'|^2 \quad (2.26)$$

satisfies the evolution equation

$$\bar{D}_t e = -u'^* w' \partial_z \bar{u} - \frac{1}{\rho_0} \nabla \cdot (\mathbf{u}'^* p') + w'^* b' + \alpha (|v'|^2 - |u'|^2) - v'^* w' \partial_z \bar{v}, \quad (2.27)$$

where the asterisks indicate complex conjugates. The different terms on the right-hand side can be interpreted as follows:

- The first term represents geostrophic shear production;
- the second term represents convergence or divergence of the energy flux;
- the third term is the release of potential energy via the buoyancy flux;
- the fourth term is the shear production by the lateral gradients of the mean flow;
- the last term represents the ageostrophic shear production.

This suggests a route to dissipation of the mesoscale flow different from the scenario associated with wave capture. In wave capture, the wave amplitude grows exponentially in time, leading to wave breaking and dissipation. If frontogenesis is present, multi-scale analysis shows that the wave action is no longer conserved, but decreases in time (Thomas 2012). In this case, the energetics of the waves are modified by the Reynolds stress introduced by the additional shear production term related to the strain-driven ageostrophic flow. Numerical simulations carried out by Thomas (2012) using the explicit expression for the deformation shear (2.19) show

that the evolution of the kinetic energy density of the IGWs is largely controlled by the last two terms in (2.27). In particular, in the presence of frontogenetic strain, i.e., when $\alpha > 0$, a perturbation satisfying $|v'| > |u'|$ acts to mix the momentum of the deformation field and extracts its kinetic energy—acting thus as an effective viscosity. In contrast, the ageostrophic shear production term is negative for all θ , resulting in a damping of the IGWs, except when the initial angle is $90^\circ < \theta_0 < 180^\circ$, for which the streamlines are tilted against the ageostrophic shear and the IGWs exhibit a transient growth that is long enough to affect the energetics.

As a note of caution, it should be noted that the results here reported are derived for a basic state that is perturbed with disturbances that are invariant in the x -direction. Studying the nonlinear stability with respect to symmetric perturbations using the Lyapunov method, Cho et al. (1993) found that the growth of the instabilities is amplified by a factor

$$\frac{\text{Ri}^{1/2} + 1}{\text{Ri}^{1/2} - 1}, \quad (2.28)$$

where Ri is the Richardson number. This result shows that significant amplification is present only when $\text{Ri} \approx 1$. Further analysis is thus required for more general, non-symmetric, perturbations.

2.4.2 Role of IGWs in Frontal Geostrophic Adjustment

Spontaneous imbalance is linked to, but is not the same as, the process of geostrophic adjustment (also known as Rossby adjustment; see Rossby 1938). Geostrophic adjustment describes the evolution of a perturbed state into a final, smooth state with the same potential vorticity through emission of waves. This mechanism is particularly active during frontogenesis. Thus, to fully understand the energetics of frontogenesis, a careful study of the relation between capture and adjustment is necessary.

The linear theory of geostrophic adjustment was initiated by Obukhov (1949) and has matured since; see Blumen (1972) for a review on the early results. The nonlinear theory, however, is much less known. Nonlinear effects during geostrophic adjustment include secondary effects such as trapping of IGWs (the production of waves that are unable to leave the frontal region) and wave breaking which modify the characteristics of the flow, in particular the PV, and thus feedback into the process of geostrophic adjustment. Zeitlin (2010) posed a series of largely open questions regarding the process of adjustment: Does a final, adjusted state, exist? Is it stable? What is the effect of the IGWs on the energy pathway of the system? Do emitted and trapped waves co-exist? And do trapped waves interact with the front?

The classic theory of frontogenesis relies on the semi-geostrophic approximation (Hoskins and Bretherton 1972), which in turn relies on balanced dynamics, thereby removing IGWs altogether. The turbulence emerging in the process of semi-geostrophic frontogenesis was studied numerically by Ragone and Badin (2016)

who found that ageostrophic advection is responsible for the growth of filamentary structures with skewed vorticity. Filamentation itself can induce a forward cascade of potential temperature/buoyancy variance and can thus be a possible route to dissipation, even if this might still be inhibited by the fact that semi-geostrophic flows are balanced. In this sense, the energy pathway in frontal regions can be much more complex than the pathway through wave capture alone.

The situation for flow far from balance is very different. Some progress has been made by analyzing the cross-frontal rotating shallow-water (RSW) equations (Zeitlin et al. 2003) and confirmed numerically by Bouchut et al. (2004). Consider the RSW equations on the f -plane with no dependence on the along-front direction y , written here in components:

$$\partial_t u + u \partial_x u - f v = -g \partial_x h, \quad (2.29a)$$

$$\partial_t v + u \partial_x v + f u = 0, \quad (2.29b)$$

$$\partial_t h + \partial_x (uh) = 0. \quad (2.29c)$$

This system can also be interpreted as the one-dimensional RSW equations with the addition of a transverse flow.

In addition to the potential vorticity $q = (\partial_x v + f)/h$, the system possesses a second Lagrangian invariant, the *geostrophic momentum* $M = v + f x$. Introducing a Lagrangian x -coordinate via $\dot{X} = u(X, t)$, we write the two momentum equations in system (2.29) as

$$\ddot{X} - f v = -g \partial_x h, \quad (2.30a)$$

$$\dot{v} + f \dot{X} = 0. \quad (2.30b)$$

The second equation can be integrated immediately, giving

$$v(a, t) + f X(a, t) = M(a), \quad (2.31)$$

where we write a to denote a Lagrangian label coordinate. The geostrophic momentum $M(a)$ is determined by the initial conditions: given the initial velocity v_0 , one has $M(a) = v_0(a) + f a$. Note that the adjusted state are no longer slowly evolving, but exactly stationary in Lagrangian coordinates. Further, in Lagrangian coordinates the problem reduces to a semi-geostrophic problem in which geostrophy is retained in (2.30a), but where (2.30b) is purely ageostrophic. The continuity equation (2.29c) in Lagrangian coordinates takes the form

$$h(X(a, t), t) = \frac{h_0(a)}{X'(a, t)}, \quad (2.32)$$

where X' denotes the derivative with respect to the label a . By the chain rule,

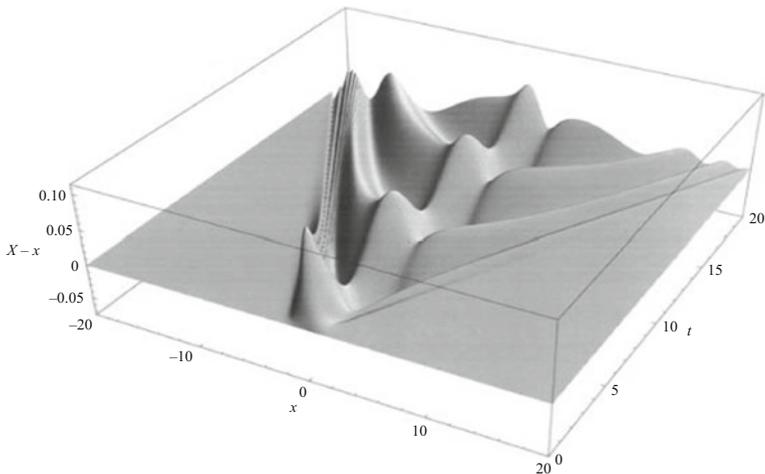


Fig. 2.2 Solution of (2.34) with initial conditions $h_0(x) = 1 + \exp(-x^2)$, $v_0(x) = -2(x + 0.2 \sin(x)) \exp(-x^2)$, $u_0(x) = 0.1 \exp(-x^2)$. The horizontal axes show the horizontal distance x and time. Surfaces show the time evolution of the particle displacement $\phi = X - x$. The solution shows emission of fast gravity waves and slowly dispersing near-inertial oscillations. A shock is visible on the left side of the front (negative values of x). From Zeitlin et al. (2003)

$$\partial_x h = \frac{h'_0}{(X')^2} - h_0 \frac{X''}{(X')^3}. \quad (2.33)$$

Writing $\phi(a, t) = X(a, t) - a$ to denote the displacement of a particle from its initial position and inserting (2.31) and (2.33) into (2.30a), we obtain a closed equation for ϕ ,

$$\ddot{\phi} + f^2 \phi + \frac{g h'_0}{(1 + \phi')^2} + \frac{g h_0}{2} \left[\frac{1}{(1 + \phi')^2} \right]' = f v_0, \quad (2.34)$$

which can be solved with appropriate initial conditions. Assuming positive potential vorticity as a condition for solvability and taking appropriate frontal initial conditions, Zeitlin et al. (2003) solved (2.34), thereby showing that to leading order in the cross-frontal Rossby number, the fast and slow dynamics split so that there are no fast oscillations trapped in the frontal region. An example of a solution to (2.34) is shown in Figure 2.2. The front is located at $x = 0$, situated in between a double jet configuration set by the initial conditions $h_0(x) = 1 + \exp(-x^2)$, $v_0(x) = -2(x + 0.2 \sin(x)) \exp(-x^2)$, and $u_0(x) = 0.1 \exp(-x^2)$. The horizontal axes show the horizontal distance x and time. Surfaces show the time evolution of the particle displacement ϕ . An initially small departure from balance evolves emitting fast IGWs as well as slowly dispersing near-inertial oscillations. Evolving toward an adjusted state which is a minimizer of the energy, the system sheds energy also via the emergence of shocks. One of these shocks is visible on the left of the figure, where the relative vorticity takes negative values.

This scenario was extended to two-layer stratification by Le Sommer et al. (2014) and to stratified rectilinear fronts by Plougonven and Zeitlin (2005). The presence of stratification allows not only for the process of baroclinic instabilities, but also for symmetric inertial instabilities. Plougonven and Zeitlin (2005) demonstrate that nonexistence of a smooth adjusted state in vertically bounded domains is generic, with the system evolving toward a singularity in finite time, even in the absence of a background confluent flow. In this setting, there are trapped modes with horizontal scales comparable to the width of the jet. The existence of shocks and the breaking trapped modes can represent a possible pathway toward dissipation.

The analysis reported here becomes more complicated when passing from the 1D RSW to the 2D RSW equations (Reznik et al. 2001) where near-inertial oscillations stay coupled to the slow vertical component for a long time and act to retard the process of adjustment with respect to the standard scenario of fast dispersion of inertia-gravity waves.

Further complication might be introduced by the presence of a background confluent flow, which might be responsible for time dependent Rossby and Richardson numbers, as discussed in Section 2.4.1. In turn, this might affect the transition toward a possibly adjusted state. A study of an initially smooth time evolving confluent flow would be also interesting for the comparison with the behavior of balance models in the same configuration, such as the surface quasi-geostrophic model, for which proof of the formation of a singularity in a finite time is linked to the formation of singularities in the 3D Euler equation and for which the problem is still open (Constantin et al. 1994).

2.5 Diagnostics

Crucial to any investigation of IGW activity is the need to consistently and efficiently split the velocity and mass fields into a balanced flow component and the residual, usually interpreted as the inertia-gravity wave field.

Flows characterized by a small Rossby number evolve on timescales much longer than the timescales of IGWs emitted by the flows. In this situation, the emitted waves can reasonably well be diagnosed using a linear time filter based on a Fourier decomposition. Linear filters are straightforward and easy to apply. Figure 2.3 shows the geographical distribution of the super-inertial kinetic energy, obtained by applying such a filter. However, it is not clear how small the Rossby number should be for ensuring adequate wave detection and whether this procedure leads to a consistent separation over the global domain. A review on time filters can be found in von Storch and Zwiers (1999).

An alternative characterization is provided by balance relations which can be derived by systematic asymptotic expansion in various ways; see, e.g., the review by McIntyre (2015). One possibility is the elimination of the vertical velocity associated with balanced motion through the Omega equation to identify the residual signal of the waves (Danioux et al. 2012). There is a variety of more sophisticated

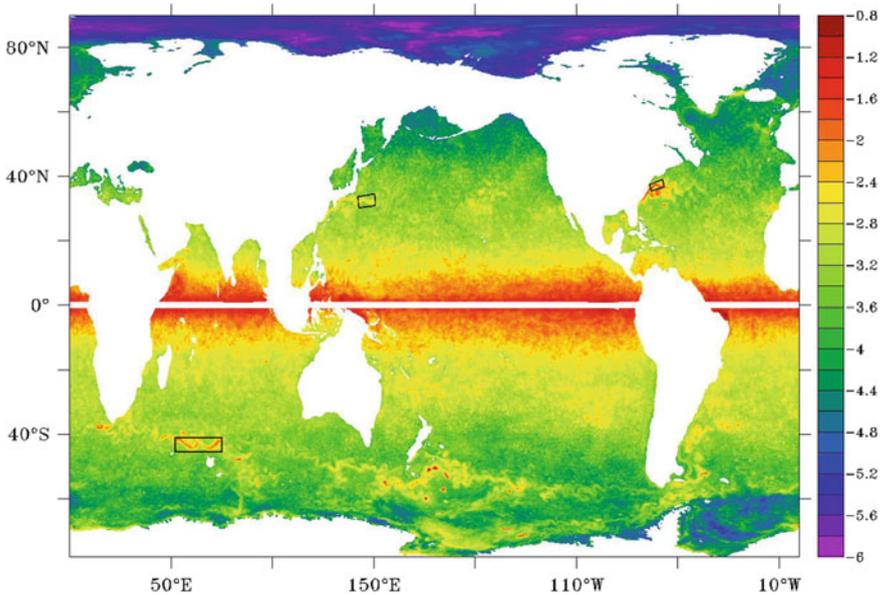


Fig. 2.3 Square root of super-inertial kinetic energy, $\sqrt{k_{\text{sup}}}$, at 291 m in a 0.1° simulation performed with the MPIOM OGCM. The super-inertial kinetic energy is obtained by Fourier decomposing the hourly zonal and meridional velocity in the frequency domain and calculating the mean variance of the zonal and meridional velocity fluctuations at frequencies larger than the local inertial frequency. The color scale is in units of m/s. Numbers given refer to $\log_{10} \sqrt{k_{\text{sup}}}$

balance relations such as the second-order correction to quasi-geostrophic theory (Muraki et al. 1999), the semi-geostrophic (Hoskins and Bretherton 1972; Hoskins 1975), surface semi-geostrophic (Badin 2013; Ragone and Badin 2016), δ - γ balance (Mohebalhojeh and Dritschel 2001), the L_1 -model and variants (Salmon 1985, 1996; McIntyre and Roulstone 2002; Oliver and Vasylyevych 2016), and many others; see, e.g., Barth et al. (1990) for a long list of intermediate models and a comparison of their performance in geostrophic turbulence and Allen et al. (2002) for a numerical comparison in a nonturbulent regime. The general advantage of balance relations is that they offer an *a priori* way of separating motion slaved to PV from other types of motion, at least to some order of approximation in the limit of small Rossby number. The drawback is that they require a careful scale analysis and that they are, by construction, restricted to a specific limited regime of validity. Nonetheless, they offer ways to “diagnose the diagnostics” as the only first-principles analysis available.

A third alternative is the equation-free “optimal potential vorticity balance” scheme pioneered by Viúdez and Dritschel (2004). Their work is based on the notion that geophysical flows are determined to a large extent by potential vorticity, so that the balanced component of the fields are those which are slaved to potential vorticity. Their procedure can be implemented on top of a purely numerical model by slowly increasing the influence of nonlinear interactions from a trivial initial state where the

splitting into balanced and imbalanced modes is explicit to the fully nonlinear final state. So long as the energy in the imbalanced degrees of freedom remains adiabatically invariant, the final state will be approximately balanced whenever the initial state is. Optimal balance has some highly desirable features, in particular, carrying no explicit assumptions on scales, only weak dependence on numerical parameters such as the time horizon of integration, and the possibility to define balance having nothing but an actual numerical code. Moreover, in simple settings, it can be proved that the procedure yields near-optimal results in the asymptotic limit of large scale separation between balanced and imbalanced motion (Cotter 2013; Gottwald et al. 2017). However, implementation and pre-asymptotic behavior in realistic configurations remains largely unexplored.

No matter whether a linear or a dynamical filter is used, we recommend to first assess the flow field by using, e.g., the Rossby number (Section 2.5.1). Some pitfalls encountered when using linear filters is briefly discussed in Section 2.5.2. The equation-free “optimal potential vorticity balance” scheme, which we hope to be applicable for a large range of dynamical regimes, is discussed in Sections 2.5.3 and 2.5.4.

2.5.1 *Characterization of Flow Regimes via the Rossby Number*

Any splitting of motions into balanced and imbalanced components, whether by linear filtering or more sophisticated techniques, relies on at least some degree of scale separation. For geophysical flow, the most important scale separation parameter is the Rossby number, informally described as the ratio of the inertial force to the Coriolis force. Small Rossby number flows are dominated by planetary rotation and can be thought of as well balanced. When the Rossby number is well below one, the timescales of the mean flow are clearly separated from IGWs. In this case, IGWs can be identified using simple Fourier filters. For larger Rossby numbers, balance relations or dynamical filters are required to separate IGWs from the full flow field. We emphasize, however, that for Rossby numbers close to or larger than unity, the concept of balance breaks down altogether and no method will lead to an unambiguous and physically meaningful splitting; see, e.g., the discussion in McIntyre (2009).

The textbook definition of the Rossby number $Ro = U/(fL)$, where U denotes the typical magnitude of the velocity field, L the typical horizontal length scale, and f the Coriolis parameter, does not work well for studying spontaneous emission. This is because flows capable of emitting waves are generally highly anisotropic and vary with time so that it is difficult to determine U and L unambiguously. This problem can be overcome by considering the Rossby number

$$Ro_{\text{Lagr}} = \frac{\pi}{\tau f} \tag{2.35}$$

where τ denotes a typical Lagrangian timescale: the time it takes a particle to travel through a half-wavelength of the dominant pattern (e.g., McIntyre 2009). The particle travel time, however, is often not easily available in a simulation, so that a more practical proxy is the local Rossby number R defined by the ratio of relative vorticity to the local Coriolis frequency f ,

$$R = \frac{|\nabla_h^\perp \cdot \mathbf{u}_h|}{|f|}. \quad (2.36)$$

Averaging can be used to eliminate possible influence of short-term fluctuations on R . Alternatively, R can be defined using the geostrophic velocity in (2.36).

A realistic simulation of the world ocean performed with the Max-Planck-Institute Ocean Model (MPIOM) at a horizontal resolution of 0.1° shows that the maximum values of the local Rossby number R derived from the monthly mean of the velocity field are about 0.15–0.25 in the Southern Ocean, about 0.3–0.4 in the Gulf Stream, and about 0.1–0.25 in the Kuroshio Extension. Extremes of about 0.5 are found in isolated spots in the Gulf Stream. R decreases quickly with increasing depth: the same numerical simulations show that at about 300 m depth, maximum values of R are reduced to about 0.1–0.15 in the Southern Ocean, about 0.2–0.3 in the Gulf Stream, and about 0.1–0.13 in the Kuroshio region. The flows outside the tropical oceans are, apart from very few exceptions, characterized by a Rossby number smaller than 0.2–0.4. The smallness of R in the extratropical regions suggests that the flows there are not far from balance and that their timescales are noticeably longer than those of IGWs.

2.5.2 Linear Filters

When the timescales of the mean flow and the IGWs are well separated, IGWs can be identified by high-pass filtering the data. There is a variety of different filters which can be used for this purpose. In the following, we shall explain the options and potential pitfalls for accurate detection of IGWs generated by spontaneous imbalance.

The flow field in which spontaneous emission occurs contains at least two components: the mean flow which emits the waves and varies predominantly at sub-inertial frequencies, and the emitted inertia-gravity waves themselves which oscillate at super-inertial frequencies. Generally, the sub-inertial variations are much stronger than the super-inertial waves. Not all filters ensure that the strong sub-inertial variations are not leaked to the super-inertial frequency range and misinterpreted as inertia-gravity waves. This can happen, for example, when using a running average filter (Figure 2.4).

The running average filter is the simplest digital filter. Denoting the original velocity time series by x_t , the $2K + 1$ -point running average \bar{x}_t is given by

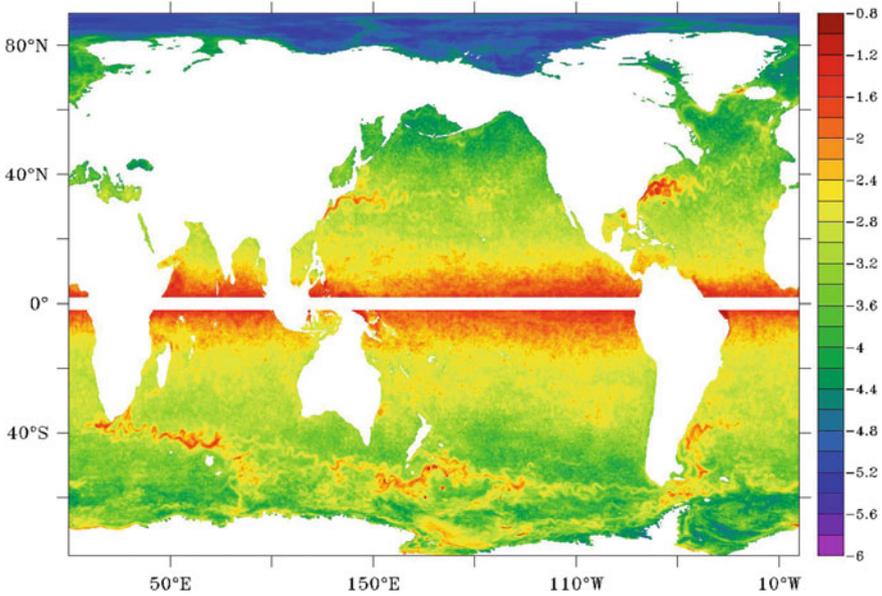


Fig. 2.4 Square root of super-inertial kinetic energy, $\sqrt{k_{\text{sup}}}$, at 291 m in the same 0.1° simulation considered in Figure 2.3. The super-inertial kinetic energy is obtained by filtering out the sub-inertial variance using a running average filter with the filter window identical to the local inertial period. The color scale is in units of m/s. Numbers given refer to $\log_{10} \sqrt{k_{\text{sup}}}$

$$\bar{x}_t = \sum_{k=-K}^{k=K} a_k x_{t+k} \quad (2.37)$$

with equal weights

$$a_k = \begin{cases} \frac{1}{2K+1} & \text{for } |k| \leq K, \\ 0 & \text{for } |k| > K. \end{cases} \quad (2.38)$$

The spectral density function of the filtered time series \bar{S} is related to the spectral density function of the original time series S by

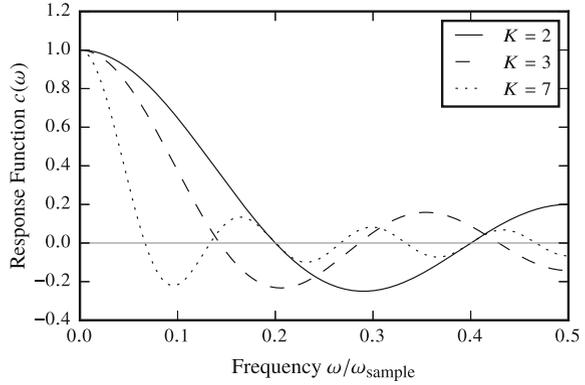
$$\bar{S}(\omega) = |c(\omega)|^2 S(\omega) \quad (2.39)$$

with frequency response function

$$c(\omega) = \frac{\sin((2K+1)\pi\omega)}{(2K+1)\sin(\pi\omega)}. \quad (2.40)$$

Figure 2.5 shows $c(\omega)$ for a running average filter with $K = 2, 3,$ and 7 . The magnitude of $c(\omega)$ is close to one at low frequencies and becomes small and oscillatory

Fig. 2.5 Response functions $c(\omega)$ of three running average low-pass filters with different filter windows. The frequency axis is labeled in units of the sample frequency ω_{sample} and shows frequencies up to the Nyquist frequency $\omega_{\text{sample}}/2$



at high frequencies, i.e., the filter acts as a low-pass filter. Further, $c(0) = 1$ and $c(\omega) = 0$ at frequencies $\omega = j/(2K + 1)$ for $j = 1, \dots, K$, i.e., the filter leaves the zero-frequency component unchanged and completely filters out frequencies that are multiples of $1/(2K + 1)$. Thus, one would typically take the local inertial period as the window length, so that variations at frequencies equal to an integer multiple of the local inertial frequency are filtered out. In the residual time series

$$x'_t = x_t - \bar{x}_t, \quad (2.41)$$

variations at these high frequencies are fully retained and variations at the zero frequency are completely removed. At all other frequencies, $c(\omega)$ is neither exactly one nor exactly zero, making the running average filter inaccurate. In particular, at frequencies close to zero, $|c(\omega)| < 1$, so that parts of the sub-inertial variations are retained in x'_t . When the sub-inertial variations are strong, this contamination can be significant.

Better results can be obtained by filtering directly in the frequency domain, where the discrete Fourier transform is used to convert the time series from time to frequency domain and back. This approach is generally accurate provided the time series x_t is periodic, which is generally not the case for a finite time series of a system with a strong slowly varying component which imposes a trend onto the finite-time window. In this case, the Fourier decomposition is inaccurate in the sense that the spectral density at a frequency ω of the finite time series does not represent the spectral density at the same frequency obtained from an infinite time series. This type of bias can be corrected by de-trending and tapering (giving the beginning and the end of a time series less weight) before taking the discrete Fourier transform (see, e.g., von Storch and Zwiers 1999).

The difference between the two filters can be seen by comparing Figure 2.3 with Figure 2.4. In each case, the filter is tuned to the local inertial period τ_1 . Figure 2.3 uses a spectral filter with response function

$$c(\omega) = \begin{cases} 0 & \text{for } \omega < 1/\tau_1, \\ 1 & \text{for } \omega \geq 1/\tau_1. \end{cases} \quad (2.42)$$

In addition, the time series was de-trended and tapered using the split-cosine-bell taper (von Storch and Zwiers 1999) where the non-unit weights make up about 12% of an hourly time series over a period of one month. Figure 2.4 is based on the same data, but with a running average filter of filter length τ_1 and subsequent computation of the residual (2.41). The two different filters lead to different results. In particular, the super-inertial kinetic energy in the Southern Ocean is much weaker in Figure 2.3 than in Figure 2.4. The stronger super-inertial kinetic energy seen with the running average filter is a result of the contamination of the residual by the imperfect filtering of strong low-frequency components. Thus, one should not use a running average filter for the purpose of separating IGWs from eddying flows.

2.5.3 Optimal Potential Vorticity Balance

In this section, we review the concept of optimal potential vorticity balance as introduced by Viúdez and Dritschel (2004). In the following Section 2.5.4, we then present a mathematical framework in which optimal balance can be understood systematically. It is our belief that the framework applies in much more general circumstances and can be used for high-accuracy diagnostics and balanced initialization.

The dynamics of large-scale geophysical flow is, to a surprising degree, determined by the evolution of a single scalar field, the potential vorticity q , alone (McIntyre and Norton 2000). Indeed, in many cases the full equations can be written, equivalently, as a material transport equation for q augmented by additional equations for the evolution of the “ageostrophic components” \mathbf{P} . For example, for the rotating Boussinesq model, the components of \mathbf{P} are the horizontal divergence δ and the ageostrophic vorticity γ , see Section 2.2 in Franzke et al. (2019). For the rotating shallow-water equations, Section 2.2.1 shows how to split off a second order in time equation using $\partial_t h$ as the imbalance variable.

In all of these cases, leading order balance is obtained by simply setting $\mathbf{P} = 0$, but an optimally balanced flow will have small, but nonzero values for these “ageostrophic” variables. Balance is then viewed as a mapping $q \mapsto \mathbf{P} = \Phi(q)$. The function Φ can be seen as defining a manifold in phase space parameterized by potential vorticity. It is generally understood that a unique invariant slow manifold does not exist except in trivial cases, but that it is possible to construct balance manifolds which are almost invariant over very long intervals of time (Vanneste 2013). The numerical task can now be described as follows. Given any potential vorticity field q^* , numerically compute one image \mathbf{P}^* of the balance map in an optimal way, where possible optimality considerations are discussed in Section 2.5.4 below.

We further require that for a state where $\mathbf{P} = 0$ there exists a stationary “rest state potential vorticity” q_r . On the sphere or on the beta-plane, q_r will only depend

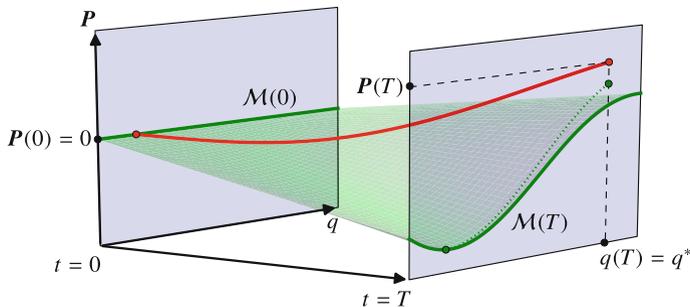


Fig. 2.6 Sketch of the ramping procedure in extended phase space

on latitude. With η denoting the flow map, let us write $\mathbf{x} = \eta(\mathbf{a}, t)$ to denote the Eulerian position of the fluid particle initially at Lagrangian label coordinate \mathbf{a} . Material conservation of potential vorticity is then expressed as

$$q(\eta(\mathbf{a}, t), t) = q_0(\mathbf{a}). \quad (2.43)$$

The central idea of optimal PV balance is to make the initial potential vorticity q_0 slowly dependent on time by writing

$$q_0(\mathbf{a}, t) = q_r(\mathbf{a}) + \rho(t/T) q_a(\mathbf{a}) \quad (2.44)$$

where q_a is a Lagrangian description of the potential vorticity anomaly, ρ is a smooth, monotonically increasing *ramp function* with $\rho(0) = 0$ and $\rho(1) = 1$, and T is the *ramp time* which we think of as comparable to the timescale of the slow, balanced motion. The introduction of an explicit dependence of q_0 on t changes the equations of motion, but a system in balance may remain approximately balanced so long as the ramp is sufficiently slow, i.e., the ramp time T is sufficiently long. In fact, the scale separation between the fast timescale and the ramp time is a small parameter playing the same role as the scale separation between fast and slow timescales in the asymptotic analysis of the problem (Gottwald et al. 2017).

This leaves us with the following task: solve the modified equations of motion with the temporal boundary conditions $\mathbf{P}(0) = 0$ and $q(T) = q^*$, then define $p^* = \mathbf{P}(T)$. Figure 2.6 shows a sketch of the situation. For every fixed t , the balance manifold is illustrated as a one-dimensional curve; the manifold itself is time-dependent. At the start of the ramp when $t = 0$, the balance manifold is exact and trivial; we know how to explicitly balance the system. At the final time $t = T$, the balance manifold is only approximate and nontrivial. A trajectory from $t = 0$ to $t = T$ which reaches the imposed potential vorticity q^* at the end of the ramp, shown in red in Figure 2.6, remains approximately balanced so long as the change of the balance manifold in time is adiabatic. The quality of balance for the final state will be affected by the choice of ramp function, the choice of the ramp time T , and the scale separation between

balanced and unbalanced motion. We will comment on the conjectured interplay of these parameters in the context of a simple model for balance in Section 2.5.4 below.

To use optimal potential vorticity balance in practice, two further problems need to be overcome. First, what is ramped in (2.44) is the initial, i.e., Lagrangian, potential vorticity q_0 . The dependence of the Eulerian potential vorticity on q_0 , however, is only straightforward in a semi-Lagrangian setting such as the CASL method by Dritschel and co-workers (Dritschel et al. 1999) or one of its variants. Second, the formulation of the problem as stated here is a boundary value problem in time. As such, a direct implementation requires specialized code. Viúdez and Dritschel (2004), however, have proposed an iterative scheme that can be performed with only an evolutionary code: Start at $t = T$ with $q(T) = q^*$ and $\mathbf{P}(T) = 0$ and integrate *backward* to $t = 0$. At $t = 0$, the solution will be generally imbalanced, so force the solution back to balance by resetting $\mathbf{P}(0) = 0$, keeping $q(0)$ unchanged. Now integrate *forward* to $t = T$. At $t = T$, the solution will still be approximately balanced, but in general $q(T) \neq q^*$. Now reset $q(T) = q^*$, keeping $\mathbf{P}(T)$ unchanged. Repeat until the iteration converges. In that case, the fixed point of the iteration is, by construction, a solution to the temporal boundary value problem as outlined earlier. Viúdez and Dritschel (2004) report fast convergence in their numerical experiments.

2.5.4 A Simple Model for Optimal Balance

A simple finite dimensional model for balance is the Hamilton system

$$\dot{q} = p, \quad (2.45a)$$

$$\varepsilon \dot{p} = Jp - \nabla V(q), \quad (2.45b)$$

where $q(t)$ and $p(t)$ are vectors of even dimension, $V(q)$ is a convex potential, and

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \quad (2.46)$$

is the canonical symplectic matrix. The corresponding ramped system, in this case, reads

$$\dot{q} = p, \quad (2.47a)$$

$$\varepsilon \dot{p} = Jp - \rho(t/T) \nabla V(q). \quad (2.47b)$$

We solve (2.47) as a boundary value problem where we impose that at $t = 0$, the system is on the trivial slow manifold and at $t = T$ the position variables take some prescribed value q^* . Thus, (2.47) is augmented with boundary conditions

$$p(0) = 0 \quad \text{and} \quad q(T) = q^*. \quad (2.48)$$

This system can be completely and rigorously analyzed by Hamiltonian normal form theory (Cotter and Reich 2006; Cotter 2013). In particular, it can be shown that the drift of the ramped trajectory away from the family of approximate balance manifolds $\mathcal{M}(t)$ is smaller than any algebraic order in ε for fixed ramp time T . When the ramp function ρ is analytic with a uniform lower bound on the radius of analyticity along the entire ramp, the drift is even exponentially small. Similarly, it can be shown that for fixed ε , the drift off the manifold decreases as T becomes large.

However, there is a second source of error besides the drift. The asymptotic series defining the family of manifolds $\mathcal{M}(t)$ for the ramped system differs from the asymptotic series for the trivial balance manifold at $t = 0$ and, similarly, it differs from the asymptotic series for the balance relation of the original system (2.45) by terms containing derivatives of ρ . More precisely, $O(\varepsilon^{n+1})$ -accuracy of the procedure requires the first n derivatives of ρ to vanish at the temporal end points. This excludes uniformly analytic ramp functions. Yet, it is still possible to get exponential bounds, albeit with a smaller power of ε in the exponent; see Gottwald et al. (2017).

The behavior of the optimal balance procedure can be demonstrated most easily in the linear case. In this case, for a frozen value of the ramping parameter, the slow manifold is an exact invariant and can be computed by diagonalization of the full system matrix. Figure 2.7 shows the quality of balance achieved by the method for the simple potential $V(q) = \frac{1}{2}|q|^2$ and different ramp functions of the type

$$\rho(\theta) = \frac{f(\theta)}{f(\theta) + f(1 - \theta)} \quad (2.49)$$

for $f(\theta) = \theta^2$ (“quadratic”), $f(\theta) = \theta^4$ (“quartic”), and $f(\theta) = \exp(-1/\theta)$ (“exponential”). The balance error is defined as the distance of $p(T)$ as computed via solving the boundary value problem of the ramped system to the slow eigenspace of the full system (2.45) computed by exact diagonalization.

The behavior for large ramp times (corresponding to good adiabatic invariance) is dominated by the error committed at the start and end point of the ramp. It shows

Fig. 2.7 Numerical test for a linear model problem. Shown is the balancing error as a function of the ramp time for ramp functions with different orders of vanishing at the end points of the ramp

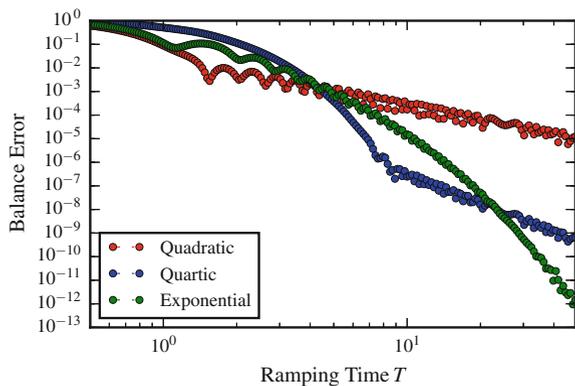
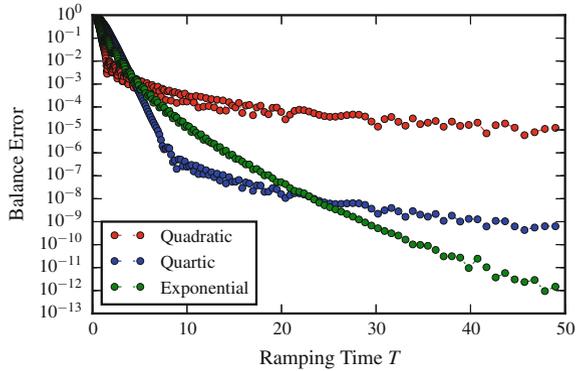


Fig. 2.8 The same data as shown in Figure 2.7 on a semi-log scale. The exponential regime of the error behavior can be seen as a straight line



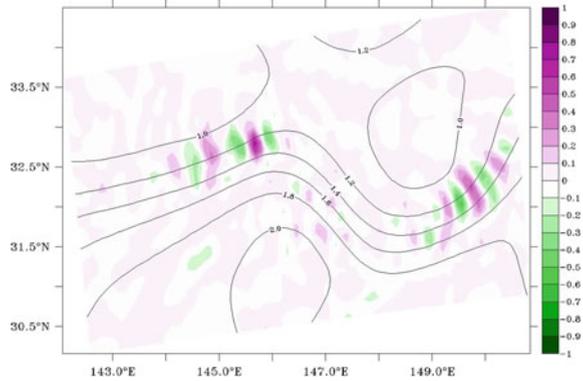
the expected power law scaling in T , given by the order of vanishing of the ramp function at the end points. The behavior for small ramp times is dominated by the deviation from adiabatic invariance and shows exponential behavior. This can be clearly seen from Figure 2.8 where the same data is plotted semi-logarithmically. The cutover point moves to larger values of T as the order of vanishing of the ramp at the end points is increased.

The results sketched above indicate that optimal balance can provide balanced flow fields which are almost as accurate as an optimally truncated asymptotic balance relation. While obtaining the latter is completely unfeasible in practice, optimal balance as a numerical technique is an easy and computationally feasible procedure. While it is more expensive than a simple evaluation of a low-order balance relation, it is not excessively expensive. Thus, in practical terms optimal balance may be the only accessible notion of balance that is accurate enough so that it can serve as the reference definition of a balanced state against which other, cheaper, diagnostics can be judged.

2.6 High-resolution Ocean General Circulation Models as a Novel Tool for Studying Spontaneous Emission

So far, modeling studies on spontaneous imbalance have been concentrated on flows with idealized or simplified configurations. Even though great progress has been made in understanding spontaneous imbalance as a phenomenon of geophysical fluids (see discussion and references Section 2.2) and attempts have been made to make the considered configurations to resemble the real atmospheric or oceanic conditions, modeling studies using idealized configurations are always challenged by the question of whether and to what extent spontaneous imbalance in the considered flow configuration matters in reality. These questions can be answered by considering observations with high spatial and temporal resolution. High resolution is essential for

Fig. 2.9 Snapshots of high-pass filtered vertical velocities (color shadings) and dynamically relevant pressure (contours) identified from a 10° simulation performed with MPIOM. Units are 10^{-4} m/s for vertical velocity and 10^{-1} bar for pressure



detecting spontaneous imbalance in the real ocean. As high-resolution observations are often not available, it is natural to use global general circulation models (GCMs) capable of realistically simulating the global ocean.

For a long time, ocean GCMs have not been suitable for studying spontaneous imbalance. State-of-the-art OGCMs, such as those contributing to the last IPCC report, are too coarse to resolve mesoscale eddies. Without mesoscale eddies, a coarse OGCM is unable to simulate realistic vortical flows capable of emitting waves. Due to the increase in computing power, global (or near-global) eddy-resolving or eddy-permitting simulations have become possible (e.g., Masumoto et al. 2004; Maltrud and McClean 2005; Lee et al. 2007; Maltrud et al. 2010; Griffies et al. 2015). At the Max-Planck-Institute, high-resolution ocean modeling was initiated by the German consortium project STORM. The 0.1° tripolar version of the MPIOM developed within the project simulates a considerable bulk of mesoscale eddies, with eddy characteristics comparable to those derived from observations (von Storch et al. 2012; Li and von Storch 2013; Li et al. 2015). This is an important prerequisite for studying the spontaneous wave emission by eddy flows in a realistic setting.

An essential feature of the eddy ocean simulated by the 0.1° MPIOM is horizontally axial flows. These flows follow a bundle of isobars in a horizontal plane, such as those indicated by the black lines in Figure 2.9. The isobars are curved, though not closed, and run parallel to each other. Flow following isobars is consistent with the idea that the eddy flows are near geostrophic balance. The shadings in Figure 2.9 show a snapshot of vertical velocity containing only super-inertial fluctuations, indicating gravity wave trains. The wave trains are concentrated along axial flows with the wave crests being perpendicular to the flow. Their vertical structures are comparable to those of low modes internal waves. They propagate along the flow as if they were captured.

These wave trains can result from the refraction of waves by the axial flow. Following equation 2.10, the effect of the refraction depends on the flow gradient tensor $\nabla \mathbf{u}$ along the path of the wave train. For a non-axial flow, $\nabla \mathbf{u}$ can be quite complicated so that \mathbf{k}_h is generally not in the direction of \mathbf{u}_h . For gravity waves, \mathbf{k}_h is in

the direction of the horizontal component of the intrinsic group velocity. For these waves, the wave trains tend to move away from the given flow, making refraction by a particular gradient tensor $\nabla \mathbf{u}$ difficult. An axial flow has the property that the shear of the flow vanishes along the core. This makes $\nabla \mathbf{u}$ to take an extremely simple form for non-divergent and horizontally axial flows. With such a $\nabla \mathbf{u}$, waves whose wavenumber vector does not have a transverse-flow component initially will stay in the core and eventually becomes refracted by the negative stretching rate such that their along-flow wavenumber increases exponentially.

Figure 2.9 provides a first indication that waves can be emitted and refracted by eddying flows in a realistic setting. Many questions are still unanswered. For instance, the 0.1° MPIOM-simulation discussed produces super-inertial kinetic energy not only in mid-latitude oceans but also, with much greater strength, in the subtropical and tropical ocean (see Figure 2.3). It is not clear whether these super-inertial fluctuations represent waves with properties similar to those found in the mid-latitude oceans shown in Figure 2.9, what is the relation of these waves to the background flows that are characterized by increasingly larger Rossby number toward the equator, and what is the role of these super-inertial fluctuations for large-scale subtropical and tropical flows.

Using high-resolution OGCMs as a tool for studying spontaneous imbalance requires further improvement of the models. In particular, a transition from primitive equation models (which resolve only long inertia-gravity waves) to non-hydrostatic models needs to be carried out to allow more complete studies on the role of spontaneous imbalance for flows in a realistic ocean.

To achieve a better understanding, it is worthwhile to combine a complex high-resolution OGCM with simpler conceptual models. For instance, one can use a high-resolution OGCM simulation as a guide for identifying the typical flow configurations likely to occur in the real ocean and study the spontaneous imbalance of these particular flow configurations in greater details using limited-area high-resolution setups. These typical flow configurations can also guide theoretical studies using, for example, matched asymptotic expansions.

2.7 Discussion

In this chapter, we have reviewed selected topics on the emission and capture of IGWs by nearly geostrophic flow. We further raised the question regarding their role as one of the routes to dissipation in the global ocean. To answer this question in general, we still need a more complete understanding of the underlying mechanisms and a careful assessment of different factors, including (i) a rigorous mathematical analysis of IGWs diagnostics, as well as improved diagnostics; (ii) scaling laws or appropriate alternative characterizations for the emission of IGWs, especially from frontal regions; and (iii) a quantitative assessment of IGWs emission and capture in realistic ocean GCMs.

Regarding point (i), one set of open questions concerns the use of balance relations, which may be interpreted as spatial filters, in addition or as an alternative time filtering for quantifying spontaneous imbalance. In particular, we have to understand whether some types of balance relations, e.g., semi-geostrophic vs. quasi-geostrophic balance or variational vs. nonvariational constructions are advantageous in a diagnostic setting. Moreover, do higher-order relations justify the added complexity? Do the diagnostics work consistently across all relevant oceanic (and eventually also atmospheric) regimes? How do the diagnostics transition from mid-latitude to equatorial balance, from weak to strong stratification, or behave near boundaries? Some of these questions are discussed in more detail in Franzke et al. (2019).

The second set of questions concerns the use of optimal balance for diagnostic purposes. While the analysis of optimal balance is well understood in the context of the finite dimensional model presented in Section 2.5.4, no rigorous results are yet available to cover even the RSW equations or any of the more complex models. To increase the applicability of optimal balance to a wide range of models, it is important to implement the ramp on top of a standard Eulerian formulation of the equations, e.g., by multiplying all nonlinear terms with the ramp function. To our knowledge, this has not been tested, but seems promising and within the framework outlined for the finite dimensional model problem in Section 2.5.4. Once this is achieved, we believe that the method has considerable potential for applications to the real ocean as a detailed understanding of the precise scalings is not necessary to apply the method so long as it is possible to adiabatically ramp the system into a simpler model for which balance can be unambiguously defined.

Regarding point (ii), a major issue is the necessity of some scale separation to be able to define any splitting into mean flow and IGW activity. This problem was raised by Saujani and Shepherd (2002) (see also Ford et al. 2002), but it clearly needs further studies in order to understand what happens when scale separation exists in large parts of the system, but is not uniform in space and time.

Regarding point (iii), studying IGW emission and capture using realistic OGCMs is novel but can be very difficult. This is because a high-resolution realistic OGCM generally resolves a much larger range of motions than a model configured for a specific flow configuration does. To be able to interpret a GCM output, the need for good diagnostics is particularly acute because of the different flow regimes present and the fact that we are starting to resolve IGWs.

Apart from diagnostics, theoretical considerations, such as those based on ray equations, are indispensable. They are needed to interpret the filtered output and to ensure that the output has proper physical meaning and does not result from numerical artifacts. In general, a number of questions are still open and two of these questions are particularly important and should be tackled first. One concerns the spontaneous imbalance in the subtropical and tropical ocean, which seems to produce much stronger super-inertial kinetic energy than that in mid-latitude oceans according to Figure 2.3. The other question concerns the difference in waves spontaneously emitted in different flow regimes, for instance, those in quasi-balanced axial flows, compared with those in a frontal system. Even though one difference, namely the difference in wave intensity, has been identified by theoretical considerations

(Vanneste 2008, 2013), the long-time behavior of waves emitted in different flow regimes and the resulting implications for the flows in different regimes have not been studied systematically. Both questions are crucial for understanding the role of spontaneous imbalance for the real ocean.

Progress on these questions will provide a deeper understanding of spontaneous emission and subsequent capture of IGWs, and lead to an assessment of the importance of the route to energy dissipation via wave emission. Eventually, this may lead to better parameterizations of IGWs in coarse-resolution models.

Acknowledgements We thank Tuba Masur for careful proof-reading of the manuscript. GB was partially supported by DFG grants FOR1740, BA-5068/8-1, and BA-5068/9-1. MO was partially supported by DFG grant OL-155/6-1.

References

- Allen, J.S., Holm, D.D., Newberger, P.A.: Toward an extended-geostrophic Euler-Poincaré model for mesoscale oceanographic flow. In: Norbury, J., Roulstone, I. (eds.) *Large-Scale Atmosphere-Ocean Dynamics*, vol. 1, pp. 101–125. Cambridge University Press (2002)
- Badin, G.: Surface semi-geostrophic dynamics in the ocean. *Geophys. Astrophys. Fluid Dyn.* **107**(5), 526–540 (2013)
- Barth, J., Allen, J., Newberger, P.: On intermediate models for barotropic continental shelf and slope flow fields. Part II: Comparison of numerical model solutions in doubly periodic domains. *J. Phys. Oceanogr.* **20**(7), 1044–1076 (1990)
- Blumen, W.: Geostrophic adjustment. *Rev. Geophys.* **10**(2), 485–528 (1972)
- Bouchut, F., Sommer, J.L., Zeitlin, V.: Frontal geostrophic adjustment and nonlinear wave phenomena in one-dimensional rotating shallow water. Part 2. High-resolution numerical simulations. *J. Fluid Mech.* **514**, 35–63 (2004)
- Broutman, D., Rottman, J.W., Eckermann, S.D.: Ray methods for internal waves in the atmosphere and ocean. *Annu. Rev. Fluid Mech.* **36**, 233–253 (2004)
- Bühler, O., McIntyre, M.E.: Wave capture and wave-vortex duality. *J. Fluid Mech.* **534**, 67–95 (2005)
- Charney, J.G.: Geostrophic turbulence. *J. Atmos. Sci.* **28**(6), 1087–1095 (1971)
- Cho, H.R., Shepherd, T.G., Vladimirov, V.A.: Application of the direct liapunov method to the problem of symmetric stability in the atmosphere. *J. Atmos. Sci.* **50**(6), 822–836 (1993)
- Constantin, P., Majda, A., Tabak, E.: Formation of strong fronts in the 2-D quasigeostrophic thermal active scalar. *Nonlinearity* **7**, 1495–1533 (1994)
- Cotter, C.: Data assimilation on the exponentially accurate slow manifold. *Philos. Trans. R. Soc. A* **371**(1991), 20120300 (2013)
- Cotter, C.J., Reich, S.: Semigeostrophic particle motion and exponentially accurate normal forms. *Multiscale Model. Simul.* **5**(2), 476–496 (2006)
- Danioux, E., Vanneste, J., Klein, P., Sasaki, H.: Spontaneous inertia-gravity-wave generation by surface-intensified turbulence. *J. Fluid Mech.* **699**, 153–173 (2012)
- Dritschel, D.G., Polvani, L.M., Mohebalhojeh, A.R.: The contour-advective semi-lagrangian algorithm for the shallow water equations. *Mon. Weather Rev.* **127**(7), 1551–1565 (1999)
- Ford, R.: Gravity wave radiation from vortex trains in rotating shallow water. *J. Fluid Mech.* **281**, 81–118 (1994)
- Ford, R., McIntyre, M.E., Norton, W.A.: Balance and the slow quasimanifold: some explicit results. *J. Atmos. Sci.* **57**(9), 1236–1254 (2000)
- Ford, R., McIntyre, M.E., Norton, W.A.: Reply. *J. Atmos. Sci.* **59**(19), 2878–2882 (2002)

- Franzke, C.L.E., Oliver, M., Rademacher, J.D.M., Badin, G.: Multi-scale methods for geophysical flows. This volume, Chapter 1 (2019)
- Gall, R.L., Williams, R.T., Clark, T.L.: Gravity waves generated during frontogenesis. *J. Atmos. Sci.* **45**(15), 2204–2219 (1988)
- Gottwald, G.A., Mohamad, H., Oliver, M.: Optimal balance via adiabatic invariance of approximate slow manifolds. *Multiscale Model. Simul.* **15**(4), 1404–1422 (2017)
- Griffies, S.M., Winton, M., Anderson, W.G., Benson, R., Delworth, T.L., Dufour, C.O., Dunne, J.P., Goddard, P., Morrison, A.K., Rosati, A., Wittenberg, A.T., Yin, J., Zhang, R.: Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *J. Clim.* **28**(3), 952–977 (2015)
- Hoskins, B.J.: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.* **32**(2), 233–242 (1975)
- Hoskins, B.J., Bretherton, F.P.: Atmospheric frontogenesis models: mathematical formulation and solution. *J. Atmos. Sci.* **29**(1), 11–37 (1972)
- Jones, W.L.: Ray tracing for internal gravity waves. *J. Geophys. Res.* **74**(8), 2028–2033 (1969)
- Klein, P., Treguier, A.M.: Dispersion of wind-induced inertial waves by a barotropic jet. *J. Mar. Res.* **53**(1), 1–22 (1995)
- Kunze, E.: Near-inertial wave propagation in geostrophic shear. *J. Phys. Oceanogr.* **15**(5), 544–565 (1985)
- Kunze, E., Boss, E.: A model for vortex-trapped internal waves. *J. Phys. Oceanogr.* **28**(10), 2104–2115 (1998)
- Le Sommer, J., Medvedev, S., Plougonven, R., Zeitlin, V.: Singularity formation during relaxation of jets and fronts toward the state of geostrophic equilibrium. *J. Fluid Mech.* **758**, 287–311 (2014)
- Le Sommer, J., Reznik, G.M., Zeitlin, V.: Nonlinear geostrophic adjustment of long-wave disturbances in the shallow-water model on the equatorial beta-plane. *J. Fluid Mech.* **515**, 135–170 (2004)
- Lee, M.-M., Nurser, A.J.G., Coward, A.C., de Cuevas, B.A.: Eddy advective and diffusive transports of heat and salt in the southern ocean. *J. Phys. Oceanogr.* **37**(5), 1376–1393 (2007)
- Li, H., von Storch, J.-S.: On the fluctuating buoyancy fluxes simulated in a OGCM. *J. Phys. Oceanogr.* **43**(7), 1270–1287 (2013)
- Li, Z., von Storch, J.-S., Müller, M.: The M_2 internal tide simulated by a $1/10^\circ$ OGCM. *J. Phys. Oceanogr.* **45**(12), 3119–3135 (2015)
- Lighthill, M.J.: On sound generated aerodynamically. I. General theory. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **211**(1107), 564–587 (1952)
- Llewellyn Smith, S.G.: Near-inertial oscillations of a barotropic vortex: trapped modes and time evolution. *J. Phys. Oceanogr.* **29**(4), 747–761 (1999)
- Lott, F., Plougonven, R., Vanneste, J.: Gravity waves generated by sheared potential vorticity anomalies. *J. Atmos. Sci.* **67**(1), 157–170 (2010)
- Lott, F., Plougonven, R., Vanneste, J.: Gravity waves generated by sheared three-dimensional potential vorticity anomalies. *J. Atmos. Sci.* **69**(7), 2134–2151 (2012)
- Maltrud, M., Bryan, F., Peacock, S.: Boundary impulse response functions in a century-long eddying global ocean simulation. *Environ. Fluid Mech.* **10**(1–2), 275–295 (2010)
- Maltrud, M.E., McClean, J.L.: An eddy resolving global $1/10^\circ$ ocean simulation. *Ocean Model.* **8**(1–2), 31–54 (2005)
- Masumoto, Y., Sasaki, H., Kagimoto, T., Komori, N., Ishida, A., Sasai, Y., Miyama, T., Motoi, T., Mitsudera, H., Takahashi, K., Sakuma, H., Yamagata, T.: A fifty-year eddy-resolving simulation of the world ocean—preliminary outcomes of OFES (OGCM for the Earth Simulator). *J. Earth Simulator* **1**, 35–56 (2004)
- McIntyre, M.: Dynamical meteorology—balanced flow. In: Pyle, J., Zhang, F. (eds.) *Encyclopedia of Atmospheric Sciences*, pp. 298–303, 2nd edn. Academic Press, Oxford (2015)
- McIntyre, M.E.: Spontaneous imbalance and hybrid vortex-gravity structures. *J. Atmos. Sci.* **66**, 1315–1326 (2009)

- McIntyre, M.E., Norton, W.A.: Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.* **57**(9), 1214–1235 (2000)
- McIntyre, M.E., Roulstone, I.: Are there higher-accuracy analogues of semi-geostrophic theory? In: Norbury, J., Roulstone, I. (eds.) *Large-Scale Atmosphere-Ocean Dynamics*, vol. 2, pp. 301–364. Cambridge University Press (2002)
- Medvedev, S.B., Zeitlin, V.: Weak turbulence of short equatorial waves. *Phys. Lett. A* **342**(3), 217–227 (2005)
- Mohebalhojeh, A.R., Dritschel, D.G.: Hierarchies of balance conditions for the f -plane shallow-water equations. *J. Atmos. Sci.* **58**(16), 2411–2426 (2001)
- Müller, P., McWilliams, J.C., Molemaker, M.J.: Routes to dissipation in the ocean: the 2D/3D turbulence conundrum. In: Baumert, H.Z., Simpson, J., Sündermann, J. (eds.) *Marine Turbulence*, pp. 397–405. Cambridge University Press (2005)
- Muraki, D.J., Snyder, C., Rotunno, R.: The next-order corrections to quasigeostrophic theory. *J. Atmos. Sci.* **56**(11), 1547–1560 (1999)
- Nadiga, B.T.: Nonlinear evolution of a baroclinic wave and imbalanced dissipation. *J. Fluid Mech.* **756**, 965–1006 (2014)
- Nagai, T., Tandon, A., Kunze, E., Mahadevan, A.: Spontaneous generation of near-inertial waves by the Kuroshio front. *J. Phys. Oceanogr.* **45**(9), 2381–2406 (2015)
- Nikurashin, M., Ferrari, R.: Global energy conversion rate from geostrophic flows into internal lee waves in the deep ocean. *Geophys. Res. Lett.* **38**(8), L08610 (2011)
- Nikurashin, M., Vallis, G.K., Adcroft, A.: Routes to energy dissipation for geostrophic flows in the southern ocean. *Nat. Geosci.* **6**(1), 48–51 (2013)
- Obukhov, A.: On the problem of geostrophic wind. *Izv. AN SSSR Geogr. Geophys. Ser.* **13**, 281–286 (1949)
- Olbers, D., Willebrand, J., Eden, C.: *Ocean Dynamics*. Springer (2012)
- Oliver, M., Vasylykevych, S.: Generalized large-scale semigeostrophic approximations for the f -plane primitive equations. *J. Phys. A: Math. Theor.* **49**, 184001 (2016)
- O’Sullivan, D., Dunkerton, T.J.: Generation of inertia-gravity waves in a simulated life cycle of baroclinic instability. *J. Atmos. Sci.* **52**, 3695–3716 (1995)
- Plougonven, R., Snyder, C.: Gravity waves excited by jets: propagation versus generation. *Geophys. Res. Lett.* **32**(18), L18802 (2005)
- Plougonven, R., Snyder, C.: Inertia-gravity waves spontaneously generated by jets and fronts. Part I: different baroclinic life cycles. *J. Atmos. Sci.* **64**(7), 2502–2520 (2007)
- Plougonven, R., Zeitlin, V.: Lagrangian approach to geostrophic adjustment of frontal anomalies in a stratified fluid. *Geophys. Astrophys. Fluid Dyn.* **99**(2), 101–135 (2005)
- Plougonven, R., Zhang, F.: Internal gravity waves from atmospheric jets and fronts. *Rev. Geophys.* **52**(1), 33–76 (2014)
- Ragone, F., Badin, G.: A study of surface semi-geostrophic turbulence: freely decaying dynamics. *J. Fluid Mech.* **792**, 740–774 (2016)
- Reznik, G.M., Zeitlin, V., Jelloul, M.B.: Nonlinear theory of geostrophic adjustment. Part 1. Rotating shallow-water model. *J. Fluid Mech.* **445**, 93–120 (2001)
- Rossby, C.-G.: On the mutual adjustment of pressure and velocity distributions in certain simple current systems, II. *J. Mar. Res.* **1**(3), 239–263 (1938)
- Salmon, R.: New equations for nearly geostrophic flow. *J. Fluid Mech.* **153**, 461–477 (1985)
- Salmon, R.: Large-scale semigeostrophic equations for use in ocean circulation models. *J. Fluid Mech.* **318**, 85–105 (1996)
- Saujani, S., Shepherd, T.G.: Comments on “balance and the slow quasimanifold: some explicit results”. *J. Atmos. Sci.* **59**(19), 2874–2877 (2002)
- Snyder, C., Muraki, D.J., Plougonven, R., Zhang, F.: Inertia-gravity waves generated within a dipole vortex. *J. Atmos. Sci.* **64**(12), 4417–4431 (2007)
- Snyder, C., Plougonven, R., Muraki, D.J.: Mechanisms for spontaneous gravity wave generation within a dipole vortex. *J. Atmos. Sci.* **66**(11), 3464–3478 (2009)

- Thomas, L.N.: On the effects of frontogenetic strain on symmetric instability and inertia-gravity waves. *J. Fluid Mech.* **711**, 620–640 (2012)
- van Haren, H.: On the polarization of oscillatory currents in the Bay of Biscay. *J. Geophys. Res. Oceans* **108**(C9), 3290 (2003)
- Vanneste, J.: Exponential smallness of inertia-gravity wave generation at small Rossby number. *J. Atmos. Sci.* **65**(5), 1622–1637 (2008)
- Vanneste, J.: Balance and spontaneous wave generation in geophysical flows. *Ann. Rev. Fluid Mech.* **45**(1), 147–172 (2013)
- Vanneste, J., Yavneh, I.: Exponentially small inertia-gravity waves and the breakdown of quasi-geostrophic balance. *J. Atmos. Sci.* **61**(2), 211–223 (2004)
- Vanneste, J., Yavneh, I.: Unbalanced instabilities of rapidly rotating stratified shear flows. *J. Fluid Mech.* **584**, 373–396 (2007)
- Viúdez, A.: Spiral patterns of inertia-gravity waves in geophysical flows. *J. Fluid Mech.* **562**, 73–82 (2006)
- Viúdez, A.: The origin of the stationary frontal wave packet spontaneously generated in rotating stratified vortex dipoles. *J. Fluid Mech.* **593**, 359–383 (2007)
- Viúdez, A., Dritschel, D.G.: Optimal potential vorticity balance of geophysical flows. *J. Fluid Mech.* **521**, 343–352 (2004)
- Viúdez, A., Dritschel, D.G.: Spontaneous generation of inertia-gravity wave packets by balanced geophysical flows. *J. Fluid Mech.* **553**, 107–117 (2006)
- von Storch, H., Zwiers, F.W.: *Statistical Analysis in Climate Research*, p. 484. Cambridge University Press (1999)
- von Storch, J.-S., Eden, C., Fast, I., Haak, H., Hernández-Deckers, D., Maier-Reimer, E., Marotzke, J., Stammer, D.: An estimate of the Lorenz energy cycle for the world ocean based on the STORM/NCEP simulation. *J. Phys. Oceanogr.* **42**(12), 2185–2205 (2012)
- Wang, S., Zhang, F.: Source of gravity waves within a vortex-dipole jet revealed by a linear model. *J. Atmos. Sci.* **67**(5), 1438–1455 (2010)
- Wang, S., Zhang, F., Snyder, C.: Generation and propagation of inertia-gravity waves from vortex dipoles and jets. *J. Atmos. Sci.* **66**(5), 1294–1314 (2009)
- Xing, J., Davies, A.M.: On the influence of a surface coastal front on near-inertial wind-induced internal wave generation. *J. Geophys. Res. Oceans* **109**(C1), C01023 (2004)
- Zeitlin, V.: Lagrangian dynamics of fronts, vortices and waves: understanding the (semi-) geostrophic adjustment. In: *Fronts, Waves and Vortices in Geophysical Flows*, pp. 109–137. Springer (2010)
- Zeitlin, V., Medvedev, S.B., Plougonven, R.: Frontal geostrophic adjustment, slow manifold and nonlinear wave phenomena in one-dimensional rotating shallow water. Part 1. Theory. *J. Fluid Mech.* **481**, 269–290 (2003)
- Zhang, F.: Generation of mesoscale gravity waves in upper-tropospheric jet-front systems. *J. Atmos. Sci.* **61**(4), 440–457 (2004)

Chapter 3

The IDEMIX Model: Parameterization of Internal Gravity Waves for Circulation Models of Ocean and Atmosphere



Dirk Olbers, Carsten Eden, Erich Becker, Friederike Pollmann
and Johann Jungclaus

Abstract The IDEMIX concept is an energetically consistent framework to describe wave effects in circulation models of ocean and atmosphere. It is based on the radiative transfer equation for an internal gravity wave field in physical and wavenumber space and was shown to be successful for ocean applications. An improved IDEMIX model for the ocean will be constructed and extended by a new high-frequency, high vertical wavenumber compartment, forcing by mesoscale eddy dissipation, anisotropic tidal forcing, and wave–mean flow interaction. It will be validated using observational and model estimates. A novel concept for gravity wave parameterization in atmospheric circulation models is developed. As for the ocean, the wave field is represented by the wave energy density in physical and wavenumber space, and its prognostic computation is performed by the radiative transfer equation. This new concept goes far beyond conventional gravity wave schemes which are based on the single-column approximation and, in particular, on the strong assumptions of a stationary mean flow and a stationary wave energy equation. The radiative transfer equation has—to our knowledge—never been considered in the atmospheric

D. Olbers (✉)

Alfred Wegener Institute for Polar and Marine Research (AWI), Bremerhaven, Germany
e-mail: dirk.olbers@awi.de

D. Olbers

MARUM/Institut für Umweltphysik, Universität Bremen, Bremen, Germany

C. Eden · F. Pollmann

Center for Earth System Research and Sustainability (CEN), Universität Hamburg,
Hamburg, Germany
e-mail: carsten.eden@uni-hamburg.de

F. Pollmann

e-mail: friederike.pollmann@uni-hamburg.de

E. Becker

Leibniz-Institut für Atmosphärenphysik e.V., Universität Rostock, Rostock, Germany
e-mail: becker@iap-kborn.de

J. Jungclaus

Max-Planck-Institut für Meteorologie (MPI-M), Hamburg, Germany
e-mail: johann.jungclaus@mpimet.mpg.de

© Springer Nature Switzerland AG 2019

C. Eden and A. Iske (eds.), *Energy Transfers in Atmosphere and Ocean*,
Mathematics of Planet Earth 1, https://doi.org/10.1007/978-3-030-05704-6_3

community as a framework for subgrid-scale parameterization. The proposed parameterization will, for the first time, (1) include all relevant sources continuously in space and time and (2) accommodate all gravity wave sources (orography, fronts, and convection) in a single parameterization framework. Moreover, the new scheme is formulated in a precisely energy-preserving fashion. The project will contribute to a transfer of knowledge from the oceanic community to the atmospheric community and vice versa. We give a brief description of the oceanic and atmospheric internal wave fields, the most important processes of generation and interactions, and the ingredients of the model IDEMIX.

3.1 Internal Waves in Ocean and Atmosphere

Internal gravity waves arise in a stably stratified fluid through the restoring force of gravity on fluid particles displaced from their equilibrium levels. Interfacial waves occurring between two superposed layers of different density are a familiar phenomenon, in particular at the upper free surface of the ocean in form of surface waves. In the continuously stratified interior of the ocean, the restoring force of gravity is much weaker (by a factor $\delta\rho/\rho = 10^{-3}$, where $\delta\rho$ is a typical density perturbation of the mean density ρ), and the periods and wavelengths of internal waves are much larger than those of surface gravity waves. In the spectrum of oceanic motions, internal gravity waves are embedded between (and partly overlap with) small-scale three-dimensional turbulence and the geostrophic balanced motion of the oceanic eddy field, as depicted in Figure 3.1. The timescales of baroclinic gravity waves are sharply defined as being in between the stability frequency $N \sim \delta\rho/\rho$, called Brunt-Väisälä frequency and related to the buoyancy force, and the Coriolis frequency f , related to the Earth rotation and Coriolis force. Spatial scales can range from global scales in case of long barotropic gravity waves down to a couple of 10 m for the baroclinic gravity wave branch. On even smaller time and space scales, the internal wave regime approaches isotropic turbulence which then connects to the regime of ultimate dissipation of energy by molecular processes. Atmospheric gravity waves obey the same constraints in frequency as oceanic ones; however, dominant wavelengths are generally larger: spectra of vertical wavelength peak at 2 to 5 km in the lower stratosphere and increase to 10 to 30 km in the mesopause. Contrary to the ocean where we always see a continuum in frequency–wavenumber space, there is ample evidence that atmospheric spectra are often composed of only a few waves (see, e.g., Fritts and Alexander 2003).

Waves are an essentially linear disturbance of the wave-carrying medium: once they are generated, they propagate almost freely along their rays, as depicted in Figure 3.2, slowly changing by non-linear effects and coupling to their supporting background, thereby slowly losing attributes acquired during their particular generation process. Strongly non-linear effects such as breaking occur only as very localized events in space and time. A linear wave is characterized by an amplitude $a(\mathbf{K})$, and

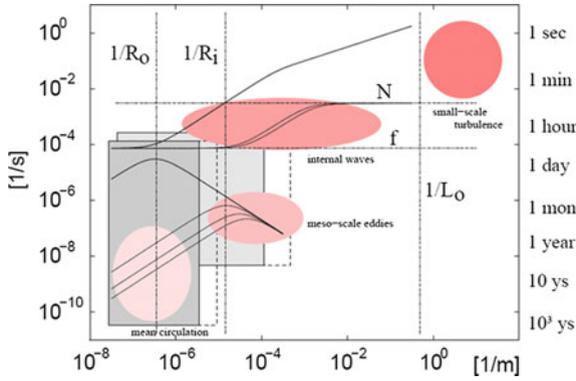
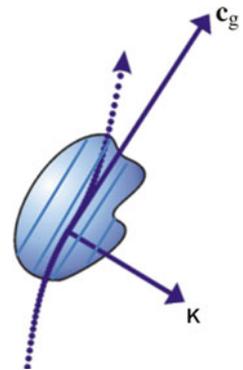


Fig. 3.1 Space-timescales of important oceanic processes (pink areas) and scales explicitly resolved by ocean models (grey rectangular areas). The lower left rectangle represents modern global ocean climate models and the upper right rectangle eddy-resolving basin-scale models. Also shown are dispersion curves (solid lines) for linear gravity waves (upper set) and planetary waves (lower set). Vertical dotted lines indicate the external (R_o) and the first internal (R_i) Rossby radii and the Ozmidov length scale L_o

Fig. 3.2 Sketch of wave packet propagation along a ray. A perpendicular orientation of \mathbf{K} and the intrinsic group velocity $\mathbf{c}_g = \partial_{\mathbf{K}}\Omega$, as depicted here, is realized by internal gravity waves. Wave crests and troughs are orthogonal to \mathbf{K} , and these phase lines show propagation along \mathbf{K}



a wave vector \mathbf{K} and an intrinsic¹ frequency ω , which are related by a dispersion relation $\omega = \Omega(\mathbf{K})$. For internal waves

$$\omega = \Omega(\mathbf{K}) = \left(\frac{N^2 k^2 + f^2 m^2}{k^2 + m^2} \right)^{1/2} \tag{3.1}$$

where the three-dimensional wave vector $\mathbf{K} = (\mathbf{k}, m)$ is split into the horizontal and vertical components, and $k = |\mathbf{k}|$. Correspondingly, we will use $\mathbf{X} = (\mathbf{x}, z)$ for the position vector. Large-scale inhomogeneities (compared to period and wavelength) of

¹We denote by ω the intrinsic frequency, i.e. the frequency observed relative to a mean flow. Then, the Doppler shifted frequency of encounter, $\omega_{enc} = \omega + \mathbf{K} \cdot \mathbf{U}$, is the one observed at a fixed position in space.

the wave-carrying background can be treated by WKB methods (see, e.g., Berry and Mount 1972; Bender and Orszag 1978). Such inhomogeneities arise via a space and time-dependent Brunt-Väisälä frequency, and mean current, and a spatially varying Coriolis frequency. Waves then appear in the form of slowly varying wavetrains which may be represented locally by wave groups (or packets) characterized by a local dispersion relation $\omega = \Omega(\mathbf{K}, \mathbf{X}, t)$ where \mathbf{X} is the spatial coordinate and t is time. A mean current $\mathbf{U}(\mathbf{X}, t)$ is included as a Doppler shift so that $\omega_{enc} = \omega + \mathbf{K} \cdot \mathbf{U}$ represents the frequency of encounter. A wave group propagates with the group velocity

$$\dot{\mathbf{X}} = \partial_{\mathbf{K}}\Omega + \mathbf{U} = \mathbf{V} + \mathbf{U} \quad (3.2)$$

where $\mathbf{V} = \partial_{\mathbf{K}}\Omega$ is the intrinsic group velocity. The wave vector changes along the trajectory (ray) according to

$$\dot{\mathbf{K}} = -\partial_{\mathbf{X}}\Omega - \mathbf{K} \cdot (\nabla\mathbf{U}) \quad (3.3)$$

The process is called refraction. Here, \mathbf{K} contracts with \mathbf{U} . The influence of the mean flow in these expressions takes place via simple advection (in (3.2)) and the gradient matrix of the mean current (in (3.3)). The vertical mean current is usually negligible so that $\mathbf{U} = (U, V, 0)$. The intrinsic group velocity \mathbf{V} for internal waves has a peculiar property: $\partial_{\mathbf{K}}\Omega$ is perpendicular to the wave vector \mathbf{K} (the group propagation is orthogonal to the phase propagation), and because the horizontal component $\partial_{\mathbf{K}}\Omega$ is aligned with the horizontal wave vector \mathbf{k} , the vertical component is opposed to the vertical wavenumber, $\partial_m\Omega \sim -\text{sign } m$. This property is important for the IDEMIX equations (see Section 3.2). The intrinsic refraction, $-\partial_{\mathbf{X}}\Omega$, mainly arises from a depth-dependent Brunt-Väisälä frequency, $N = N(z)$. The gradient term $\nabla\mathbf{U}$ in (3.3) leads to the occurrence of critical layers and is also responsible for wave capture effects (see Section 3.3.2).

Writing the dispersion relation for the frequency of encounter as

$$\omega_{enc} = \Omega_{enc}(\mathbf{K}, \mathbf{X}, t) = \Omega(\mathbf{K}, z) + \mathbf{K} \cdot \mathbf{U}(\mathbf{X}, t) \quad (3.4)$$

we have, according to WKB theory, for the rate of change of frequencies along the ray

$$\begin{aligned} \dot{\omega}_{enc} &= \partial_t\Omega_{enc} = \partial_t\Omega + \mathbf{K} \cdot \partial_t\mathbf{U} \\ \dot{\omega} &= \partial_t\Omega + \mathbf{U} \cdot \partial_{\mathbf{X}}\Omega - \mathbf{V} \cdot (\mathbf{K} \cdot \partial_{\mathbf{X}}\mathbf{U}) \end{aligned} \quad (3.5)$$

Note that \mathbf{K} contracts with \mathbf{U} and \mathbf{V} contracts with $\partial_{\mathbf{X}}$ in this expression. These frequency relations are consistent with $\omega = \omega_{enc} - \mathbf{K} \cdot \mathbf{U}$ as integral of the motion along the ray.

If the background medium is time dependent (slow in the WKB sense), resulting in a time-dependent $\Omega(\mathbf{K}, \mathbf{X}, t)$, the frequency of encounter changes as given by

(3.5). The wave energy $E \sim |a|^2$ is then not conserved, but wave action $A(\mathbf{X}, t) = E(\mathbf{X}, t)/\omega$ (energy over *intrinsic* frequency) is an adiabatic invariant (Landau and Lifshitz 1982; Bretherton and Garrett 1968; Whitham 1970),

$$\partial_t A + \nabla_{\mathbf{X}} \cdot (\dot{\mathbf{X}} A) = 0 \tag{3.6}$$

This property is fundamental for the radiation balance discussed further below.

In a realistic geophysical situation, the wave field is more likely described by a superposition of a great number of wave packets, each localized in physical space and having a dominant wave vector, frequency, and amplitude, which slowly change as a consequence of propagation, refraction, and reflection according to the above ray equations. When two wave packets occupy the same volume, they might interact resonantly for a short finite time and build up a third wave component. This is the process of wave–wave interactions. Wave packets may interact with background fields, e.g. the mean flow, in which they propagate, and new packets may be introduced in the wave ensemble by forcing mechanisms to be discussed in later sections. Instead of using the energy or action of single waves, we describe such a wave field by its energy (power) spectrum² $\mathcal{E}(\mathbf{K}, \mathbf{X}, t) = \omega \mathcal{A}(\mathbf{K}, \mathbf{X}, t)$, defined such that the integral over wavenumbers yields the local energy density $E(\mathbf{X}, t)$. Because the action spectrum is now written as function of \mathbf{K} in addition to previous independent variables \mathbf{X} and t , and the wave vector is also slowly changing, the action conservation reads for the random wave field (Hasselmann 1968)

$$\partial_t \mathcal{A} + \nabla_{\mathbf{X}} \cdot (\dot{\mathbf{X}} \mathcal{A}) + \nabla_{\mathbf{K}} \cdot (\dot{\mathbf{K}} \mathcal{A}) = \mathcal{S} \tag{3.7}$$

Here, \mathcal{S} is a source, not yet considered in (3.6), representing all processes that may lead to a change of the action spectrum, except for the propagation and refraction processes which are explicitly accounted for on the left-hand side. The vertical energy flux $\mathbf{F} = \dot{z} \mathcal{E}$ must be specified at the top and bottom boundaries, assuming for simplicity that these are horizontal surfaces. At the surface

$$\mathbf{F}(\mathbf{k}, m) + \mathbf{F}(\mathbf{k}, -m) = \Phi_{surf}(\mathbf{k}, m) \tag{3.8}$$

must hold, and similarly for the bottom with a net flux $\Phi_{bot}(\mathbf{k}, m)$. The condition accounts for reflection, in which a wave with vertical wavenumber m is reflected into one with $-m$, and an energy source $\Phi_{surf}(\mathbf{k}, m)$ by a wave-maker situated at the surface as, e.g. wind stress fluctuations, or tidal conversion at the bottom in case of $\Phi_{bot}(\mathbf{k}, m)$.

Application of the radiative transfer will be done in wavenumber coordinates different from the Cartesian ones (\mathbf{k}, m) . This is because dominant forcing functions are more easily embedded in frequency space, e.g. tidal forcing and near-inertial wave radiation. Also, directional spreading of horizontal wavenumbers is more easily

² $\mathcal{E}(\mathbf{K}, \mathbf{X}, t) d^3 K$ is the wave energy (density in physical space) in a small wavenumber volume $d^3 K$ at \mathbf{K} at the position \mathbf{X} at time t .

formulated in angular coordinates. We therefore transform the radiation balance (3.7) into more convenient coordinates. For the balance of $\tilde{\mathcal{E}} = \tilde{\mathcal{E}}(m, \omega, \phi)$, we find

$$\partial_r \tilde{\mathcal{E}} + \nabla \cdot \dot{\mathbf{x}} \tilde{\mathcal{E}} + \partial_z \dot{z} \tilde{\mathcal{E}} + \partial_\omega \dot{\omega} \tilde{\mathcal{E}} + \partial_m \dot{m} \tilde{\mathcal{E}} + \partial_\phi \dot{\phi} \tilde{\mathcal{E}} = \omega \tilde{\mathcal{S}} + \dot{\omega} \frac{\tilde{\mathcal{E}}}{\omega} \quad (3.9)$$

Here, $\dot{\omega}$ is the change of intrinsic frequency along the ray, given by (3.5). The term $\dot{\omega} \tilde{\mathcal{E}}/\omega$ contains the energy exchange between the waves and the mean flow. We abandon the tilde in the further discussions.

The knowledge about the structure and importance of the oceanic internal wave field is strongly based on experimental evidence of the wave motion. Amplitudes of internal gravity waves are remarkably large, of the order of 10 m (occasionally, they may be an order of magnitude larger), and current speeds are typically 5 cm/s. The wave motion is therefore not difficult to observe; in fact, it is the dominant signal in many oceanic measurements. The first attempt to provide a unified picture of the internal wave field was made by Garrett and Munk (1972) who synthesized a model of the complete wavenumber–frequency spectrum (GM model) of the motion in the deep ocean on the basis of linear theory and the available observations by horizontally or vertically separated moored instruments or dropped sondes. Except for inertial internal waves and baroclinic internal tides, this model is believed to reflect the spectral features of the internal wave climate in the deep ocean and to possess a certain global validity. Most data were in good agreement with the GM model or could be incorporated by slight modifications (Garrett and Munk 1975; Cairns and Williams 1976; Müller et al. 1978; Munk 1981). In a broad-brush view, the GM spectrum is characterized by a ω^{-2} decay of energy power in frequency space with a minor peak at $\omega = f$, and a m^{-2} decay in vertical wavenumber space with a roll-off at $m = m_*$ to a plateau at low wavenumbers; see Figure 3.3. The wavenumber m_* of the roll-off is referred to as the bandwidth of the spectrum. Note that GM is horizontally isotropic, vertically symmetric, and of the factorized form $\mathcal{E}(m, \omega, \phi, z) = E(z)A(|m|)B(\omega)/2\pi$.

Atmospheric gravity wave spectra are reviewed by Fritts and Alexander (2003), and a working base for our purpose of constructing an atmospheric IDEMIX model is presented in Section 3.2.2.2. The general functional class considered for the spectrum is in fact very similar to GM except for a distinct azimuthal distribution of wave propagation directions, $\mathcal{E}(m, \omega, \phi, z) = E(z)A(|m|)B(\omega)\Phi(\phi)$. The power laws in frequency and vertical wavenumber are different from GM and seen as not as universal as the oceanic counterpart. The energy level $E(z)$ and peak (or roll-off) wavenumber m_* are not following WKB scaling as GM does. Also, as Fritts and Alexander (2003) emphasize, this canonical spectrum does not capture the true complexity of the gravity wave field in altitude and does not account for considerable variability. To a certain degree, this drawback applies as well to GM.

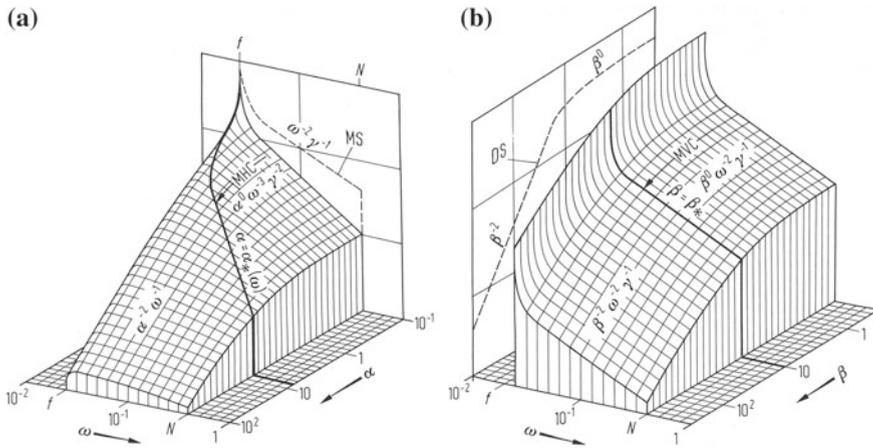


Fig. 3.3 GM76 model, displayed in (a) as $\mathcal{E}(k, \omega)$ and in (b) as $\mathcal{E}(m, \omega)$. The coordinates are plotted logarithmically so that plane surfaces represent power laws, some of which are indicated in the graphs. The partially integrated forms MS and DS of the moored and dropped spectra, respectively, are displayed as respective projections, and the moored coherences MHC and MVC are related to the corresponding bandwidths, as indicated. All quantities are normalized with reference to the scale b and N_0 of the Brunt-Väisälä frequency profile $N(z) = N_0 e^{z/b}$. In the figure, the notation $\alpha = k$, $\beta = m$, and $\gamma = (1 - f^2/\omega^2)^{1/2}$ is used. After Garrett and Munk (1975), Olbers (1986)

3.2 The IDEMIX Model

Müller and Natarov (2003) suggested to base a model for the propagation and dissipation of internal waves on the radiative transfer equation (3.7) of weakly interacting waves in the 6-dimensional phase space. However, theoretical, practical, and numerical limits hamper the realization of such a comprehensive model. Olbers and Eden (2013) discussed a drastic simplification of the concept which they called IDEMIX (Internal wave Dissipation, Energy and Mixing). Instead of resolving the detailed wave spectrum as suggested by Müller and Natarov (2003), they integrate the spectrum in wavenumber and frequency domain, leading to conservation equations for integral energy compartments in physical space. These equations can be closed with a few simple but reasonable parameterizations. IDEMIX describes the generation, interaction, propagation, and dissipation of the internal gravity wave field and can be used in ocean general circulation models to account for vertical mixing (and friction) in the interior of the ocean. In its simplest version, IDEMIX consists of two compartments of interacting up- and downward propagating waves (Olbers and Eden 2013). In a more complex version, low vertical mode compartments at near-inertial frequencies and frequencies of tidal constituents are added which also account for horizontally anisotropic wave propagation (Eden and Olbers 2014). For some more details, we refer to Section 3.2.1. Eden et al. (2014) demonstrate how an energetically consistent ocean model can be constructed connecting IDEMIX with other

energy-based parameterizations for the unresolved dynamical regimes of mesoscale eddies and small-scale turbulence. IDEMIX is central to the concept of an energetically consistent ocean model, since it enables to link all sources and sinks of internal wave energy, and furthermore all parameterized forms of energy in an ocean model without spurious sources and sinks of energy.

3.2.1 Details of the Oceanic IDEMIX

Both the simple and extended versions of IDEMIX are based on the radiative transfer equation (3.9) and the boundary condition (3.8) which describe the evolution in time of the energy spectrum $\mathcal{E}(m, \omega, \phi, \mathbf{x}, z, t)$ of an ensemble of weakly interacting gravity waves in wavenumber and physical space. Using the 6-dimensional space $(m, \omega, \phi, \mathbf{x}, z)$ is too difficult, and integrated energy compartments are thus considered instead. In the simple version of IDEMIX by Olbers and Eden (2013), \mathcal{E} is integrated over frequency ω , azimuth ϕ , and over vertical wavenumber m , but separately for positive m (yields E^- , downward propagating waves) and negative m (yields E^+ , upward propagating waves). Defining total energy $E = E^+ + E^-$ and energy asymmetry $\Delta E = E^+ - E^-$, the projection of (3.7) leads to

$$\begin{aligned} \partial_t E + \partial_z c_0 \Delta E &= -F_{diss} = -\mu E^2 \\ \partial_t \Delta E + \partial_z c_0 E &= F_{ww} = -\Delta E / \tau_v \end{aligned} \quad (3.10)$$

This applies to the simplest IDEMIX model where horizontal homogeneity is assumed. The crosswise form of the vertical energy fluxes in (3.10) derives from the vertical group velocity $\dot{z} = \pm c$ of up- and downward waves being opposed to the vertical wavenumber m . However, a parameterization of the integrated group velocity is needed, and this is done in a typical way for parameters in IDEMIX energy balances: here, the mean vertical propagation with group speed c_0 is calculated analytically by assuming a spectrum of the gravity wave field of fixed shape but unknown amplitude E , the factorized GM76 spectrum (which may vary with E in space and time). The form of the energy balances is thus exact, but the group velocity c_0 (modulus) is that of a related GM spectrum. IDEMIX consequently assumes that the actual wave spectrum is always close to the GM spectrum with respect to the shape; the unknown energy is then given by E^\pm or E and ΔE and governed by (3.10).

Surface and bottom reflections lead to a flux from E^\pm to E^\mp , respectively, and forcing by tides and near-inertial pumping is added as a surface and bottom flux $\sim \Delta E$ into the total wave energy E , resulting from the integration of (3.8). Wave-wave interactions in the simple IDEMIX version are parameterized as damping of the differences of up- and downward propagating waves with a timescale of a few days, leading to the relaxation term F_{ww} in the asymmetry balance of (3.10). This simple closure is supported by the observation of a nearly symmetric wave field in m (as the GM spectrum) and the evaluation of the wave-wave interactions for

slightly perturbed GM spectra (McComas 1977). There is no corresponding term in the balance of total energy since wave–wave interactions conserve energy. The closure for the dissipation of gravity waves in (3.10) by the term $F_{diss} = \mu E^2$ follows the method of finestructure estimates of dissipation rates (Gregg 1989; Kunze and Smith 2004) and is given by a quadratic functional in the total wave energy E , as found by Olbers (1976), McComas and Müller (1981) from the scattering integral of resonant triad interactions. The parameter μ is a known function of N , f , and the GM bandwidth m_* .

The simple IDEMIX version can be extended to horizontal inhomogeneity conditions but cannot treat well lateral propagation of waves. This issue is resolved in the extended IDEMIX version by Eden and Olbers (2014), where low vertical mode energy compartments E_n at fixed frequency ω_n (e.g. tidal frequency) are separated from the rest (the wave continuum) and which resolve horizontal wave propagation. The E_n are accordingly governed by a corresponding radiative transfer equation

$$\partial_t E_n + \nabla \cdot \mathbf{c}_g E_n + \partial_\phi \dot{\phi} E_n = W_n + T_n \quad (3.11)$$

where \mathbf{c}_g denotes the lateral group velocity at ω_n , and $\dot{\phi}$ denotes the refraction of the wavenumber angle ϕ with $\mathbf{k} = k(\cos \phi, \sin \phi)$. The E_n are functions of \mathbf{x} and ϕ (and time) and are chosen as tides M_2 , S_2 or local near-inertial waves. They are represented by vertically integrated low vertical modes, while all other frequencies and higher vertical modes are still contained in the vertically resolved wave continuum E^+ and E^- .

Equation (3.11) describes the lateral propagation and refraction of low-mode baroclinic tidal or near-inertial energy compartments, the scattering into the wave continuum by rough topography by the term T_n , and the wave–wave interaction with the continuum by the term W_n , which were derived analytically by Eden and Olbers (2014). They show up with opposite signs in the conservation equation for $E = E^+ + E^-$. Tidal forcing enters in the extended version of IDEMIX partly as a bottom flux for the wave continuum as in the simple IDEMIX version, but also in the energy compartment of the respective tidal constituent where it will laterally propagate over considerable distances before it is transferred to the wave continuum and to dissipation. Both the simple and extended versions of IDEMIX are available as stand-alone versions with prescribed stratification and forcing without feedback on the circulation, and coupled to a general circulation model (<https://wiki.zmaw.de/ifm/TO/pyOM2>).

The link from the wave energy balance to mixing is as follows. Knowledge of the wave energy E enables us to compute the wave dissipation term $F_{diss} = \mu E^2$ which is a source of turbulent kinetic energy (TKE). Assuming steady state and a few other simplifications, the TKE balance reads (Osborn and Cox 1972)

$$F_{diss} + \overline{b'w'} - \varepsilon = 0 \quad (3.12)$$

Here, $\overline{b'w'}$ denotes the vertical turbulent buoyancy flux, i.e. the exchange with potential energy, and ε the dissipation rate of TKE by molecular friction, i.e. the exchange

with internal energy ('heat'). Assuming a conventional downgradient turbulent buoyancy flux $\overline{b'w'} = -K_\rho N^2$, where K_ρ denotes a vertical diffusivity, and a constant mixing efficiency $\delta = K_\rho N^2 / \varepsilon$ (usually taken equal 0.2), the diffusivity K_ρ can be computed from the wave energy E and given δ . More elaborate coupling of TKE and E is considered in Eden et al. (2014).

3.2.2 *The IDEMIX Concept Applied to Atmospheric Gravity Waves*

IDEMIX is not yet realized for atmospheric internal gravity waves. We describe here how an atmospheric IDEMIX can be built. Unlike the ocean case where parameterization of mixing is a first goal, the atmospheric IDEMIX should aim at parameterization of wave drag, i.e. the wave-induced Reynolds stress.

We restrict ourselves to the single-column approximation (i.e. we assume a horizontally homogeneous background flow at each geographical location, which is analogous to the plane-parallel approximation in radiative transfer computations) such that the coordinate dependence is reduced to that on height z . Furthermore, we assume gravity wave propagation only in particular azimuthal directions. In the actual gravity wave parameterization, we then add up the contributions from 4 or 8 equally distributed azimuths, denoted by the index j . Note that these approximations are made as well in any conventional gravity wave scheme used in climate models. Regarding the dependence on the wavenumber vector in horizontal and vertical directions, we express the spectral energy density with regard to the horizontal wavenumber k in terms of the intrinsic frequency,³ denoted as ω . The total spectral energy is given by $\mathcal{E}(m, \omega, \phi, z, t) = \sum_j \mathcal{E}_j(m, \omega, z, t) \delta(\phi - \phi_j)$. The radiative transfer equation

$$\partial_t \mathcal{E}_j + \partial_z (\dot{z} \mathcal{E}_j) + \partial_m (\dot{m} \mathcal{E}_j) + \partial_\omega (\dot{\omega} \mathcal{E}_j) = S_j + \dot{\omega} \mathcal{E}_j / \omega - D m^2 \mathcal{E}_j \quad (3.13)$$

for the wave energy compartment for the azimuth direction j is derived from (3.9) after integration over the azimuth angle. Our sign convention is such that $\omega > 0$ and $m < 0$ for upward group propagation (downward phase propagation), as before for the oceanic case. In (3.13), $D = D(z, t)$ denotes a vertical diffusion coefficient which describes damping due to wave breaking. This coefficient must be computed according to some dynamic stability criterion and should in principle also include the diffusion coefficient computed by the vertical diffusion scheme of the model (e.g. the boundary layer diffusion will damp orographic gravity waves). Furthermore, $S_j = S_j(m, \omega, z, t)$ is a source function that needs to be specified in order to describe the various generation mechanisms of atmospheric gravity waves such as flow over a rough surface, convection, frontal activity, and secondary gravity waves. Wave-wave

³Conventional parameterization schemes often assume a fixed k for the assumed gravity wave spectrum.

interactions would also add to S_j , but are not further discussed here since they play a less dominant role compared to the ocean. The other terms in (3.13) describe wave propagation, wave refraction, and interactions (reversible and irreversible) with the mean flow.

For the ocean case, energy compartments of up- and downward propagating waves are considered. This is important because of the nearly vertical symmetry of the wave field in the ocean, but less so for the atmosphere since surface reflection and strong wave–wave interactions, which lead to the symmetry, are missing. Reflection at the top is entirely absent. We will therefore dispense with this differentiation in compartments of up- and downward propagating waves for the atmosphere. We propose to use four directional compartments as a starting point. Conventional gravity wave schemes for atmosphere models often use 8 azimuthal directions.

3.2.2.1 Wave–Mean Flow Interaction and Energy Conservation

Regarding the effects of the parameterized gravity wave spectrum on the mean flow, we resort to the general theoretical framework of the two-way Reynolds average approach to filter out small-scale turbulence and mesoscale gravity waves as presented in Becker (2004) and Becker and McLandress (2009). A more detailed discussion can also be found in Shaw and Shepherd (2009). For a prognostic gravity wave scheme as envisioned in our case, the following framework applies. The rate of change of mean momentum and energy by the wave-induced stress is

$$(\rho \partial_t \mathbf{v})_{GW} = -\partial_z \mathbf{F} \quad (3.14)$$

$$(\rho c_p \partial_t T)_{GW} = -(\partial_z F_p + \mathbf{F} \cdot \partial_z \mathbf{v}) - \partial_t e_{GW} \quad (3.15)$$

Here, \mathbf{v} denotes the horizontal wind vector of the mean flow, and hence, the left-hand side of (3.14) simply describes the familiar gravity drag. The in situ air temperature is denoted by T , and the direct heating that accompanies the gravity wave drag is given by the right-hand side of (3.15), where the first term is the so-called energy deposition (e.g. Hines 1997). The energy deposition involves both the vertical flux of horizontal momentum, \mathbf{F} , and the pressure flux, F_p , induced by the wave field. The second term is the tendency of the total energy⁴ of the gravity wave field, denoted as e_{GW} . This energy

$$e_{GW}(z, t) = \sum_j \int_{-\infty}^{-m_0} dm \int_{\omega_0}^{\omega_1} d\omega \mathcal{E}_j(m, \omega, z, t) \quad (3.16)$$

is obtained by integrating (3.13) over wavenumber and frequency and by summation over all azimuths. Here, m_0 defines a minimum absolute vertical wavenumber, ω_0 a

⁴This direct heating term vanishes for a conventional parameterization due to the assumption of stationary wave energy.

minimum intrinsic frequency that is of the order of the Coriolis parameter at middle latitudes, and ω_1 a maximum intrinsic frequency that is of the order of the background buoyancy frequency N .

Energy conservation is trivially fulfilled since F_p and \mathbf{v} vanish at the surface, $z = z_s$, and since \mathbf{F} and F_p vanish for $z \rightarrow \infty$. Hence,

$$\int_{z_s}^{\infty} dz \partial_t e_{GW} + \int_{z_s}^{\infty} dz \mathbf{v} \cdot (\rho \partial_t \mathbf{v})_{GW} + \int_{z_s}^{\infty} dz (\rho c_p \partial_t T)_{GW} \equiv 0 \quad (3.17)$$

The momentum flux and pressure flux are obtained by using the polarization relations for gravity waves having vertical wavelengths not larger than the scale height, yielding

$$\mathbf{F}(z, t) = \sum_j F_j (\cos \phi_j \mathbf{e}_x + \sin \phi_j \mathbf{e}_y) \quad (3.18)$$

with

$$F_j(z, t) = \int_{-\infty}^{-m_0} dm \int_{\omega_0}^{\omega_1} d\omega \frac{\omega}{N} \mathcal{E}_j(m, \omega, z, t)$$

$$F_p(z, t) = \sum_j \int_{m_0}^{\infty} dm \int_{\omega_0}^{\omega_1} d\omega \frac{\omega}{m} \mathcal{E}_j(m, \omega, z, t) \quad (3.19)$$

Here, ϕ_j is the angle of the azimuthal direction j with the unit vector in eastward direction denoted as \mathbf{e}_x , while \mathbf{e}_y is the unit vector in northward direction. In these expressions, the mid-frequency approximation was made which can, however, be relaxed as for the oceanic IDEMIX version.

3.2.2.2 Factorization of the Spectrum and Prognostic Equations

Since the treatment of (3.13) in the 3-dimensional space (z, m, ω) will be computationally too expensive, we follow the IDEMIX concept and assume a simple factorization of the wave spectrum with unknown amplitude $E_j(z, t)$. Having $\mathcal{E}_j(z, t, m, \omega)$ in such a form, the integrals over m and ω for the fluxes (3.19) in the model equations (3.14) and (3.15) are readily calculated as functionals of $E_j(z, t)$. The assumption of a factorized wave spectrum with unknown amplitude is key to the IDEMIX concept which we also use here, but we also allow for variations of the shape parameters.

While a generic spectral shape is observed in the ocean with only rare exceptions, this is not expected for the atmosphere. We therefore attempt to introduce additional prognostic equations to characterize the spectral shape. This concept was originally

introduced by Hasselmann et al. (1973) for the simulation of the energetics of surface wind waves. In the atmospheric case, the generalized Desaubies spectrum according to Fritts and VanZandt (1993) has proven to work well in advanced gravity wave schemes using the conventional framework (Scinocca 2003). Here, we apply the Desaubies spectrum (Desaubies 1976) (this actually is a derivative of GM; see also Müller et al. 1978) for each azimuthal direction as:

$$\mathcal{E}_j(m, \omega, z, t) = E_j(z, t) n(\omega_0, m_j^*(z, t), q, r, s) \left(\frac{\omega_0}{\omega}\right)^q \frac{(m/m_j^*(z, t))^s}{1 + (m/m_j^*(z, t))^{r+s}} \quad (3.20)$$

Like in IDEMIX, $E_j(z, t)$ is the total energy contained in the spectrum at (z, t) ; i.e. n is a normalization factor such that $E_j = \int d\omega \int dm \mathcal{E}_j$. The shape of the spectrum regarding its m -dependence is allowed to vary with height and time via the parameter $m_j^*(z, t)$. The exponent q used in (3.20) can be chosen within $1 \leq q \leq 5/3$, while typical values for s and r are 1 and 3, respectively. As a starting point for the new parameterization, we set $q = 1$, $s = 1$, and $r = 3$. For large m , the spectrum is then proportional to m^{-3} . Such a behaviour is consistent with the scaling laws of stratified turbulence.⁵

To compute the temporal evolution of $E_j(z, t)$ and $m_j^*(z, t)$, we have to solve (3.13) for each azimuth. To this end, we have to compute \dot{z} and $\dot{\omega}$ from the dispersion relation. For the most simple case of mid-frequency gravity waves with upward group propagation (i.e. for $m < 0$), we get $\dot{z} = -\omega/m$, $\dot{m}_j = N^{-1}\omega (m \partial_z u_j + \partial_z N)$ and $\dot{\omega}_j = -\omega^2 N^{-1} \partial_z u_j$, where u_j is the projection of \mathbf{v} into the direction of ϕ_j . Furthermore, we have to specify the diffusion coefficient and some source functions. In IDEMIX, the resulting radiative transfer equation is then integrated over m and ω to obtain a prognostic equation for just the amplitude of the spectrum, i.e. for $E_j(z, t)$.

To allow, however, for variations of the spectral shape with altitude and time, we use another method to solve (3.13). We propose to apply the Gaussian variation principle in order to determine $\partial_t E_j(z, t)$ and $\partial_t m_j^*(z, t)$ from the functional derivatives

$$\frac{\delta \chi_j^2(z, t)}{\delta (\partial_t E_j(z, t))} = 0, \quad \frac{\delta \chi_j^2(z, t)}{\delta (\partial_t m_j^*(z, t))} = 0 \quad (3.21)$$

that are obtained from requiring that

$$\chi_j^2(z, t) = \int_{-\infty}^{-m_0} dm \int_{\omega_0}^{\omega_1} d\omega \left[\partial_t \mathcal{E}_j + \partial_z (\dot{z} \mathcal{E}) + \partial_m (\dot{m}_j \mathcal{E}) + \partial_\omega (\dot{\omega}_j \mathcal{E}) - S_j - \dot{\omega}_j \mathcal{E}_j / \omega + D m^2 \mathcal{E}_j \right]^2 \quad (3.22)$$

⁵The aspect ratio for stratified turbulence, $k \propto m^3$, converts a spectral energy density proportional to m^{-3} into a $k^{-5/3}$ -spectrum with respect to the horizontal wavenumber.

is minimum. The procedure is straightforward. All integrals in (3.22) for m and ω can be solved analytically.

3.2.2.3 Dissipation and Forcing

A particular and important aspect of the project is to set up a proper parameterization of the diffusion coefficient, $D(z, t)$, and the source function $S_j(m, \omega, z, t)$ to force gravity waves. Regarding the diffusion, we plan to rely on the saturation assumption of Lindzen (1981) which has proven to work equally well for non-orographic and orographic gravity waves (see the model description in Garcia et al. 2007). This has, however, to be further specified for the present entirely new framework for an atmospheric gravity wave scheme. Also note that, like in the generalized Doppler spread parameterization of Becker and McLandress (2009), the diffusion coefficient should be the same for the entire gravity wave field.

While wave–wave interactions are of importance in the oceanic case, this appears less so for the atmosphere. Wave–wave interactions will thus be neglected in the first phase. The dissipation in the oceanic case is inferred from the flux in wavenumber space due to wave–wave interactions to large vertical wavenumbers and is a quadratic functional of the total wave energy (Olbers 1976; Müller et al. 1986), which is also used for observational estimates of dissipation rates (Gregg 1989; Kunze and Smith 2004). It remains to be seen how this process relates to the aforementioned non-linear wave-breaking theories and whether it can eventually be incorporated in the new atmospheric gravity wave scheme.

Regarding the source function, we plan to use specifications for gravity wave generation by orography, frontal activity, and convection that have proven to work well in conventional schemes used in global climate models. We will take advantage of the specification of orographic sources as proposed by McFarlane (1987) and use gravity wave sources due to frontal activity and convection as outlined by Charron and Manzini (2002) and Richter et al. (2010).

3.3 Oceanic Processes in Present and Future IDEMIX

Internal gravity waves have a major share of the energy contained in oceanic motions (e.g. Wunsch and Ferrari 2004). When they break by shear or gravitational instability, they feed their energy to small-scale turbulence in the interior of the ocean and thus contribute to density mixing. This process, transferring wave energy to large-scale potential energy, becomes an important part of the oceanic energy cycle. It is furthermore thought to be an important driver of the ocean circulation (e.g. Wunsch and Ferrari 2004; Kunze and Smith 2004).

The part of the gravity wave field which is prone to break is by far unresolved by ocean model components of climate models, since a lateral and vertical resolution of less than 100 and 10 m, respectively, is required. Consistent parameterizations to

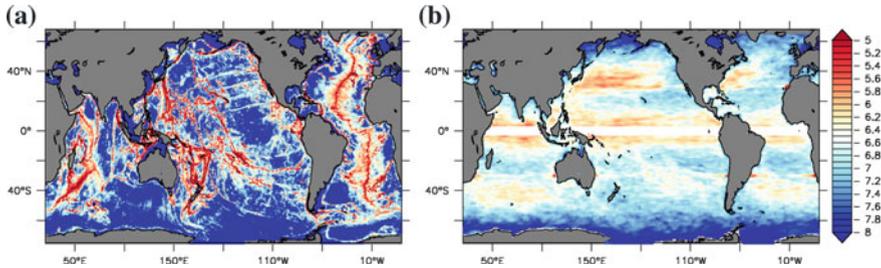


Fig. 3.4 (a) Tidal forcing and (b) near-inertial wave forcing in $\log_{10} F/[\text{m}^3 \text{s}^{-3}]$

include the effects of wave breaking have been missing for a long time. A varying diffusivity was introduced for the ocean interior by Bryan and Lewis (1979) with a prescribed depth function and, among others, by Cummins et al. (1990) with a dependence on the stability frequency. The energy source for mixing was not specified in these and other early approaches although the vertical diffusivity was supposed to parameterize the effect of breaking internal waves. Jayne and St. Laurent (2001) suggested to link the conversion of barotropic tidal energy into internal waves as simulated by a barotropic tidal model to a profile of vertical diffusivity adapted to observations. However, the lateral spreading of baroclinic tides and the effect of the other sources of gravity waves are left unconsidered, and the parameterization remains energetically inconsistent. The recently developed IDEMIX concept (Olbers and Eden 2013; Eden and Olbers 2014), on the other hand, implements this spreading and treats gravity wave sources in an energetically consistent way. For details, see Section 3.2.1.

Prominent forcing mechanisms of internal waves occur very localized in the frequency domain. Near-inertial waves with frequencies slightly above the local Coriolis frequency are excited by wind stress fluctuations at the surface (Alford 2001; Rimac et al. 2013) and can propagate over large distances in horizontal direction, while slowly propagating down into the interior ocean (Garrett 2001). A further monochromatic source is related to the scattering of the barotropic tide at topography, predominantly at the continental shelf, the mid-ocean ridges but also at the more random-type small-scale roughness of the seafloor (Nycander 2005). The flux from the barotropic tide into the internal wave field occurs at the ocean bottom, shown in Figure 3.4(a). The flux from radiation of near-inertial waves out of the surface mixed layer is depicted in Figure 3.4(b). These fluxes are currently the main drivers implemented in the IDEMIX versions.

Several other generation processes have been discussed (see, e.g., Thorpe 1975; Müller and Olbers 1975; Olbers 1983, 1986; Polzin and Lvov 2011) occurring over a broad range of frequencies, e.g. dissipation of mesoscale eddies by spontaneous wave emission or other processes (Ford et al. 2000; Molemaker et al. 2010; Tandon and Garrett 1996; Eden and Greatbatch 2008a; Brüggemann and Eden 2015), resonantly interacting surface gravity waves (Olbers and Herterich 1979; Olbers and Eden 2016), the generation of lee waves by large-scale currents or mesoscale eddies flowing over

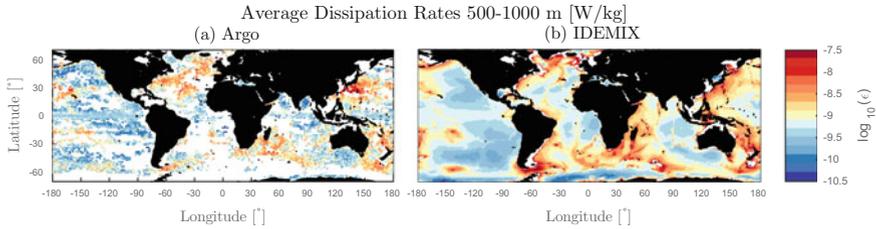


Fig. 3.5 Validation of the simple IDEMIX version: (a) observational estimate of the dissipation of gravity wave energy calculated from density profiles of ARGO floats following Whalen et al. (2012) averaged between 500 and 1000 m. (b) Same as (a) but simulated by IDEMIX in a fully coupled mode. From Pollmann et al. (2017)

topography (Nikurashin and Ferrari 2011), and wave–mean flow interaction (Müller 1976; Polzin 2008). There are also indications that mixed-layer turbulence generates waves close to the local stability frequency below the mixed layer (Bell 1978).

Both IDEMIX versions show agreement with observational estimates in first diagnostics, but also biases: Figure 3.5 shows that magnitude and lateral pattern of the simulated dissipation rates of internal waves in the simple IDEMIX version agree with observational estimates (more details in Section 3.3.5). It turns out that the dissipation of mesoscale eddy energy is important for the maxima in dissipation rates seen in the western boundary currents and the Antarctic Circumpolar Current (ACC). Without the eddy dissipation, some of these maxima are not simulated (not shown). In the Southern Ocean, however, IDEMIX simulates too much dissipation within the ACC and thus too large diffusivities (as shown in Eden et al. 2014), when all eddy energy is dissipated locally and injected into the wave field as assumed in the mesoscale eddy closure by Eden and Greatbatch (2008b). This points towards the need to better understand the dissipation of mesoscale eddy energy and its relation to the internal wave field. More details are given in Section 3.3.1.

When waves are propagating in a vertically sheared mean flow, they exchange energy with the mean flow and even can break due to critical layer absorption or wave capture. The former effect is also called gravity wave drag in the atmospheric literature (see Section 3.2.2.1), where it is of importance for the dynamics of the upper atmosphere. The direction of the energy exchange can be from the mean flow to the waves or vice versa. When waves break in a critical layer, on the other hand, their energy is transferred to small-scale turbulence. In Section 3.2.2.1, we propose an extension of IDEMIX to incorporate the energy exchange with the mean flow for the atmospheric case. The effect on the mean flow figures in this concept as a divergence of a vertical eddy momentum flux due to wave activity in the residual momentum equation, similar to vertical friction, but can accelerate and decelerate the mean flow. It can be shown that the energy transfer between waves and large-scale mean flow is present in the ocean, but amounts to only a fraction of, e.g. the energy exchange between mean flow and mesoscale eddies (see Section 3.3.2). However, the effect of critical layers where waves break and contribute to density mixing has

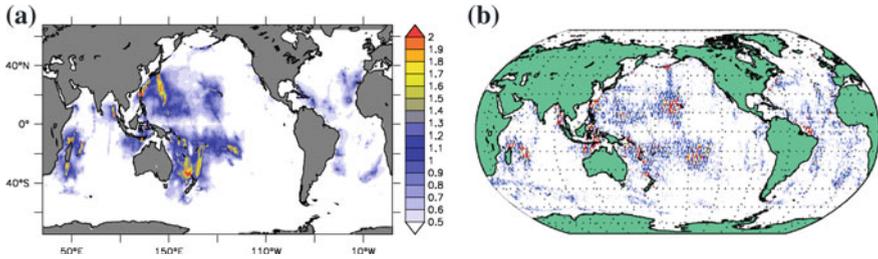


Fig. 3.6 Validation of the extended IDEMIX version: (a) equivalent surface elevation ζ due to the baroclinic M_2 tide in cm simulated by IDEMIX. (b) Observed surface elevation ζ in m of M_2 tide taken from Müller et al. (2012) on the same colour scale as in (a). Taken from Eden and Olbers (2014)

not been discussed so far in IDEMIX, although this might be the more important effect in the ocean. Effects similar to critical layers occur when horizontally sheared mean currents are present; the process is called wave capture (see, e.g., Jones 1969; Bühler and McIntyre 2005). This points towards the need to include wave–mean flow interaction and the effect of critical layers and wave capture into IDEMIX. We expand this issue further in Section 3.3.2.

Figure 3.6 shows that the extended version of IDEMIX simulates well the generation of the low-mode M_2 tide and also its propagation but that there are also biases. It remains at the moment unclear whether these differences are due to errors in the observational estimates—the identification of the baroclinic tidal signal in the altimeter data is rather difficult (Dushaw et al. 2011)—but it is clear that IDEMIX also has shortcomings. Most important is the forcing by the barotropic tide, which is taken at the moment for simplicity as isotropic for the wave propagation direction. On the other hand, anisotropic wave generation is most likely responsible for many features seen in the observational estimates, such as the energy maximum between the Aleutian and the Hawaiian Islands which is not reproduced by IDEMIX. This points towards the need to include anisotropic wave generation in IDEMIX and a detailed comparison with observations and direct simulations of baroclinic tides in ocean models. Details are found in Section 3.3.3.

Unlike the internal wave energy generated by near-inertial motions in the mixed layer and the tides which have low frequencies and propagate through the entire water column, internal waves generated by resonant interactions of surface waves are of high frequency (Olbers and Herterich 1979; Olbers and Eden 2016). The same is true for the wave generation by mixed-layer turbulence near the stability frequency (Bell 1978; Polton et al. 2008). Since the stability frequency at larger depth is normally much smaller than within the seasonal thermocline close to the mixed layer or the upper permanent thermocline, waves generated by these processes have shallow turning points and are thus likely to be trapped in the upper ocean. They could contribute to mixing just below the mixed layer. Recent estimates of the global energy transfer from the surface waves to the internal wave field are given by Olbers

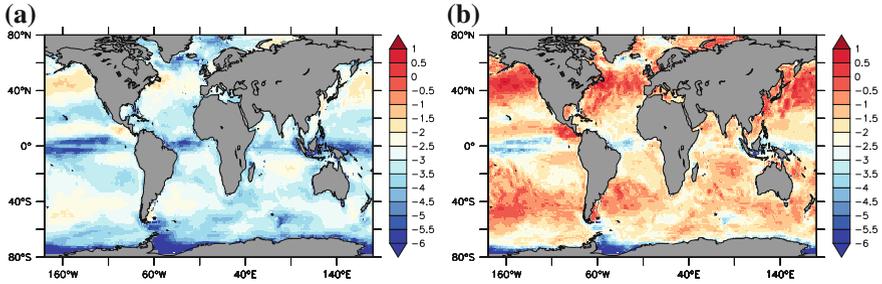


Fig. 3.7 (a) Mean flux from surface waves to internal waves in $\log_{10}(\Phi_{tot}/[\text{mW}/\text{m}^2])$ in 2010. (b) Same as (a) but maximum of the year. From Olbers and Eden (2016)

and Eden (2016). Figure 3.7 shows the annual mean total flux and its maximum during the year. The largest fluxes show up in the storm track regions of the oceans, while towards the equator the flux and its maximum almost vanish. The implied dissipation rates are found to reach magnitudes comparable to observational estimates close to the mixed layer, in particular during strong wind events. This points towards the need to include in IDEMIX surface–internal wave interactions and waves generated by mixed-layer turbulence. See Section 3.3.4 for more details.

3.3.1 Including Energy Transfers from Mesoscale Eddies to Internal Waves

Dissipation of balanced flow is thought to happen on different routes; lee wave generation, bottom friction, and loss of balance are often considered as important processes. Other processes which have been discussed are topographic inviscid dissipation of balanced flow (Dewar and Hogg 2010), direct generation of unbalanced ageostrophic instabilities (Molemaker et al. 2005), or geostrophic adjustment of balanced flow (Wang and Zhang 2010). Further, wave–mean flow interaction can lead to an energy transfer between the wave field and the vertically sheared mesoscale and large-scale mean flow, a process called gravity wave drag in the atmospheric literature. This ‘drag’, however, can take both directions.

3.3.1.1 Lee Wave Generation

Scott et al. (2011) estimate an energy flux of 0.34 to 0.49 TW as the global transfer of energy by internal lee wave generation of eddies and mean flow over varying topography, while Nikurashin and Ferrari (2011) estimate a lower energy flux of 0.2 TW by the process. However, both numbers are only a fraction of the total eddy energy production: von Storch et al. (2012) report 0.83 TW global eddy energy

production by conversion from mean to eddy available potential energy (baroclinic instability) and additionally a transfer of 0.1 TW from mean to eddy kinetic energy (barotropic instability) in agreement with observational estimates.

Linear theory (Bell 1975) shows that the energy flux into the internal wave field for subcritical topographic slopes is given by

$$F_{lee} = \frac{\rho_0}{4\pi^2} \int_{f^2 < (\mathbf{U} \cdot \mathbf{k})^2 < N^2} d^2k P(\mathbf{k}) \mathbf{U} \cdot \mathbf{k} / |\mathbf{k}| \sqrt{N^2 - (\mathbf{U} \cdot \mathbf{k})^2} \sqrt{(\mathbf{U} \cdot \mathbf{k})^2 - f^2} \quad (3.23)$$

where $P(\mathbf{k})$ is the topography spectrum, \mathbf{U} the balanced bottom flow, N is the stability frequency at the bottom, and \mathbf{k} the horizontal wave vector. We will implement the effect of lee wave generation using for (3.23) estimates of topographic spectra from Goff and Arbic (2010) and magnitudes of balanced bottom flow from the model of von Storch et al. (2012). To describe the topography spectrum, we use (as in Eden and Olbers 2014) digital maps of geophysical parameters given by Goff and Arbic (2010) for the root mean square topographic height h_{rms} of abyssal hills and $\bar{k} = 2\pi/\lambda_n$, where λ_n is a characteristic width and \bar{k} a mean wavenumber. These maps are based on estimates of palaeo-spreading rates of the mid-oceanic ridge system and sediment thickness. Specific formulae for h_{rms} and λ_n and data sources are given in Eden and Olbers (2014). Assuming an isotropic spectrum and a fixed power law for $P(k) = P_0 k^{1-\mu}$, it is possible to evaluate the integral in (3.23) for $|f| \ll N$ analytically as

$$F_{lee} \approx \frac{\rho_0 P_0}{\pi} N |\mathbf{U}|^{5/2} f^{-1/2} \left(\frac{9}{5} - \frac{7}{3} |f/N|^{1/2} + O(f^2/N^2) \right) \quad (3.24)$$

written here for simplicity for the parameter $\mu = 7/2$, as suggested by Nikurashin and Ferrari (2011). Note that P_0 relates to h_{rms}^2 , if evaluated for the characteristic range of lee waves $P_0 = 4h_{rms}^2/\lambda_n \sqrt{|U|/f}$. We will compute the flux F_{lee} from the bottom flow \mathbf{U} of the global eddy-resolving STORM model (von Storch et al. 2012), which provides a map of energy injection due to lee wave generation as lower boundary condition for a stand-alone version of IDEMIX (in addition to the tidal forcing). For IDEMIX coupled to a non-eddy-resolving general circulation model, we will use the mean flow from that model, and in addition the (square root of) EKE from the closure by Eden and Greatbatch (2008b) for \mathbf{U} in (3.24).

3.3.1.2 Bottom Friction

The transfer by bottom friction is likely of similar order of magnitude as the one by lee wave generation: the dissipation of eddy energy by bottom friction was estimated by Arbic et al. (2009) from different model bottom velocities as a global energy transfer ranging between 0.14 and 0.65 TW by a simple quadratic drag law. In the model by von Storch et al. (2012), the global transfer is only 0.12 TW (Brüggemann and Eden

2015). Here, we will assume that bottom friction generates small-scale turbulence in a bottom boundary layer only, without energy transfer to internal waves.

3.3.1.3 Loss of Balance

Williams et al. (2008) estimate 1.5 TW energy extracted from quasi-geostrophic flow and transferred into internal gravity waves by Lighthill radiation (Lighthill 1952) based on a laboratory experiment extrapolated to the global ocean. Although their extrapolation is presumably quite crude, the experiment suggests that dissipation by Lighthill radiation might play an important role in the ocean energy cycle. Brüggemann and Eden (2015) find in idealized model experiments of forced-dissipative baroclinic instability that a kinetic energy cascade towards smaller scales is generated for dynamical conditions characterized by a low Richardson number (Ri) or large Rossby number (Ro), i.e. for ageostrophic conditions, while the familiar inverse energy cascade towards larger scales (and to bottom friction) dominates for $Ri \gg 1$ or $Ro \ll 1$. From a simple fit of the energy transfer towards smaller scales as a function of the local Ri , Brüggemann and Eden (2015) estimate the fraction of the local baroclinic eddy production rate which would go into the forward cascade in the model by von Storch et al. (2012) (Figure 3.8). The global transfer by this process amounts to 0.31 ± 0.23 TW, i.e. only a fraction of the total production but comparable to the generation by lee waves. The flux by the forward cascade is not only large in the surface mixed layer where the stratification vanishes, or near the equator where geostrophic balance becomes a weaker constraint, but also at high latitudes in western boundary currents and the Southern Ocean, where the vertical shear becomes large.

We will assume that this energy transfer by the forward energy cascade at large Ri in the interior of the ocean generates gravity waves, while the remainder of the total eddy energy production is assumed to be dissipated by bottom friction. We will thus provide the flux by the forward cascade as an interior source to IDEMIX. The flux in the surface mixed layer will be assumed to contribute to small-scale

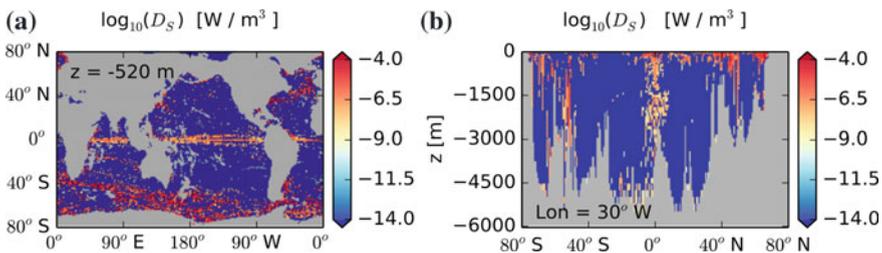


Fig. 3.8 Horizontal and meridional sections of the implied energy transfer D_s due to the forward energy cascade in the high-resolution ocean model by von Storch et al. (2012). Taken from Brüggemann and Eden (2015)

turbulence there, while the flux near the equator within 5° latitude will be disregarded since here the inference from the simulations of mid-latitude baroclinic instability by Brüggemann and Eden (2015) might not hold. To include the interior flux in a version of IDEMIX coupled to a general circulation model, the dissipation of EKE implied by the mesoscale eddy closure by Eden and Greatbatch (2008b) will be used. Other proposed sources of eddy dissipation will be neglected in the first phase of this project, but will be considered in later phases if necessary.

Both, the bottom flux due to lee wave generation and the interior flux due to the forward energy cascade will be prescribed in a stand-alone version of the simple version of IDEMIX. Experiments with tidal and surface forcing with and without the flux due to the eddy dissipation will be compared with observational estimates of dissipation rates as in Figure 3.5 to identify and to validate the effect of eddy dissipation.

3.3.2 Including Wave–Mean Flow Interaction

Waves propagating in a vertically sheared mean flow exchange energy with the mean flow and can even break when they hit a critical layer or run into a wave capture domain. In the presence of a mean shear, the radiative transfer equation (3.9) for the energy spectrum correspondingly contains an exchange term with the mean flow (last term on the right-hand side). This exchange depends on the magnitude of the shear, internal wave energy, the direction of the wave propagation, and other wave properties. Assuming as before a fixed (but locally varying) form of the spectrum of unknown amplitude, it becomes possible to evaluate the effect of the wave–mean flow energy exchange for the energy compartments E^\pm integrated in wavenumber space, as shown in Olbers and Eden (2017), Eden and Olbers (2017). An extension of IDEMIX including the wave–mean flow energy exchange requires only the split of E^\pm or equivalently E and ΔE into energy components which propagate in four different lateral directions, while all other parameters and closures remain identical to before. The corresponding IDEMIX model becomes

$$\begin{aligned}
 \partial_t E_e + \partial_z(c_0 \Delta E_e) &= -\tilde{U}_z \Delta E_e - \mu E_e (E_e + E_w) \\
 \partial_t \Delta E_e + \partial_z(c_0 E_e) &= -\tilde{U}_z E_e - \Delta E_e / \tau_v \\
 \partial_t E_w + \partial_z(c_0 \Delta E_w) &= \tilde{U}_z \Delta E_w - \mu E_w (E_e + E_w) \\
 \partial_t \Delta E_w + \partial_z(c_0 E_w) &= \tilde{U}_z E_w - \Delta E_w / \tau_v
 \end{aligned}
 \tag{3.25}$$

for eastward ($E_e, \Delta E_e$) and westward propagating waves ($E_w, \Delta E_w$), with $\tilde{U}_z = (\sqrt{8}/\pi)\Lambda \partial_z U$, and similar for northward ($E_n, \Delta E_n$) and southward propagating waves ($E_s, \Delta E_s$), except that $\partial_z U$ is replaced with $\partial_z V$. The wave–mean flow interaction terms are those involving the mean shear. They derive from the last term in the radiation balance (3.9). The interaction coefficient Λ is computed in IDEMIX manner using the GM spectral shape (Olbers and Eden 2017).

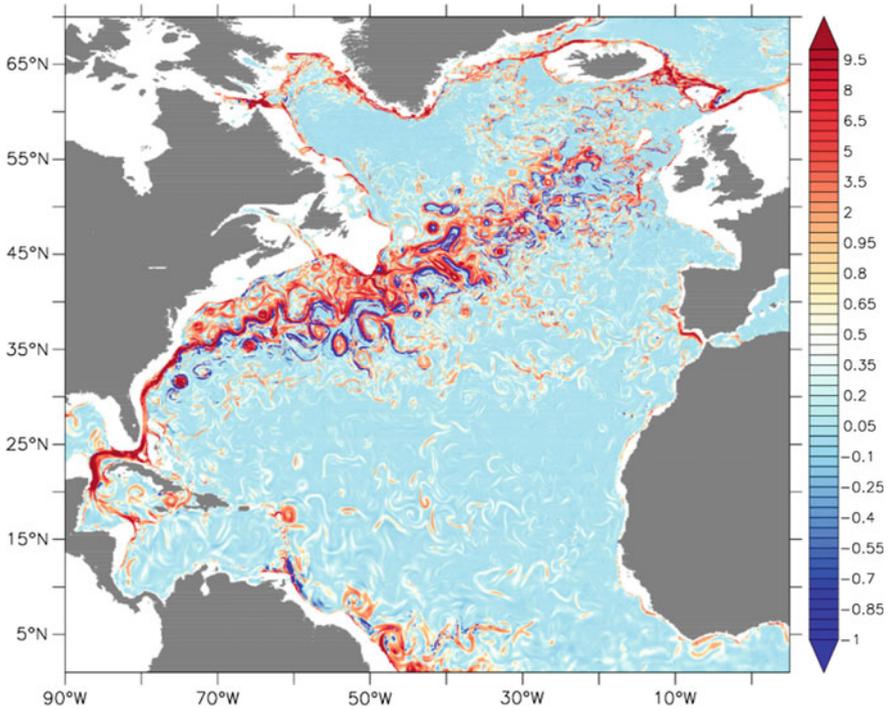


Fig. 3.9 Instantaneous energy transfer from mean flow to waves in 10^{-6} W/m³ at 300 m depth in a high-resolution model of the North Atlantic Ocean. From Eden and Olbers (2017)

The effect on the mean flow figures then as a divergence of a vertical eddy momentum flux due to wave activity in the residual momentum equation, similar to vertical friction, although the direction of the energy exchange can be from the mean flow to the waves but also vice versa, as demonstrated in Figure 3.9. The figure shows the energy transfer from the mean flow to the waves due to wave drag at 300 m depth of a simulation with a realistic, eddying North Atlantic Ocean model for a snapshot in September. The energy transfer due to the wave drag is significant for the kinetic energy balance of the mean flow. At 300 m depth, the horizontally integrated energy transfer from the mean flow to the waves is 19.9×10^6 W/m, while it is 149.8×10^6 W/m for the dissipation due to lateral biharmonic friction. This ratio of 10 to 20% can also be found at other depths. The energy transfer from mean flow to small-scale turbulence by the parameterized vertical friction at 300 m depth amounts to 56.9×10^6 W/m but is only significant within the mixed layer, and very low below; i.e. in the interior of the model, only lateral friction and the wave drag act as dissipation of mean kinetic energy. The depth-integrated values of the transfer due to lateral friction, wave drag, and vertical friction are 0.224×10^{12} W, 0.018×10^{12} W, and 0.115×10^{12} W, respectively. For the balance of internal wave energy itself,

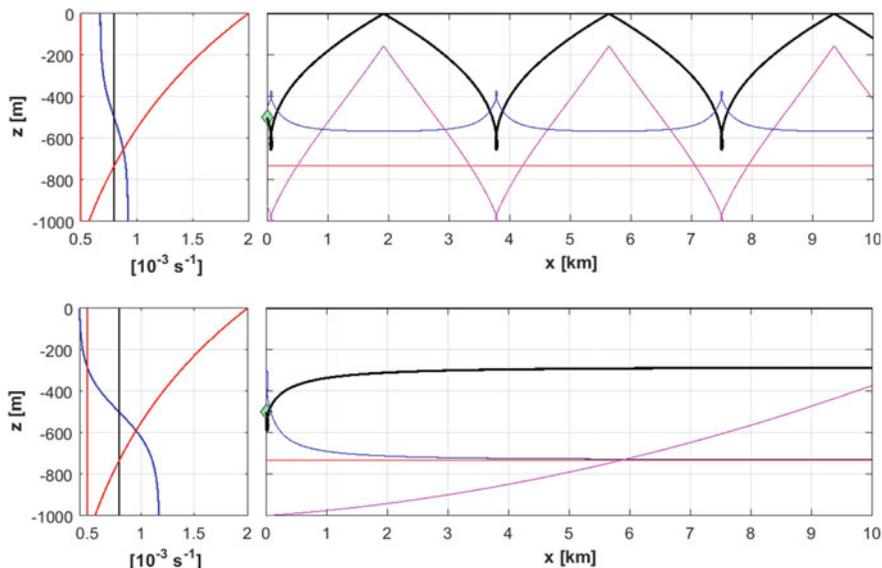


Fig. 3.10 Rays for propagation in a vertical shear flow. Upper row: weak mean shear, $U_0 = 0.1$, lower row: strong mean shear, $U_0 = 0.3$. Left: f and N (red), ω_{enc} (black), $\omega(z) = \omega_{enc} - kU(z)$ (blue). Right: ray (black), energy $E(t)$ (blue), minimum available energy $fE(t_0)/\omega(t_0)$ (red), shear $E(t)m^2(t)$ (magenta). Energy and shear have an unspecified scale. The horizontal wavelength is 5000 m. Other parameters are $f = 5 \times 10^{-4} \text{ s}^{-1}$, $N_0 = 2 \times 10^{-3} \text{ s}^{-1}$

the energy transfer due to the wave drag is less important, i.e. on the order of 5%, since the tidal forcing amounts to more than $0.4 \times 10^{12} \text{ W}$ in the North Atlantic.

The effect of critical layers and wave capture, where waves break and contribute to density mixing, has not been discussed so far. The conditions for a generic case of a critical layer is $N = \text{const}$ and a purely vertical shear flow $\mathbf{U} = \mathbf{U}(z)$ so that the vertical wavenumber changes according to

$$\dot{m} = -\mathbf{k} \cdot \partial_z \mathbf{U} \quad (3.26)$$

such that $|m|$ increases in time (linearly as \mathbf{k} remains constant) and the frequency ω of the wave ultimately approaches f where the intrinsic group velocity goes to zero and the intrinsic propagation comes to a halt. At the same time the vertical wavelength diminishes, the wave is prone to break (in fact, as action is conserved, $E(t)/\omega(t) = \text{const}$, but the shear $E(t)m^2(t)$ tends to infinity). We demonstrate the behaviour in an exponentially increasing Brunt-Väisälä frequency and a tanh-shaped mean flow in Figure 3.10 with different current amplitudes. The rays start at a depth of 500 m (at the zero of the mean current) and a downward group velocity. With a weak mean flow (upper panels), the shear does not yet allow a critical layer to appear. Note that the rays' cycle, energy, and shear are completely periodic. The situation changes dramatically (lower panels) when the shear is increased to an amplitude such that

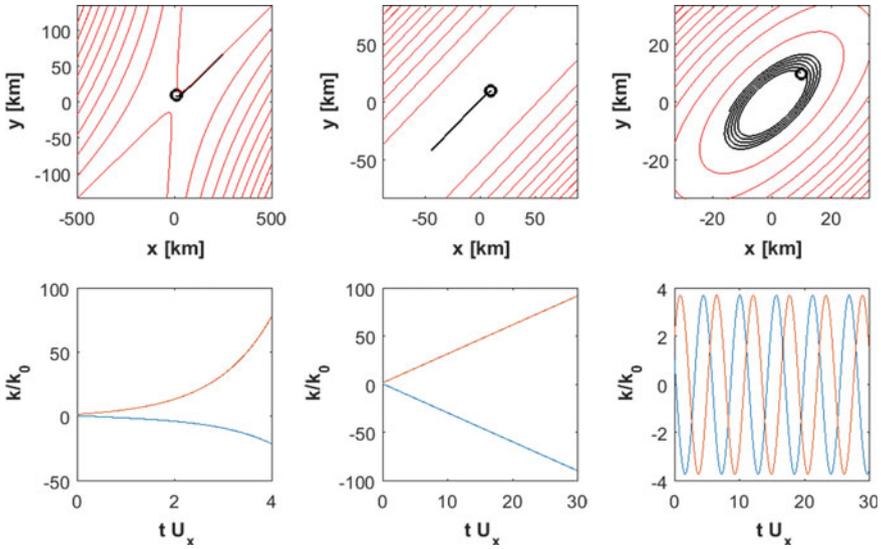


Fig. 3.11 Generic cases of wave capture, after Bühler and McIntyre (2005). The streamfunction of the horizontal flow is $\psi = -axy + c(x^2 + y^2)/2$ with $U = -\psi_y = ax - cy$, $V = \psi_x = -ay + cx$. Hence, $V_x + U_y = 0$, $U_x + V_y = 0$, $\beta = a^2 - c^2$. Left: $\beta > 0$, hyperbolic case for $c = a/2$. Middle: $\beta = 0$, parabolic case. Right: $\beta < 0$, elliptic case for $c = 3a/2$. Upper panels: streamfunction (red) and ray (black). Lower panels: k_i/k_0 , $i = 1, 2$ as function of time

at a certain level the intrinsic frequency can approach the local Coriolis frequency, i.e. $\omega(z) = \omega_{enc} - \mathbf{k} \cdot \mathbf{U}(z) \rightarrow f$. At such a level, the vertical group velocity goes to zero while the energy approaches the minimum value, however residing in ever decreasing vertical scales. The wave must eventually break.

A similar behaviour occurs in a flow with horizontal and vertical shear, $\mathbf{U} = \mathbf{U}(\mathbf{x}, z)$ where the horizontal wave vector increases exponentially

$$\dot{\mathbf{k}} = -\mathbf{k} \cdot \partial_{\mathbf{x}} \mathbf{U} \quad (3.27)$$

under certain conditions on the mean flow gradient. Equation (3.26) still holds such that $|m|$ also increases exponentially. This is the generic case of wave capture. If $U_z \sim U/H$, $U_x \sim U/L$, then the aspect ratio of the wavenumbers approaches $m/k \sim L/H$ and the ratio of wavelength (vertical to horizontal) approaches H/L . The frequency of the wave in this state goes to $\omega^2(t) \rightarrow f^2 + N^2(H/L)^2$ which is close to f^2 (Bühler and McIntyre 2005). The group velocity is exponentially decreasing, and the wave will be captured (trapped) in the mean flow and break. Note, however, that the captured wave is still moving with the mean current. In Figure 3.11, we show wave capture for three generic cases of mean flow: a hyperbolic flow with a saddle point (the wavenumber increases exponentially), a parabolic case of parallel currents (the wavenumber increases linearly), and an elliptic case with closed streamlines (the

wavenumber shows a cyclic behaviour). In the first two cases, the wave is prone to breaking.

The critical layer process will be implemented by calculating the fraction of the vertical (up or down) flux of wave energy at any level that will eventually be absorbed in critical layers (at shallower or deeper levels) and contribute to mixing there. As for the other closures in IDEMIX, the assumption of a fixed spectral shape for the local wave energy will be used. The flux portion which is absorbed depends on the mean flow profile, on f/N and the total wave energy at the respective level. We will convert this flux portion directly into dissipation over the respective water column, obtaining a parameterization of the mixing induced by critical layer absorption. We attempt a similar treatment for wave capture.

3.3.3 Including Anisotropic Tidal Forcing

The linear theory by Bell (1975) can also be used to generate maps of energy transfers from the barotropic tide to internal waves, similar to the lee wave generation of a mean flow in (3.23). Two effects can lead to anisotropy in the wave generation: anisotropy in the barotropic flow (tidal ellipse) and anisotropy in the topography. A realistic estimate of the barotropic tidal velocity field is provided by the TPXO.7 model for eight tidal constituents (M_2 , S_2 , etc.) with a resolution of $1/4^\circ$ (Egbert et al. 2010). Anisotropic abyssal hill spectra are provided by the digital maps given by Goff and Arbic (2010), similar as used for (3.23), but to be implemented here in its anisotropic form on a 2 min horizontal grid. However, it is clear that also topographic variations with scales larger than about 10 km, not covered by the data set by Goff and Arbic (2010), will contribute to anisotropic baroclinic tide generation. Following Nycander (2005), we will also use the latest satellite-derived topography by Smith and Sandwell (1997) on a 2 min grid to extend the spectra by Goff and Arbic (2010) to smaller wavenumbers taken from the observed topography. Local two-dimensional spectral estimates of subregions of $X \text{ km} \times X \text{ km}$ size from the satellite-derived topography will be averaged and blended with the anisotropic spectrum by Goff and Arbic (2010), averaged over the same region. X and Y will be of order 50 to 100 km but varied to obtain the sensitivity on these parameters.

Based on this estimate of the topography spectrum and using the tidal velocity averaged over the same region for each tidal constituent, linear theory predicts the energy transfer to the baroclinic tides. We will use the spectral tapering method by Nycander (2005) to account for finite depth effects in the linear theory by Bell (1975). The flux Φ_{tide} into the barotropic tide of frequency ω_n and velocity amplitude $|\mathbf{U}|$ for an isotropic tidal ellipse can be written as

$$\Phi_{tide}(\mathbf{k}) = \frac{1}{2} \rho_0 \frac{|\mathbf{U}|^2 N}{\omega_n \pi k_n k_s} \frac{h_{rms}^2 (\mu/2 - 1) (\omega_n^2 - f^2)^{1/2} k}{(1 + k^2/k_s^2 \cos^2(\theta - \theta_s) + k^2/k_n^2 \sin^2(\theta - \theta_s))^{\mu/2}} \tag{3.28}$$

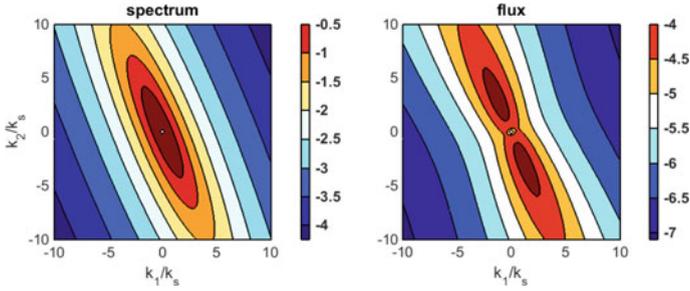


Fig. 3.12 Example of an anisotropic tidal forcing (left) due to anisotropic topographic spectrum (right) given by the form of Goff and Arbic (2010). Both are shown as normalized functions of $\mathbf{k} = (k_1, k_2)$ on a logarithm scale

as density in wavenumber space, where h_{rms} and μ are defined above in Section 3.3.1.1 and where k_n , k_s , and θ_s define the anisotropy of the spectrum by Goff and Arbic (2010). The case with an anisotropic tidal ellipse is analogous but more involved. The resulting flux Φ_{tide} is shown in Figure 3.12 as function of wavenumber vector \mathbf{k} for an artificial but typical topography spectrum based on the parameters given by Goff and Arbic (2010). The ϕ -dependency of the forcing will generate a baroclinic tide propagating predominantly in the direction anticipated from the figure, which is clearly different from the isotropic flux used before by Eden and Olbers (2014), and we expect a corresponding effect of the improved forcing function for IDEMIX.

We will use Φ_{tide} as interior forcing in (3.11) for the corresponding energy compartment in the extended version of IDEMIX. Expressing the \mathbf{k} -dependency of Φ_{tide} in terms of vertical wavenumber m or vertical mode number, using the dispersion relation, it becomes also possible to determine from (3.28) the amount of energy transferred to the low modes and the amount transferred to the wave continuum. This will be an improvement of Eden and Olbers (2014) who simply transfer 50% to the continuum and 50% to the low modes. The flux Φ_{tide} from (3.28) will be included in a stand-alone extended version of IDEMIX, and the results will be compared with observations.

3.3.4 Including High-Frequency Compartments

Waves forced at the base of the mixed layer by resonant surface–internal wave interaction (Olbers and Herterich 1979; Olbers and Eden 2016), mixed-layer turbulence (Bell 1978), and convective pumping (Polton et al. 2008) are of high frequency and thus trapped close to the surface. Their energy will thus predominantly contribute to near-surface mixing. Similar to the low-mode tidal and near-inertial energy compartments of the extended IDEMIX version by Eden and Olbers (2014), we

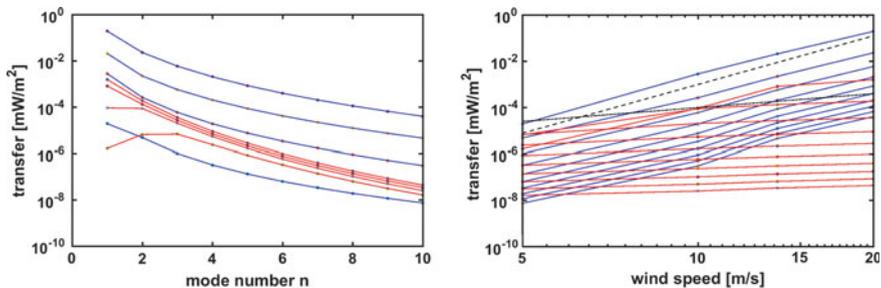


Fig. 3.13 Transfer for SC (blue) and MI (red) mechanisms as function of mode number (left) and wind speed (right). The respective lowest curves are for wind speed 5 m/s or mode 10, and the uppermost is for wind speed 20 m/s and mode 1. The curves are for $N_0 = 0.0157\text{s}^{-1}$. The black dashed lines (right panel) display a U^2 and U^7 dependence. From Olbers and Eden (2016)

will split energy compartments of high-frequency and high vertical modes from the wave continuum and will treat them separately from the spectral continuum. This energy compartment couples to the wave continuum by induced diffusion, a process of wave–wave interactions (McComas 1977), across the boundaries in frequency domain and to the low vertical mode near-inertial waves. The waves described by this compartment are thought to break predominantly in the upper ocean and are thus of central importance for dissipation in and above the main thermocline.

Resonant triad interactions between surface gravity waves (SW, with frequencies $\omega_j = \sqrt{gk_j}$, $j = 1, 2$) and internal gravity waves (IW, with frequency ω) are hampered by the extreme difference of their frequencies in the resonance condition $\omega_1 - \omega_2 - \omega = 0$. The scattering cross section is of order $(\omega^2/gk_j)^2 \ll 1$. Two processes compete: spontaneous creation (SC) in which two SW generate an IW and modulation interaction (MI) where a preexisting internal wave is modulated by a surface wave and interacts with another one (Watson 1990; Olbers and Eden 2016). SC is always a source of IW energy, and MI might establish a damping of the internal wave field, thus acting against SC. Interesting is the dependence of the respective transfer rates on the wind speed U (via the spectral energy content of the SW spectrum). For MI, the rate varies as U^2 , for SC however as U^7 , so that the dominant process at low wind speeds (roughly $U < 10$ m/s) is MI, while at stronger winds the interactions between surface and internal gravity waves are always controlled by SC (see Figure 3.13).

Figure 3.7 shows the annual mean total SC flux Φ_{tot} and its maximum during the year entering the internal wave field diagnosed from a simulation with an ocean model which contains a mixed-layer parameterization after Gaspar et al. (1990). It is driven by wind stress forcing; this and also U are taken from the year 2010 of the reanalysis by Kalnay et al. (1996). The total SC flux is computed from a parameterization of the SC scattering integral, as explained in Olbers and Eden (2016), depending on the simulated mixed-layer depth, the Brunt-Väisälä frequency at the mixed-layer base, and the wind speed U . The largest fluxes show up in the storm track regions of the oceans, while towards the equator the flux and its maximum almost vanish.

The flux varies a lot in space and time. In the global integral, it is only about 0.5 to 1×10^{-3} TW, i.e. two orders of magnitude smaller than the flux due the inertial pumping. Locally in space and time, however, it can reach similar magnitudes. Since the internal waves generated by interaction with surface waves are of high frequency, their turning points lie close to the mixed layer. It is therefore likely that the flux Φ_{tot} is also dissipated close to the mixed layer. This is different from the fluxes due to inertial pumping and the tides which generate low-frequency waves penetrating the entire depth range of the ocean. Our estimate of the implied dissipation shows that it sometimes reaches magnitudes comparable to observational estimates close to the mixed layer, in particular during strong wind events, but stays in general below them.

3.3.5 Evaluation with Available Observations

The results of the improved versions of IDEMIX need to be evaluated against observations. Measurements that resolve turbulence, however, are to date not available on a global scale. For the evaluation of IDEMIX, we therefore estimate dissipation rates of turbulent kinetic energy (TKE) and internal gravity wave energy from finestructure data, i.e. observations on 10–100 m vertical scales. The fundamental assumption of this ‘finestructure method’ is that TKE dissipation can be identified with the spectral energy transport caused by non-linear wave–wave interactions, which manifest themselves in the finescale variability of the internal wave field (Gregg 1989; Polzin et al. 1995). Based on the approach by Kunze et al. (2006) and Whalen et al. (2012), TKE-dissipation rates can be estimated from finescale strain information derived from Argo CTD profiles,⁶

$$\varepsilon = \varepsilon_0 \frac{\overline{N^2 \langle \xi_z^2 \rangle^2}}{N_0^2 \langle \xi_{z,GM}^2 \rangle^2} h(R_\omega) L(f, N), \quad (3.29)$$

where the observed strain variance $\langle \xi_z^2 \rangle$ is scaled by the corresponding value for the GM model, $\langle \xi_{z,GM}^2 \rangle$, $L(f, N)$ is a function correcting for latitudinal variations, and ε_0 is a constant. Due to the lack of shear information, the shear-to-strain ratio R_ω has to be set constant, reducing the function $h(R_\omega)$ to unity for the GM value of $R_\omega = 3$. Under the assumption that the observed internal gravity wave energy has the same wavenumber and frequency dependence as the GM model, it, too, can be estimated from finescale strain information (see Pollmann et al. 2017).

Figure 3.14 shows a comparison of Argo-derived and IDEMIX-based internal gravity wave energy levels. Like for dissipation rates (cf. Figure 3.5), IDEMIX well reproduces the magnitude as well as the spatial variations of the Argo-based estimates. Inconsistencies are, for example, found at high latitudes, e.g. in the northern

⁶The Argo programme maintains almost 4,000 freely drifting floats that profile temperature and salinity down to 2,000 m depth every 10 days, making the data publicly available within hours after their collection (www.argo.ucsd.edu).

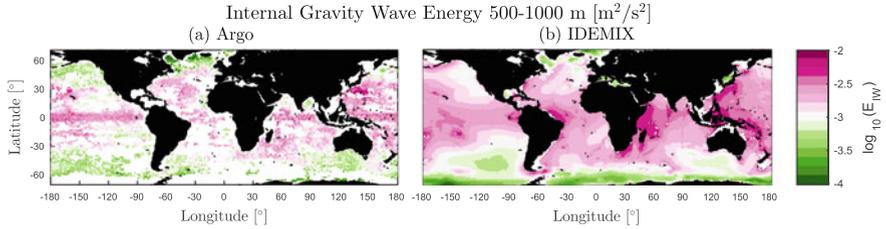


Fig. 3.14 Same as Figure 3.5 but for internal gravity wave energy. From Pollmann et al. (2017)

Pacific or Drake Passage, where IDEMIX simulates too high energy levels. In the 250–500 m depth range (not shown), this discrepancy is less pronounced, underlining the need to better understand and implement the depth dependence of the different forcing functions (e.g. the interaction with the mesoscale eddy field). Global estimates of internal gravity wave energies based on finescale strain information alone, like Figure 3.14, have to our knowledge not been attempted before.

The uncertainty of the Argo-derived dissipation rate and energy level estimates with respect to parameter choices and assumptions inherent in the finestructure method is shown in Figure 3.15. With variations amounting to more than a factor of two, the dissipation rate estimates prove more sensitive to modifications of the parameter settings than energy levels. Particularly, the value of the shear-to-strain

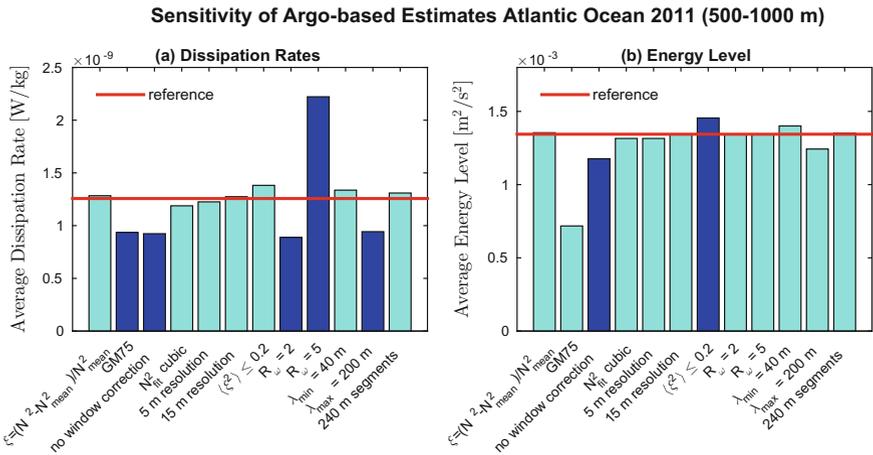


Fig. 3.15 Sensitivity of Argo-derived dissipation rate and energy level estimates: each bar represents a different scenario, where one parameter at a time was changed from the reference settings. These are given by $\xi_z = (N^2 - N_{fit}^2)/N_{mean}^2$, GM76 with $A(m) \propto (m^2 + m_*^2)^{-1}$, a resolution of 10 m, $\langle \xi_z^2 \rangle \leq 0.1$, $R_\omega = 3$, $\lambda_{min} = 10$ m, $\lambda_{max} = 100$ m and segments of 200 m length. Bars are shown in blue when the null hypothesis of a Welch's t-test, assuming equal mean dissipation rates in the reference case and the scenario in question, can be discarded for a significance level of $\alpha = 0.05$; cyan bars denote the failure to do so. Modified after Pollmann et al. (2017)

ratio, which is observed to vary between 2 and 7 in the global ocean (Kunze et al. 2006), strongly affects the dissipation rates' magnitude. In total, technical and statistical uncertainties can accumulate to a factor of 6–8 uncertainty, but compensations between the different forms are possible. Since dissipation rates vary globally by three orders of magnitude, a general comparison to IDEMIX is still feasible in any case. We aim to reduce the bias of these comparisons, for example, by evaluating IDEMIX against improved dissipation rate estimates that are based on a combination of shear and strain information.

3.4 Atmospheric Processes in IDEMIX

Important processes which generate gravity waves in the atmosphere are large-scale flow over orography, convection as well as spontaneous emission from jet streams, fronts, and other balanced flows. The waves propagate vertically and laterally over large distances and interact with the mean flow, and they feed energy to small-scale turbulence when they break. While in the ocean the density-mixing effect during wave breaking is considered to be most important since it generates potential energy and thus drives large-scale flow, it is the wave–mean flow interaction known as momentum and energy deposition that is most important in the atmosphere (e.g. Miller et al. 1989). Regarding the atmosphere, the momentum deposition, also known as gravity wave drag, is in the focus of parameterization schemes since it strongly contributes to driving the residual circulation in the middle atmosphere, to constrain the tropospheric jets, and to induce the quasi-biennial oscillation in the tropical stratosphere. The turbulent frictional heating that accompanies wave breaking becomes important in the middle atmosphere and needs consistent treatment as part of the energy deposition (Becker 2004). In contrast, this heating can be safely neglected in the ocean (McDougall 2003; Olbers et al. 2012; Eden 2015).

In general circulation models of the atmosphere, the gravity wave field is only marginally resolved and its effects on the mean flow need to be parameterized. The first theory that gave rise to simple gravity wave parameterizations in global circulation models is the saturation concept by Lindzen (1981). Here, upward propagating monochromatic waves are assumed to be damped by turbulent vertical diffusion above a certain critical height such that convective instability is marginally avoided. This wave damping leads to non-conservative wave–mean flow interaction according to the non-acceleration theorem (Andrews and McIntyre 1976). Other damping mechanisms are possible as well. For example, gravity wave schemes may employ a spectrum of waves that is truncated with height due to some criterion at the high vertical wavenumber end (Hines 1997; Alexander and Dunkerton 1999; Warner and McIntyre 2001). Lindzen's simple wave saturation concept has also been applied to orographic gravity waves by McFarlane (1987), and corresponding schemes are currently used in many climate models in order to simulate realistic jets in the troposphere and winter stratosphere. The response of orographic gravity waves is particularly essential in climate scenarios with regard to changes in the Brewer–Dobson

circulation (McLandress and Shepherd 2009). A basic overview of the wave driving of the middle atmosphere is given by Becker (2012).

Non-orographic gravity waves are sometimes also parameterized by concepts which go beyond linear theory, such as, for example, the Doppler spread theory (Hines 1997), where not only the Doppler shift by the mean flow but also that by the entire spectrum of parameterized waves is considered as a criterion for spectral truncation. Another example is the non-linear saturation theory of Medvedev and Klaasen (e.g. Yiğit et al. 2008), where non-linear wave interactions are assumed to trigger convective instabilities leading to a wave-induced eddy diffusion, as a first step towards accounting for the effect of resonant wave–wave interaction. Overviews of the gravity wave schemes currently used in global models are given in Fritts and Alexander (2003) and Alexander et al. (2010).

We emphasize that all conventional gravity wave schemes are based on the framework of the single-column approximation, as well as on the strong assumptions of a stationary background state and a stationary wave energy equation (see discussion in Becker and McLandress 2009). Conservative (reversible) wave–mean flow interactions, which may take place when a wave packet propagates through a vertically variable background atmosphere, are excluded in such a framework. Furthermore, the wave parameters have to be specified at a particular launch level; sources that are continuous in space and time, e.g. due to convection or frontal activity, cannot be incorporated in a consistent fashion. The latter restriction holds even for more sophisticated parameterizations that are based on ray tracing (Dunkerton and Fritts 1984; Warner and McIntyre 1996; Preusse et al. 2009; Senf and Achatz 2011). Conventional gravity wave schemes often also lack a consistent representation of scale interactions and energetics (Becker 2004; Becker and McLandress 2009). Particularly, the stationarity assumptions lead to parameterization errors when the interaction between gravity waves and thermal tides (having planetary scales and periods of 8, 12, and 24 hours) is considered. As shown by Senf and Achatz (2011), gravity wave–tidal interaction changes significantly when a non-stationary parameterization based on the full ray tracing equations is used instead of the conventional framework. One particular reason is that the horizontal phase speed (or ground-based frequency) of a gravity wave changes when the background wind is temporally variable (Fritts and Dunkerton 1984). In addition, horizontal propagation and refraction play an important role for gravity wave propagating from the lower to the middle atmosphere (Preusse et al. 2009; Fritts et al. 2006), but is neglected in the single-column approximation.

Müller and Natarov (2003) suggested to base a model for internal waves in the ocean on the radiative transfer equation of weakly interacting gravity waves (Hasselmann 1968), a concept which has—to our knowledge—never been considered for gravity wave parameterizations for the atmosphere. This equation describes (1) the propagation and refraction of the gravity wave field in physical and wavenumber space, (2) the wave–mean flow interaction, (3) non-linear wave–wave interactions, and (4) the forcing and dissipation of the waves. Olbers and Eden (2013) considered a drastic simplification of the concept which they called IDEMIX (Internal wave Dissipation, Energy and Mixing). Instead of resolving the detailed wave spectrum

as proposed in Müller and Natarov (2003), they integrate the radiative transfer equation in the wavenumber and frequency domain, leading to conservation equations for integral energy compartments in physical space, which can be closed with a few simple but reasonable parameterizations. The IDEMIX concept yields an energetically consistent framework to describe wave effects and has been shown to be successful for ocean applications. We propose to follow this approach and to base a new, energetically consistent gravity wave parameterization for atmosphere models on the radiative transfer equation. Applying the IDEMIX concept will also foster transfer of knowledge from the oceanic community to the atmospheric community and vice versa.

Integrating the radiative transfer equation in wavenumber space yields prognostic equations in physical space for energy compartments describing the mean wave propagation, refraction, and wave–mean flow interaction, and the integrated effects of wave–wave interaction, forcing and dissipation. The key to describe the mean propagation, refraction, and wave–mean flow interaction of the compartments in IDEMIX is the assumption that the actual wave spectrum is close to a wave spectrum of known shape but unknown amplitude. In the ocean, this would be the well-known Garrett–Munk spectrum (e.g. Garrett and Munk 1972). In the atmosphere, it would be a different shape as described, for example, by Fritts and VanZandt (1993) or Nastrom and Gage (1985), which was also assumed in the scheme by Warner and McIntyre (2001). The free parameters of the energy compartments are then used as additional prognostic variables in a circulation model. Olbers and Eden (2013) use only two energy compartments for the up- and downward propagating part of the full spectrum in a simple IDEMIX version. The wave–wave interaction is parameterized as linear decay of the asymmetry in the vertical propagation, and the dissipation of wave energy is given by a quadratic functional of total wave energy (Olbers 1976; Gregg 1989; Müller et al. 1986). In the atmospheric case, we will have different compartments for different directions of horizontal phase propagation, and it is sufficient to consider only upward group propagation.

Recently, IDEMIX was extended by single-column approximation of wave–mean flow interaction (Olbers and Eden 2017; Eden and Olbers 2017). Figure 3.16 shows wave energy compartments in a one-dimensional (z, t) idealized simulation using IDEMIX with waves propagating through a mean flow with vertical shear that is concentrated in the upper part of the domain. The waves are forced from below with a magnitude representative of oceanic tidal forcing. Shown are energy compartments of east- and westward and up- and downward propagating waves. The governing equations are given by (3.25). While up- and eastward (Figure 3.16b) and down- and westward (Figure 3.16c) propagating waves lose energy to the mean flow, down- and eastward (Figure 3.16a) and up- and westward (Figure 3.16d) propagating waves gain energy from the mean flow, while they propagate through the shear zone. Note that changes in wave energy outside the shear zone are only due to the variable buoyancy frequency. Without dissipation, the energy exchange of the waves with the mean flow would be completely reversible. Up- and eastward propagating wave lose energy, but are reflected at the surface to down- and eastward propagating waves and regain the same amount of energy; then, they are reflected at the bottom, lose

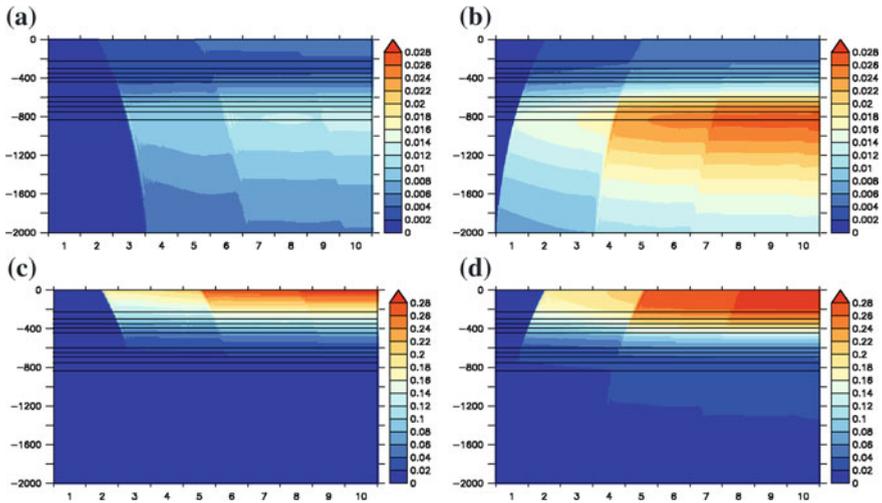


Fig. 3.16 Energy compartments in an idealized IDEMIX simulation as a function of height z and time t . Solid lines denote the shear zone. (a) Energy compartment of down- and eastward propagating, (b) of up- and eastward propagating, (c) down- and westward propagating, and (d) up- and westward propagating waves. Note the different colour scales in the upper and lower row. Energy magnitudes are representative of oceanic waves forced by the barotropic tidal flow at the bottom. From Olbers and Eden (2017)

energy, and so forth. The same is true for the westward propagating waves. Only if there is dissipation, as in the simulation shown in Figure 3.16, this symmetry is broken and a net energy exchange with the mean flow takes place. Note that this behaviour of IDEMIX is fully consistent with the non-acceleration theorem of Andrews and McIntyre (1976). In the atmosphere, surface reflection is missing and reversible energy exchanges with the mean flow are less likely than in the ocean, such that irreversible wave–mean flow interaction is of significant importance for the large-scale circulation. For a mean shear as shown in Figure 3.16, the energy transfer is from the mean flow to the waves, which is akin to a wave drag, but other configurations (jet-like situations) can also lead to an acceleration of the mean flow by the waves.

3.5 Summary

The recently proposed parameterization module IDEMIX (Internal wave Dissipation Energy and Mixing) describes the generation, propagation, interaction, and dissipation of the internal gravity wave field and can be used in ocean general circulation models to account for vertical mixing (and friction) in the interior of the ocean. It is based on the radiative transfer equation of a weakly interacting internal wave field,

for which spectrally integrated energy compartments are used as prognostic model variables. IDEMIX is central to the concept of an energetically consistent ocean model, since it enables to link all sources and sinks of internal wave energy and furthermore all parameterized forms of energy in an ocean model without spurious sources and sinks of energy.

In its simplest version, IDEMIX consists of two energy compartments of interacting up- and downward propagating waves; in a more detailed version, low vertical mode compartments at near-inertial and frequencies of tidal constituents are added which also account for horizontally anisotropic wave propagation. It is proposed to extend IDEMIX

- with forcing mechanism by mesoscale eddy dissipation, such as lee wave generation or spontaneous wave emission of balanced flow
- with the effect of the interaction of the internal wave field with mesoscale and large-scale mean flow including critical layers and wave capture
- with anisotropic forcing of the low-mode tidal energy compartments
- and with additional energy compartments for high-frequency, high-wavenumber waves which are in particular effective for upper ocean mixing.

All these processes have never been implemented in ocean models but have an important effect on mixing and the energy transfers in the ocean. We will validate the simple and more complex versions of IDEMIX and the new version using available fine- and microstructure data sets. The simple and more complex IDEMIX versions will be implemented into the ICON and FESOM ocean models. By linking the wave sources with the wave dissipation using IDEMIX both in the ocean and atmosphere, we envision to close the energy cycle in a consistent way for the coupled system.

Gravity waves are an important part of the energy cycle of the atmosphere and exchange momentum and energy with the mean flow due to wave breaking and wave refraction. Wave breaking and the resulting mean flow effects need special parameterization in global climate models as they usually resolve at most a small part of the full spectrum of gravity waves. Here, we apply the IDEMIX concept to develop corresponding gravity wave schemes for atmospheric circulation models. We propose to base a new, energetically consistent gravity wave parameterization on the radiative transfer equation for a field of waves. This method is fundamentally different from conventional schemes which describe the superposition of monochromatic waves launched at a particular level and which make the strong assumption of a stationary mean flow.

The IDEMIX concept was shown to be successful for ocean applications, but instead of focussing on the mixing effect by breaking waves as for the oceanic case, the focus in the atmospheric application is on the wave–mean flow interaction, i.e. the gravity wave drag and the energy deposition. We will extend the concept of energetically consistent closures to atmospheric gravity wave closures. We will derive gravity wave source functions due to flow over orography and due to the forcing by frontal dynamics and convection. We will include the effects of transience of the mean flow on the gravity wave field, and we will allow to incorporate sinks and sources that are continuous in space and time. During the first phase of the project, we will

stick to the single-column approximation in order to allow for easy implementation of the derived gravity wave scheme in global circulation models.

There are several ways to generalize the proposed gravity wave scheme. The most obvious one is to use the general dispersion relation for gravity waves such as to incorporate also high-frequency and inertial-gravity waves. In the latter case, the momentum flux must be substituted by the pseudo-momentum flux in order to account for the Stokes drift when ω approaches the Coriolis parameter (Fritts and Alexander 2003), as is in principle always the case close to a critical level. Furthermore, we may dispense with discrete azimuths and rather use a continuous directional representation, thereby modifying also the radiative transfer equation (3.13). Furthermore, one should consider existing theories to account for non-linear effects and wave–wave interaction in the wave-breaking process (Yiğit et al. 2008) and incorporate this insight into the new framework. Simple closures for the wave–wave interaction in the oceanic case are outlined in Olbers and Eden (2013) and may be considered also for atmospheric case. Finally, we may relax the single-column approach and consider wave propagation and refraction in all three directions.

References

- Alexander, M., Dunkerton, T.: A spectral parameterization of mean-flow forcing due to breaking gravity waves. *J. Atmos. Sci.* **56**(24), 4167–4182 (1999)
- Alexander, M., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., Sato, K., Eckermann, S., Ern, M., Hertzog, A., et al.: Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Q. J. R. Meteorol. Soc.* **136**(650), 1103–1124 (2010)
- Alford, M.: Internal swell generation: the spatial distribution of energy flux from the wind to mixed layer near-inertial motions. *J. Phys. Oceanogr.* **31**(8), 2359–2368 (2001)
- Andrews, D.G., McIntyre, M.E.: Planetary waves in horizontal and vertical shear: the generalized Eliassen-Palm relation and the zonal mean acceleration. *J. Atmos. Sci.* **33**, 2031–2048 (1976)
- Arbic, B.K., Shriver, J.F., Hogan, P.J., Hurlburt, H.E., McClean, J.L., Metzger, E.J., Scott, R.B., Sen, A., Smedstad, O.M., Wallcraft, A.J.: Estimates of bottom flows and bottom boundary layer dissipation of the oceanic general circulation from global high-resolution models. *J. Geophys. Res.* **114**(C2) (2009)
- Becker, E.: Direct heating rates associated with gravity wave saturation. *J. Atmos. Sol. Terr. Phys.* **66**(6), 683–696 (2004)
- Becker, E.: Dynamical control of the middle atmosphere. *Space Sci. Rev.* **168**(1–4), 283–314 (2012)
- Becker, E., McLandress, C.: Consistent scale interaction of gravity waves in the Doppler spread parameterization. *J. Atmos. Sci.* **66**(5), 1434–1449 (2009)
- Bell, T.H.: Topographically generated internal waves in the open ocean. *J. Geophys. Res.* **80**(3), 320–327 (1975)
- Bell, T.H.: Radiation damping of inertial oscillations in the upper ocean. *J. Fluid Mech.* **88**(02), 289–308 (1978)
- Bender, C.M., Orszag, S.A.: *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory*. McGraw-Hill Book Company, New York (1978)
- Berry, M.V., Mount, K.: Semiclassical approximations in wave mechanics. *Rep. Prog. Phys.* **35**(1), 315–397 (1972)

- Bretherton, F.P., Garrett, C.J.: Wavetrains in inhomogeneous moving media. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 302, pp. 529–554. The Royal Society (1968)
- Brüggemann, N., Eden, C.: Routes to dissipation under different dynamical conditions. *J. Phys. Oceanogr.* **45**(8), 2149–2168 (2015)
- Bryan, K., Lewis, L.: A water mass model of circulation cells and the source waters of the Pacific equatorial undercurrent. *J. Phys. Oceanogr.* **28**, 62–84 (1979)
- Bühler, O., McIntyre, M.E.: Wave capture and wave-vortex duality. *J. Fluid Mech.* **534**, 67–95 (2005)
- Cairns, J.L., Williams, G.O.: Internal wave observations from a midwater float, 2. *J. Geophys. Res.* **81**(12), 1943–1950 (1976)
- Charron, M., Manzini, E.: Gravity waves from fronts: Parameterization and middle atmosphere response in a general circulation model. *J. Atmos. Sci.* **59**(5), 923–941 (2002)
- Cummins, P.F., Holloway, G., Gargett, A.E.: Sensitivity of the GFDL ocean general circulation model to a parameterization of vertical diffusion. *J. Phys. Oceanogr.* **20**, 817–830 (1990)
- Desaubies, Y.J.: Analytical representation of internal wave spectra. *J. Geophys. Res.* **6**(6), 976–981 (1976)
- Dewar, W.K., Hogg, A.M.: Topographic inviscid dissipation of balanced flow. *Ocean Model.* **32**(1), 1–13 (2010)
- Dunkerton, T.J., Fritts, D.C.: Transient gravity wave-critical layer interaction. Part I: convective adjustment and the mean zonal acceleration. *J. Atmos. Sci.* **41**(6), 992–1007 (1984)
- Dushaw, B.D., Worcester, P.F., Dzieciuch, M.A.: On the predictability of mode-1 internal tides. *Deep-Sea Res.* **58**(6), 677–698 (2011)
- Eden, C.: Revisiting the energetics of the ocean in Boussinesq approximation. *J. Phys. Oceanogr.* **45**(3), 630–637 (2015)
- Eden, C., Czeschel, L., Olbers, D.: Towards energetically consistent ocean models. *J. Phys. Oceanogr.* **44**(8), 2093–2106 (2014)
- Eden, C., Greatbatch, R.J.: Diapycnal mixing by mesoscale eddies. *Ocean Model.* **23**, 113–120 (2008a)
- Eden, C., Greatbatch, R.J.: Towards a mesoscale eddy closure. *Ocean Model.* **20**, 223–239 (2008b)
- Eden, C., Olbers, D.: An energy compartment model for propagation, non-linear interaction and dissipation of internal gravity waves. *J. Phys. Oceanogr.* **44**, 2093–2106 (2014)
- Eden, C., Olbers, D.: Interaction of internal waves with a mean flow. Part II: wave drag. *J. Phys. Oceanogr.* **47**, 1403–1412 (2017)
- Egbert, G.D., Erofeeva, S.Y., Ray, R.D.: Assimilation of altimetry data for nonlinear shallow-water tides: quarter-diurnal tides of the Northwest European Shelf. *Cont. Shelf Res.* **30**(6), 668–679 (2010)
- Ford, R., McIntyre, M., Norton, W.: Balance and the slow quasimanifold: some explicit results. *J. Atmos. Sci.* **57**(9), 1236–1254 (2000)
- Fritts, D.C., Alexander, M.J.: Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.* **41**(1) (2003)
- Fritts, D.C., Dunkerton, T.J.: A quasi-linear study of gravity-wave saturation and self-acceleration. *J. Atmos. Sci.* **41**(22), 3272–3289 (1984)
- Fritts, D.C., Vadas, S.L., Wan, K., Werne, J.A.: Mean and variable forcing of the middle atmosphere by gravity waves. *J. Atmos. Sol. Terr. Phys.* **68**(3), 247–265 (2006)
- Fritts, D.C., VanZandt, T.E.: Spectral estimates of gravity wave energy and momentum fluxes. Part I: energy dissipation, acceleration, and constraints. *J. Atmos. Sci.* **50**(22), 3685–3694 (1993)
- Garcia, R., Marsh, D., Kinnison, D., Boville, B., Sassi, F.: Simulation of secular trends in the middle atmosphere, 1950–2003. *J. Geophys. Res.* **112**(D9) (2007)
- Garrett, C.: What is the near-inertial wave band and why is it different from the rest of the internal wave spectrum? *J. Phys. Oceanogr.* **31**, 962–71 (2001)
- Garrett, C., Munk, W.: Space-time scales of internal waves. *Geophys. Astrophys. Fluid Dyn.* **3**(1), 225–264 (1972)

- Garrett, C., Munk, W.: Space-time scales of internal waves: a progress report. *J. Geophys. Res.* **80**, 291–297 (1975)
- Gaspar, P., Gregoris, Y., Lefevre, J.-M.: A simple eddy kinetic energy model for simulations of the oceanic vertical mixing: tests at station PAPA and Long-Term Upper Ocean Study site. *J. Geophys. Res.* **95**, 16179–16193 (1990)
- Goff, J.A., Arbic, B.K.: Global prediction of abyssal hill roughness statistics for use in ocean models from digital maps of paleo-spreading rate, paleo-ridge orientation, and sediment thickness. *Ocean Model.* **32**(1), 36–43 (2010)
- Gregg, M.C.: Scaling turbulent dissipation in the thermocline. *J. Geophys. Res.* **94**(C7), 9686–9698 (1989)
- Hasselmann, K.: Weak-interaction theory of ocean waves. *Basic Dev. Fluid Dyn.* **2**, 117–182 (1968)
- Hasselmann, K., Barnett, T., Bouws, E., Carlson, H., Cartwright, D., Enke, K., Ewing, J., Gienapp, H., Hasselmann, D., Krusemann, P., Meerburg, A., Müller, P., Olbers, D., Richter, K., Sell, W., Walden, H.: Measurements of wind-wave growth and swell decay during the Joint North Sea Wave Project (JONSWAP). *Dtsch. Hydrogr. Z.* **12**, 1–95 (1973)
- Hines, C.O.: Doppler-spread parameterization of gravity-wave momentum deposition in the middle atmosphere. Part 1: basic formulation. *J. Atmos. Sol. Terr. Phys.* **59**(4), 371–386 (1997)
- Jayne, S., St. Laurent, L.: Parameterizing tidal dissipation over rough topography. *Geophys. Res. Lett.* **28**(5), 811–814 (2001)
- Jones, W.L.: Ray tracing for internal gravity waves. *J. Geophys. Res.* **74**(8), 2028–2033 (1969)
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D.: The NCEP/NCAR 40-years reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996)
- Kunze, E., Firing, E., Hummon, J.M., Chereskin, T.K., Thurnherr, A.M.: Global abyssal mixing inferred from lowered ADCP shear and CTD strain profiles. *J. Phys. Oceanogr.* **36**(8), 1553–1576 (2006)
- Kunze, E., Smith, S.L.: The role of small-scale topography in turbulent mixing of the global ocean. *Oceanography* **17**(1), 55–64 (2004)
- Landau, L.D., Lifshitz, E.M.: *Mechanics, Course of Theoretical Physics*, vol. 1. Pergamon Press Ltd. (1982)
- Lighthill, M.J.: On sound generated aerodynamically. I. General theory. In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 211, pp. 564–587. The Royal Society (1952)
- Lindzen, R.S.: Turbulence and stress owing to gravity wave and tidal breakdown. *J. Geophys. Res.* **86**(C10), 9707–9714 (1981)
- McComas, C., Müller, P.: Time scales of resonant interactions among oceanic internal waves. *J. Phys. Oceanogr.* **11**, 139–147 (1981)
- McComas, C.H.: Equilibrium mechanisms within the oceanic internal wave field. *J. Phys. Oceanogr.* **7**, 836–845 (1977)
- McDougall, T.J.: Potential enthalpy: a conservative oceanic variable for evaluating heat content and heat fluxes. *J. Phys. Oceanogr.* **33**(5), 945–963 (2003)
- McFarlane, N.: The effect of orographically excited gravity wave drag on the general circulation of the lower stratosphere and troposphere. *J. Atmos. Sci.* **44**(14), 1775–1800 (1987)
- McLandress, C., Shepherd, T.G.: Simulated anthropogenic changes in the brewer-dobson circulation, including its extension to high latitudes. *J. Clim.* **22**(6), 1516–1540 (2009)
- Miller, M., Palmer, T., Swinbank, R.: Parametrization and influence of subgridscale orography in general circulation and numerical weather prediction models. *Meteorol. Atmos. Phys.* **40**(1–3), 84–109 (1989)
- Molemaker, M.J., McWilliams, J.C., Yavneh, I.: Baroclinic instability and loss of balance. *J. Phys. Oceanogr.* **35**, 1505 (2005)
- Molemaker, M.J., McWilliams, J. C., Capet, X.: Balanced and unbalanced routes to dissipation in an equilibrated Eady flow. *J. Fluid Mech.* **654**, 35–63 (2010)

- Müller, M., Cherniawsky, J., Foreman, M., Storch, J.-S.: Global M2 internal tide and its seasonal variability from high resolution ocean circulation and tide modeling. *Geophys. Res. Lett.* **39**(19) (2012)
- Müller, P.: On the diffusion of momentum and mass by internal gravity waves. *J. Fluid Mech.* **77**(04), 789–823 (1976)
- Müller, P., Holloway, G., Henyey, F., Pomphrey, N.: Nonlinear interactions among internal gravity waves. *Rev. Geophys.* **24**(3), 493–536 (1986)
- Müller, P., Natarov, A.: The internal wave action model (IWAM). In: Proceedings, 13th Aha Huliko‘a Hawaiian Winter Workshop, School of Ocean and Earth Science and Technology, Special Publication (2003)
- Müller, P., Olbers, D.J.: On the dynamics of internal waves in the deep ocean. *J. Geophys. Res.* **80**, 3848–3860 (1975)
- Müller, P., Olbers, D.J., Willebrand, J.: The IWEX spectrum. *J. Geophys. Res.* **83**, 479–500 (1978)
- Munk, W.: Internal waves and small-scale processes. In: Warren, B.A., Wunsch, C. (eds.) *Evolution of Physical Oceanography*, pp. 264–291. MIT Press, Cambridge, MA (1981)
- Nastrom, G., Gage, K.S.: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.* **42**(9), 950–960 (1985)
- Nikurashin, M., Ferrari, R.: Global energy conversion rate from geostrophic flows into internal lee waves in the deep ocean. *Geophys. Res. Lett.* **38**, L08610 (2011)
- Nycander, J.: Generation of internal waves in the deep ocean by tides. *J. Geophys. Res.* **110**(C9), C10028 (2005)
- Olbers, D., Eden, C.: A global model for the diapycnal diffusivity induced by internal gravity waves. *J. Phys. Oceanogr.* **43**, 1759–1779 (2013)
- Olbers, D., Eden, C.: Revisiting the generation of internal waves by resonant interaction with surface waves. *J. Phys. Oceanogr.* **46**, 2335–2350 (2016)
- Olbers, D., Eden, C.: Interaction of internal waves with a mean flow. Part I: energy conversion. *J. Phys. Oceanogr.* **47**, 1389–1401 (2017)
- Olbers, D., Willebrand, J., Eden, C.: *Ocean Dynamics*. Springer, Heidelberg (2012)
- Olbers, D.J.: Nonlinear energy transfer and the energy balance of the internal wave field in the deep ocean. *J. Fluid Mech.* **74**, 375–399 (1976)
- Olbers, D.J.: Models of the oceanic internal wave field. *Rev. Geophys. Space Phys.* **21**, 1567–1606 (1983)
- Olbers, D.J.: Internal gravity waves. In: Stündermann, J. (ed.) *Landolt-Börnstein—Numerical Data and Functional Relationships in Science and Technology—New Series, Group V, vol. 3a*, pp. 37–82. Springer, Berlin (1986)
- Olbers, D.J., Herterich, K.: The spectral energy transfer from surface waves to internal waves. *J. Fluid Mech.* **92**, 349–379 (1979)
- Osborn, T.R., Cox, C.S.: Oceanic fine structure. *Geophys. Astrophys. Fluid Dyn.* **3**, 321–345 (1972)
- Pollmann, F., Eden, C., Olbers, D.: Evaluating the global internal wave model IDEMIX using a finescale strain approximation. *J. Phys. Oceanogr.* **47**, 2267–2289 (2017)
- Polton, J.A., Smith, J.A., MacKinnon, J.A., Tejada-Martínez, A.E.: Rapid generation of high-frequency internal waves beneath a wind and wave forced oceanic surface mixed layer. *Geophys. Res. Lett.* **35**(13) (2008)
- Polzin, K.L.: Mesoscale eddy-internal wave coupling. Part I: symmetry, wave capture, and results from the mid-ocean dynamics experiment. *J. Phys. Oceanogr.* **38**(11), 2556–2574 (2008)
- Polzin, K.L., Lvov, Y.V.: Toward regional characterizations of the oceanic internal wavefield. *Rev. Geophys.* **49**(4), RG4003 (2011)
- Polzin, K.L., Toole, J.M., Schmitt, R.W.: Finescale parameterizations of turbulent dissipation. *J. Phys. Oceanogr.* **25**(3), 306–328 (1995)
- Preusse, P., Eckermann, S.D., Ern, M., Oberheide, J., Picard, R.H., Roble, R.G., Riese, M., Russell, J.M., Mlynczak, M.G.: Global ray tracing simulations of the saber gravity wave climatology. *J. Geophys. Res.* **114**(D8) (2009)

- Richter, J.H., Sassi, F., Garcia, R.R.: Toward a physically based gravity wave source parameterization in a general circulation model. *J. Atmos. Sci.* **67**(1), 136–156 (2010)
- Rimac, A., von Storch, J.-S., Eden, C., Haak, H.: The influence of high-resolution wind stress field on the power input to near-inertial motions in the ocean. *Geophys. Res. Lett.* **40**(18), 4882–4886 (2013)
- Scinocca, J.F.: An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *J. Atmos. Sci.* **60**(4), 667–682 (2003)
- Scott, R., Goff, J., Naveira Garabato, A., Nurser, A.: Global rate and spectral characteristics of internal gravity wave generation by geostrophic flow over topography. *J. Geophys. Res.* **116**(C9) (2011)
- Senf, F., Achatz, U.: On the impact of middle-atmosphere thermal tides on the propagation and dissipation of gravity waves. *J. Geophys. Res.* **116**(D24) (2011)
- Shaw, T.A., Shepherd, T.G.: A theoretical framework for energy and momentum consistency in subgrid-scale parameterization for climate models. *J. Atmos. Sci.* **66**(10), 3095 (2009)
- Smith, W.H.F., Sandwell, D.T.: Global seafloor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1957–1962 (1997)
- Tandon, A., Garrett, C.: On a recent parameterization of mesoscale eddies. *J. Phys. Oceanogr.* **26**, 406–416 (1996)
- Thorpe, S.: The excitation, dissipation, and interaction of internal waves in the deep ocean. *J. Geophys. Res.* **80**(3), 328–338 (1975)
- von Storch, J.-S., Eden, C., Fast, I., Haak, H., Hernández-Deckers, D., Maier-Reimer, E., Marotzke, J., Stammer, D.: An estimate of the Lorenz energy cycle for the world ocean based on the STORM/NCEP simulation. *J. Phys. Oceanogr.* **42**(12), 2185–2205 (2012)
- Wang, S., Zhang, F.: Source of gravity waves within a vortex-dipole jet revealed by a linear model. *J. Atmos. Sci.* **67**(5), 1438–1455 (2010)
- Warner, C., McIntyre, M.: On the propagation and dissipation of gravity wave spectra through a realistic middle atmosphere. *J. Atmos. Sci.* **53**(22), 3213–3235 (1996)
- Warner, C., McIntyre, M.: An ultrasimple spectral parameterization for nonorographic gravity waves. *J. Atmos. Sci.* **58**(14), 1837–1857 (2001)
- Watson, K.M.: The coupling of surface and internal gravity waves: revisited. *J. Phys. Oceanogr.* **20**, 1233–1248 (1990)
- Whalen, C., Talley, L., MacKinnon, J.: Spatial and temporal variability of global ocean mixing inferred from argo profiles. *Geophys. Res. Lett.* **39**, L18612 (2012)
- Whitham, G.: Two-timing, variational principles and waves. *J. Fluid Mech.* **44**(02), 373–395 (1970)
- Williams, P.D., Haine, T.W., Read, P.L.: Inertia–gravity waves emitted from balanced flow: observations, properties, and consequences. *J. Atmos. Sci.* **65**(11) (2008)
- Wunsch, C., Ferrari, R.: Vertical mixing, energy and the general circulation of the oceans. *Annu. Rev. Fluid Mech.* **36**, 281–314 (2004)
- Yiğit, E., Aylward, A.D., Medvedev, A.S.: Parameterization of the effects of vertically propagating gravity waves for thermosphere general circulation models: Sensitivity study. *J. Geophys. Res.* **113**(D19) (2008)

Chapter 4

Observations and Models of Low-Mode Internal Waves in the Ocean



Christian Mertens, Janna Köhler, Maren Walter, Jin-Song von Storch and Monika Rhein

Abstract The generation of internal gravity waves in the ocean is largely driven by tides, winds, and interaction of currents with the seafloor. Models and observations indicate a global energy supply for the internal wave field of about 1 TW by the conversion of barotropic tides at mid-ocean ridges and abrupt topographic features. Winds acting on the oceanic mixed layer contribute 0.3–1.5 TW, and mesoscale flow over rough topography adds about 0.2 TW. Globally, 1–2 TW are needed to maintain the observed stratification of the deep ocean by diapycnal mixing that results from the breaking of internal waves. Ocean circulation models show significant impact of the spatial distribution of internal wave dissipation and mixing on the ocean state, e.g., thermal structure, stratification, and meridional overturning circulation. Observations indicate that the local ratio of generation and dissipation of internal waves is often below unity, and thus, the energy available for mixing must be redistributed by internal tides and near-inertial waves at low vertical wavenumber that can propagate thousands of kilometers from their source regions. Eddy-permitting global ocean circulation models are able to quantify the different sources of energy input and can also simulate the propagation of the lowest internal wave modes. However, the variation of the internal wave energy flux along its paths by wave–wave interaction, topographic scattering, and refraction by mesoscale features as well as its ultimate fate by dissipation remains to be parameterized.

C. Mertens (✉) · J. Köhler · M. Walter · M. Rhein
MARUM/Institut für Umweltphysik, Universität Bremen, Bremen, Germany
e-mail: cmertens@uni-bremen.de

J. Köhler
e-mail: jannak@uni-bremen.de

M. Walter
e-mail: maren.walter@uni-bremen.de

M. Rhein
e-mail: mrhein@uni-bremen.de

J.-S. von Storch
Max-Planck-Institut für Meteorologie (MPI-M), Hamburg, Germany
e-mail: jin-song.von.storch@mpimet.mpg.de

4.1 Introduction

Internal gravity waves are generated in stratified fluids by perturbations from a state of equilibrium. These perturbations cause currents that rotate in a two-dimensional plane that is inclined to the horizontal. The restoring forces are the buoyancy force in the vertical and the Coriolis force in the horizontal. The angle of inclination, which is a monotonic function of the frequency, sets the direction of phase propagation. The upper and lower limits of the frequency range of propagating linear internal waves are determined by the local buoyancy and inertial frequencies. Low-frequency waves have a nearly vertical and high-frequency waves a nearly horizontal phase propagation. The group velocity and thus the direction of energy radiation is perpendicular to the phase propagation, nearly horizontal for low-frequency waves and nearly vertical for high-frequency waves.

The supply of energy into the oceanic internal wave field comes primarily from two sources: Tides that flow over sloping topography and generate internal waves at tidal frequencies, known as internal tides, and winds that act on the sea surface and generate near-inertial waves. Other generation processes include lee wave generation by geostrophic flow impinging on topography and spontaneous emission through loss of balance. A cascade of dynamical processes ultimately leads to turbulent mixing and viscous dissipation at small scales in the interior and the boundaries of the ocean. Observations indicate that the local ratio of generation and dissipation of internal waves is often below unity (Waterhouse et al. 2014), and thus, the energy available for mixing must be redistributed by internal tides and near-inertial waves at low vertical wavenumber that can propagate thousands of kilometers from their source regions.

Low-mode internal waves, with horizontal wavelengths of typically 100–200 km, are crucial for the oceanic energy pathways and ocean interior mixing, since they hold a large amount of the kinetic energy of the entire internal wave field (e.g., Wunsch 1975). It is still a challenge to quantify the fluxes associated with these waves on a global scale and to properly include them in ocean general circulation models. This is particularly true regarding the observation and simulation of the fate of these waves away from their generation sites.

Observations and numerical models indicate a global input of about 1 TW (10^{12} W) from tides available for deep ocean mixing (Egbert and Ray 2000; Müller 2013; Nycander 2005; Simmons et al. 2004). Power estimates from wind forcing vary between 0.3 and 1.5 TW transferred into near-inertial motions in the surface mixed layer (Waterhouse et al., 2014). From observations at Ocean Station Papa in the northwest Pacific, it was found that 12–33% of the energy input were transmitted into the deep ocean (Alford et al. 2012). The total energy flux from lee wave generation is about 0.2 TW (Nikurashin and Ferrari 2011). The fractions of the individual sources of power input vary in the literature, but in total they sum up to account for the roughly 1–2 TW necessary to maintain the deep ocean stratification (Munk and Wunsch 1998; Wunsch and Ferrari 2004).

To accurately simulate the ocean circulation, not only the total power input, but also the geographic distribution of deep mixing needs to be known. A number of model studies show that besides the strength of the climate-relevant meridional overturning circulation (MOC), the deep ocean stratification and the distribution of passive tracers respond to changing patterns of imposed deep mixing (Eden et al. 2014; Exarchou et al. 2012, 2014; Melet et al. 2013, 2014). The importance of the representation of lee wave-driven mixing in climate models was emphasized by Nikurashin and Ferrari (2013). Up to now, it is unclear where and when exactly the waves break and lead to mixing.

This paper reviews observational and modeling studies on internal wave energetics, relevant for the energy transfer in the ocean. Numerical model studies (Section 4.2), sorted according to wave generation mechanisms, are used to derive quantitative estimates of wave sources in the global ocean. Dissipation mechanisms, which are less well observed and extremely difficult to be properly treated in numerical models, are discussed in Section 4.3. The observational studies (Section 4.4), sorted according to measurement methods, are devoted mainly to detecting coherent signals of low-mode internal tides and to quantifying wave energy fluxes and turbulent dissipation, especially in terms of diapycnal mixing. A summary and the outlook are given in the final section.

4.2 Numerical Modeling

The generation of internal waves, both with respect to the globally integrated generation rates and to the spatial distribution of the generation, has been studied extensively using numerical models.

4.2.1 Wind

Temporally varying winds at the sea surface generate near-inertial waves. The early estimates of wind power input into near-inertial motions were obtained by calculating the wind-induced inertial motions $\mathbf{u}_i = (u_i, v_i)$ from a slab ocean model and estimating the wind-induced inertial energy flux across the sea surface using $\boldsymbol{\tau} \cdot \mathbf{u}_i$, where $\boldsymbol{\tau} = (\tau_x, \tau_y)$ is the wind stress obtained from a reanalysis. To obtain $\mathbf{u}_i = (u_i, v_i)$, the approach of D'Asaro et al. (1995) and Pollard and Millard (1970) is used, in which the momentum equations are given by

$$\frac{dZ}{dt} + \omega Z = \frac{X}{H}. \quad (4.1)$$

H is the mixed layer depth, ρ density, $Z = (u + iv)$, $X = (\tau_x + i\tau_y)/\rho$, and $\omega = (r + if)$, with f being the local inertial frequency and r an empirical damping

constant. The solution of Eq. (4.1) can be written as the sum of a time-varying Ekman transport $Z_E = X/\omega H$ and an inertially rotating current $Z_i = u_i + iv_i = Z - Z_E$, which is the solution to

$$\frac{dZ_I}{dt} + \omega Z_I = -\frac{1}{\omega} \frac{d(X/H)}{dt}. \quad (4.2)$$

The solution of Equation (4.2), which can be obtained spectrally, is then used to estimate the wind power input to the near-inertial motions. Integrated globally, this power input is estimated to be about 0.7 TW in Watanabe and Hibiya (2002) and 0.47 TW in Alford (2003a). The strongest wind power input is found in the western portion of each ocean basin at mid-latitudes, especially in the winter season, confirming the role of mid-latitude storms for providing the bulk of the near-inertial energy flux.

Plueddemann and Farrar (2006) pointed out that slab ocean models do not account for the energy lost due to vertical shear instability and may therefore systematically overestimate the wind work. The later estimates are hence based on three-dimensional OGCMs in which the effect of shear instability is parametrized. Furuichi et al. (2008) used the Princeton Ocean Model (POM, Mellor 2003) on a grid with horizontal spacing of 0.15° in longitude and 0.125° in latitude, but without simulating mesoscale flows. The global ocean was divided into three subregions, and numerical simulations were carried out separately for each subregion. By applying a high-pass filter to retain fluctuations at frequencies larger than $0.8 f$, they estimated the annual mean global wind energy input to be 0.4 TW. Simmons and Alford (2012) used the Generalized Ocean Layer Dynamics (GOLD) model, an evolution of the Hallberg Isopycnal Coordinate Model (HIM, Hallberg 1997), with a nominal horizontal resolution of $1/8^\circ$. The model simulates mesoscale eddies along with the large-scale circulation. The wind power input to near-inertial motions is again estimated from the band-passed wind stress and surface velocities. Averaged over different seasons, the global input is about 0.36 TW. The small difference between this estimate and the results obtained by Furuichi et al. (2008) was taken as an indication that the effect of mesoscale eddies is small.

Jiang et al. (2005) showed that the near-inertial wind power input is sensitive to the wind products used. Rimac et al. (2013) investigated the detailed role of spatial and temporal resolution of the wind stress forcing for the near-inertial wind power input using a series of numerical experiments. Considerably higher values are obtained when forcing the model with high-frequency (higher than 6 hourly) wind stress data. Generally, the near-inertial wind-induced power is transferred downward and equatorward (e.g., Alford and Zhao 2007). However, a large fraction of the wind power input dissipates within the mixed layer. The recent estimates of the fraction that enters the deep ocean, about 10% (Rimac et al. 2016), are much lower than the early one of about 37% (D'Asaro et al. 1995).

4.2.2 Tides

Barotropic tides generate internal waves at tidal frequencies over rough topography. The higher mode internal tides have short wavelengths and break near the generation sites and dissipate locally (e.g., Klymak et al. 2008). This results in enhanced mixing near the generation sites. This effect of tides can be parametrized in an OGCM following the scheme by St. Laurent et al. (2002). The low modes of internal tides have long wavelengths of a few tens to about 100–200 km (group velocities are in the order of cm/s). They carry most of the energy converted from the tides and can radiate thousands of kilometers away from their sources. Their effect is normally parametrized in an OGCM in terms of a uniform mixing coefficient.

Global studies on the tidal generation of internal waves have followed different approaches. Sjöberg and Stigebrandt (1992) calculated the tidal generation in terms of barotropic to baroclinic energy transfer due to baroclinic wave drag. To fulfill the condition of vanishing normal velocity at non-horizontal topography, the approach assumes that baroclinic waves have to be generated so that the sum of the baroclinic and barotropic modes satisfies the kinematic boundary condition. Using this approach and a state-of-the-art numerical tidal model, Gustafsson (2001) estimated the total loss of barotropic tidal energy in the deep ocean (between 70° S and 70° N and at depth greater than 1000 m) to be about 0.7 TW from the M_2 tide. The approach is purely local, although wave generation is a non-local process. The approach following Bell (1975a, b) and Llewellyn Smith and Young (2002), on the other hand, is non-local. It is based on linear wave theory. A direct consequence of linear wave theory is that the solutions are linear transformations of the bottom topography, so that in phase space they are proportional to the Fourier transform of the topography. The total energy flux from tides to internal waves at a location is a convolution integral around that location. Nycander (2005) studied the tidal generation of internal waves following this approach. Using a bottom topography having a nominal resolution of two minutes and the tidal velocity field from the TPXO.6 model (Egbert and Erofeeva 2002), he estimated the tidal generation over the area below 500 m to be about 1.2 TW.

These early studies on tidal generation rely on barotropic tidal models and linear wave theory, as the OGCMs were unable to resolve internal tides. With increasing computer power, it becomes possible to resolve low-mode internal tides using an OGCM. Two groups have achieved to perform concurrent simulations of the eddy-circulation and internal tides. One is the HYCOM group, based on the HYbrid Coordinate Ocean Model with a hybrid isopycnal/terrain-following vertical coordinate and a horizontal resolution of about $1/12.5^\circ$ (Arbic et al. 2010, 2012). The other is the STORMTIDE group, based on the 0.1° Max-Planck Institute Ocean Model (MPIOM) developed with the German consortium project STORM (Müller et al. 2012). In both the HYCOM and STORMTIDE models, barotropic tides are reasonably well simulated. The simulated internal tides' signatures in the sea surface height compare also well with those obtained from the altimeter data. Li et al. (2015) show further that away from strong currents such as the Antarctic Circumpolar

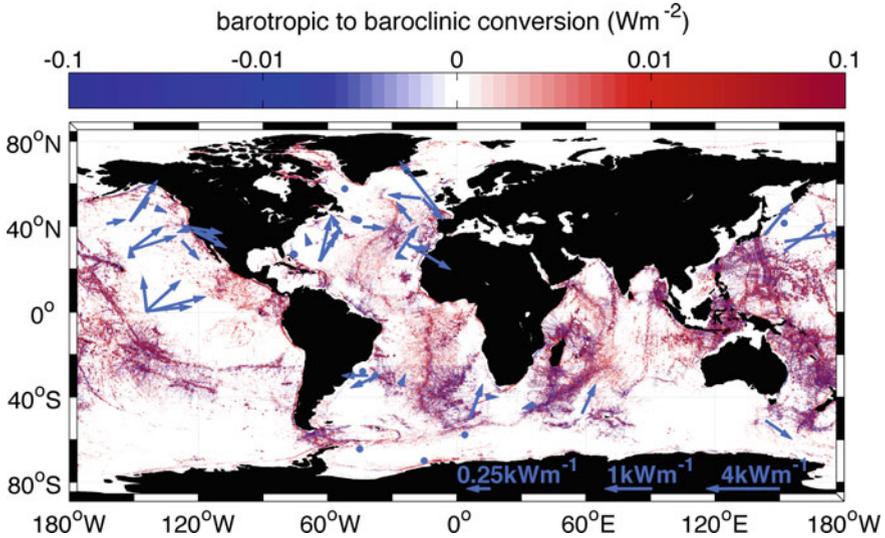


Fig. 4.1 Barotropic to baroclinic tidal energy conversion in W m^{-2} (color scale is logarithmic) for the semidiurnal M_2 tide from the high-resolution ocean circulation model STORMTIDE (Müller 2013). Arrows denote energy flux (taken from Alford 2003b) for low-mode internal tides from historical mooring records

Currents, the Gulf Stream, and the Kuroshio, the baroclinic velocities at M_2 frequency simulated by STORMTIDE represent indeed the M_2 internal tides.

In OGCMs in which internal tides are resolved, the tidal generation of internal waves can be derived from the barotropic to baroclinic tidal energy conversion (Kang and Fringer 2012). This approach is based on the energy equations for barotropic and baroclinic motions. The barotropic to baroclinic conversion rate, which couples the two energy equations, is determined by the product $\rho' W$ with ρ' being the perturbation density due to wave motions and W the barotropic vertical velocity. Following this approach, Müller (2013) estimated the energy conversion from tides (including M_2 , S_2 , K_1 , and O_1) to internal waves from the STORMTIDE simulation and found the tidal generation to be 1.2 TW (0.8 TW) over the ocean with depths greater than 1000 m (2000 m). The geographical distribution of this generation of internal waves from the M_2 tide is shown in Figure 4.1.

4.3 Dissipation

Low-mode internal wave energy can be dissipated in various ways. One mechanism to cause energy loss is non-linear interaction with the ambient internal wave field (e.g., parametric subharmonic instability, PSI) (McComas 1977; McComas and Bretherton 1977; Polzin and Lvov 2011). Candidates for enhanced dissipation are mid-basin

topographic features (e.g., Johnston et al. 2003), or continental slopes and shelves, where low-mode waves may reflect and/or scatter into higher mode waves that are more likely to break (Kelly et al. 2012, 2013). In general, wave energy flows toward higher vertical wave numbers and likely leads to wave breaking and dissipation (Lvov et al. 2010; McComas 1977; Olbers 1976). PSI is potentially important where the subharmonic of the M_2 tidal frequency is equal or smaller than the local inertial frequency (Nikurashin and Ferrari 2011), i.e., in the range between approximately 29° N and 29° S for the semidiurnal tides. Observational evidence was found in hydrographic and current measurements (Hibiya and Nagasawa 2004; Kunze et al. 2006), as well as in satellite altimetry (Tian et al. 2006). However, observations dedicated to PSI north of Hawaii found only slightly elevated mixing (MacKinnon et al. 2013) and no change in the amplitude of the mode 1 internal tide (Dushaw et al. 2011). Other processes are refraction of low-mode waves through, for instance, eddies and other mesoscale features.

In addition to the spatial inhomogeneity of dissipation, the wave field itself is variable. The forcing and the generation mechanisms for internal waves vary in time and space. Thus, it is plausible that they produce a regionally and temporally heterogeneous wave field (Polzin and Lvov 2011). There are only a few ocean observations of the temporal variability of mixing as a sink of energy—in a region with a tidal spring–neap cycle and topographically trapped waves (Shag Rocks Passage, Damerell et al. 2012), in the presence of eddies or variable background flow (North Atlantic Current, Walter and Mertens 2013), or of an annual cycle (Whalen et al. 2012)—but these observations show diapycnal diffusivities that vary by up to two orders of magnitude.

4.4 Observations

In situ observations of propagating low-mode internal waves are only sparsely available because they require time series measurements from a research vessel or mooring (Figure 4.2). While a mooring stays in the water for extended periods of up to several months, measurements from a ship are only carried out over few tidal cycles. Satellite altimetry measurements can be analyzed for barotropic and low-mode baroclinic tides but lack information about higher modes and incoherent signals. One experiment combined with models to estimate wave generation, fluxes, and dissipation was carried out in the South China Sea (Alford et al. 2015). An agreement within a factor of two between modeled and observed energy fluxes was found that allowed to establish an energy budget for the region that is supported by observations.

4.4.1 Satellite Altimetry

The long-term satellite altimetry measurements available today permit the extraction of the coherent part of the low-mode internal tide signals despite the repeat times of

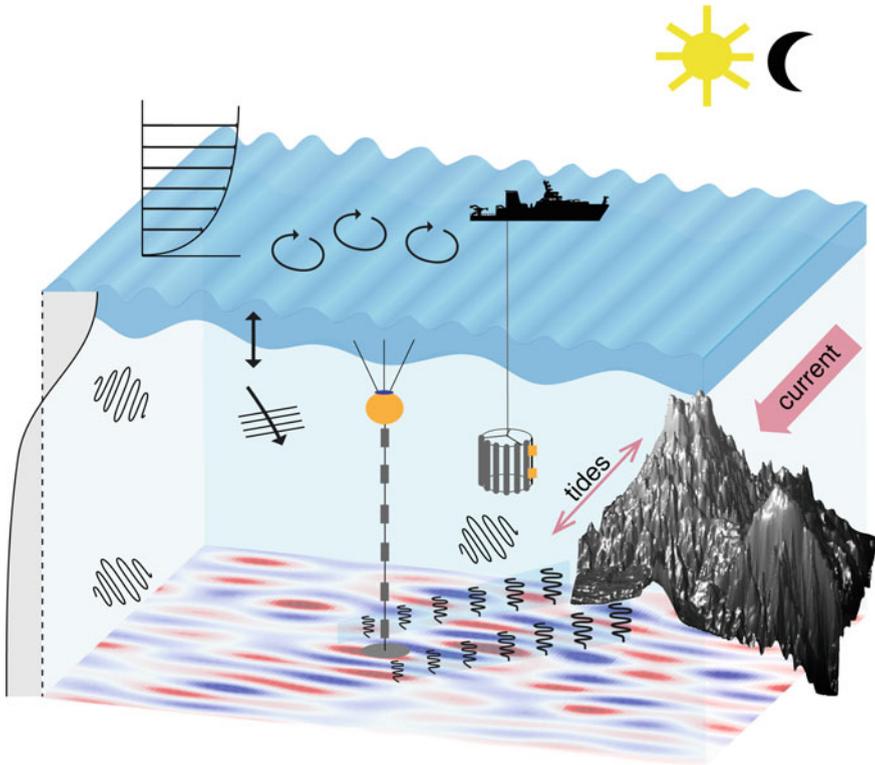


Fig. 4.2 Processes and observational techniques associated with the generation, propagation, and dissipation of near-inertial waves and internal tides. Internal tides are induced by barotropic tides over topographic features (depicted by the bathymetry of a seamount southeast of the Azores). Wind generates inertial oscillations in the surface mixed layer that result in the generation of near-inertial waves below. Both high and low modes are excited by wind and tides. Low-mode waves propagate long distances, while higher modes have stronger shear that results in local dissipation and mixing. The pattern of vertical displacement of an internal M2 tide as inferred from satellite altimetry is shown at the bottom of the figure (data provided by B. Dushaw, pers. comm.). Measurements of internal wave energy fluxes are carried out by moored instruments or by repeatedly lowering the instruments from a ship over the duration of one or two tidal cycles

several days as the tidal signals alias to longer periods (approximately 62 days for M2 from TOPEX/Poseidon).

Different methods are currently successfully used to derive sea surface displacements induced by internal tides (≈ 2 cm). Using TOPEX/Poseidon data, a spatially continuous frequency–wavenumber analysis was carried out by Dushaw et al. (2011) and a good agreement between the sea surface height expression of the mode 1 internal tide deduced from satellite altimetry and from acoustic tomography measurements, e.g., from the Hawaiian Ocean mixing experiment (HOME, Pinkel and Rudnick 2006) is found. This study was later extended to global coverage (Figure 4.3).

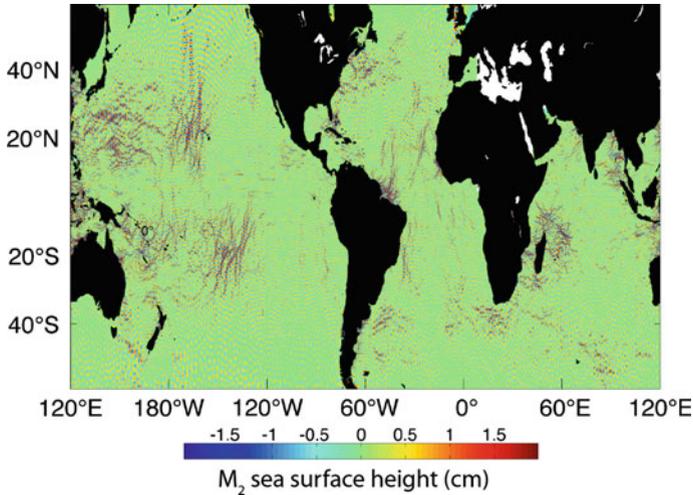


Fig. 4.3 Sea surface height amplitude (cm) of internal M_2 tide from AVISO satellite altimetry (Dushaw et al. 2011; Dushaw 2015). Data provided by B. Dushaw

Using local plane wave fits to TOPEX/Poseidon data, Zhao and Alford (2009) showed that the direction of their satellite altimetry derived energy fluxes in the central North Pacific is in general agreement with energy fluxes derived from mooring time series. Due to the low spatial resolution of about 250 km (Zhao and Alford 2009) of the satellite data, the flux magnitude was lower compared to energy fluxes from mooring time series or the Princeton Ocean Model (POM). Ray and Zaron (2016) and Zhao et al. (2011) use data from multiple satellites in an empirical mapping approach and a local plane wave fit method in $120 \text{ km} \times 120 \text{ km}$ subregions, respectively. The use of multi-satellite altimetry improves the spatial resolution of internal wave energy fluxes calculated from the local plane wave method and their agreement with the POM as shown in a regional study around the Hawaiian Ridge (Zhao et al. 2011). Recently, this method was extended to global coverage (Zhao et al. 2016) and compared to historical mooring data and model results.

The study from Zhao and Alford (2009) revealed internal tidal beams as narrow as 200 km and stretching more than 3500 km across the North Pacific Ocean. Off the Hawaiian Ridge, Zhao et al. (2010) combined altimeter data with six profiling moorings and showed that the coherent mode 1 internal tide dominated the energy flux over the whole distance (1400 km) of the array. However, at the northern end of the array phase differences were found and attributed to refraction by mesoscale flows. A decrease of mode 1 internal tide amplitude by less than 20% over the 2000 km between the Hawaiian Ridge at 20° N and 40° N was found by Dushaw et al. (2011). In general, the satellite observed internal tides compare well with mooring measurements and numerical models. The studies so far indicate that low-mode internal tides propagate hundreds to thousands of kilometers before they dissipate.

4.4.2 Shipboard Observations

Repeated top-to-bottom CTD and lowered ADCP profiles, taken for periods of 24–36 h to capture semidiurnal to diurnal and inertial signals, can be used to measure low-mode internal wave energy fluxes. The data are decomposed into mean, diurnal, and semidiurnal components by harmonic analysis of velocity and displacement at each depth, and the energy flux is then computed (Nash et al. 2005). The separation of energy flux constituents from short time series can be difficult. However, in their analysis of shipboard observations in Luzon Strait Alford et al. (2011) found that the total energy flux measured nearly equaled the sum of diurnal and semidiurnal components.

Turbulent dissipation can also be determined from shipboard observations and related to the observed loss in internal wave energy. Direct measurements of turbulent dissipation or diapycnal mixing with microstructure instrumentation are difficult and expensive, and thus sparse. Therefore, finescale parameterizations are used to estimate the diapycnal diffusivity from routinely observed properties (Gregg et al. 2003; Kunze et al. 2006): the vertical shear of horizontal flow and strain of the density field. These empirical methods rely on the observation that the variance of these properties on the finescale (order of tens to hundreds of meters), the so-called finestructure variance, is related to energy dissipation and mixing (Gregg 1989). The underlying assumption is that in steady state the internal wave spectral energy on the finescale is subject to a continuous spectral transport toward higher wavenumbers.

Using finescale parametrization with large data sets, e.g., North Atlantic shipboard observations (Figure 4.4, Walter et al. 2005), the WOCE hydrographic data (Kunze et al. 2006), or ARGO floats (Whalen et al. 2012), have led to a substantial increase in the database of diapycnal mixing estimates. Regional and global comparisons between these results and microstructure measurements showed that the diapycnal diffusivities based on the finescale parameterization are on average accurate within a factor of 2–3 (Polzin et al. 2014; Waterhouse et al. 2014), but can locally deviate from microstructure measurements by a factor of up to 10 (e.g., Waterman et al. 2013).

An alternative method of using hydrographic (temperature/density) data to estimate energy dissipation and diapycnal diffusivities is the overturn or Thorpe scale method (Thorpe 1977). Here, instabilities detected in the density profiles are used to determine characteristic length scales for individual turbulent patches, which then are used to calculate the dissipation rate ε . However, in regions of weak stratification, the detections of instabilities are impaired by the poor signal-to-noise ratio, and therefore, the uncertainties associated with diffusivity estimates derived with the overturn method increase (Frants et al. 2013). For mixing regimes that are dominantly driven by convection, the overturn method may overestimate mixing (Mater et al. 2015).

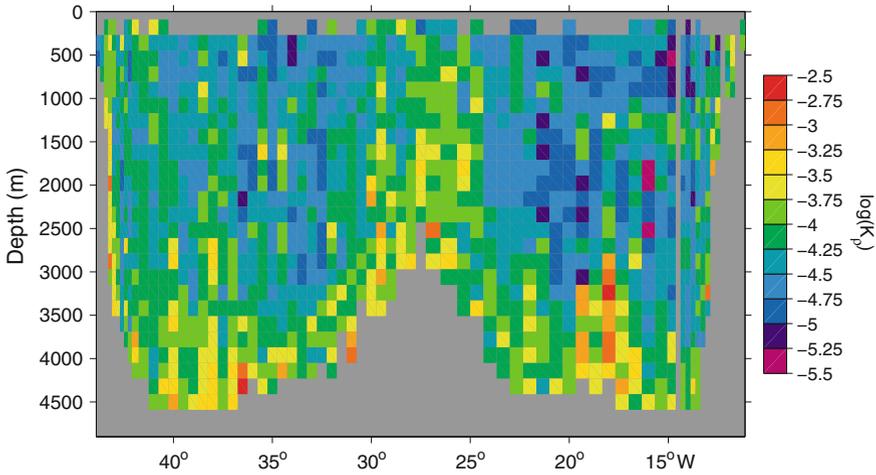


Fig. 4.4 Diapycnal diffusivity K_ρ ($\text{m}^2 \text{s}^{-1}$) estimated from repeated finescale measurements of velocity shear and density strain along a section across the North Atlantic at about 47°N . The calculation was carried out as described in Walter et al. (2005)

4.4.3 Moorings

Altimetric estimates of the internal tide cannot detect the portion of the signal that has lost coherence with astronomic forcing, and near-inertial waves have no coherent component at all. Therefore, time series measurements that have a higher temporal resolution are required to study the variability of phase, energy, and fluxes of this part of the wave field.

Using 80 globally distributed moored data records mainly from the 1980s and 1990s, Alford (2003b) showed that fluxes of the low-mode internal tides are generally away from regions of rough topography (Figure 4.1). Historical mooring data are archived by the Oregon State University Buoy Group. The archive contains approximately 2000 current records from the World Ocean Circulation Experiment (WOCE) and more than 7000 non-WOCE records. Using their criterion of at least four current meters in a specific depth range, Alford (2003b) found only 80 moorings with a sampling interval of 3 h or less and a record length of at least 180 days. They bandpass-filtered the historical velocity and temperature-depth data to extract the near-inertial and semidiurnal components and obtained full-depth profiles of modal velocity and displacement for the first two modes by performing a least squares inverse for each time step. The analysis showed that the fluxes in both frequency bands are of the order of 1 kW m^{-1} and directed away from their respective source regions (Figure 4.1). The tidal fluxes are uniform throughout the year, while the near-inertial fluxes are generally stronger in winter.

4.5 Summary and Outlook

Because of the cascades of dynamical processes involved before the low-mode internal waves break and lead to mixing, and the different—and partly unknown—geographical distribution of generation and dissipation, it is evident that the parameterization of mixing in ocean circulation and climate models cannot simply be a constant vertical diffusion coefficient as shown by Olbers and Eden (2013). As mentioned before, the MOC, the deep ocean stratification, and the distribution of passive tracers (for instance anthropogenic carbon) respond to changing patterns of imposed deep mixing. Olbers and Eden (2013) and Eden and Olbers (2014) propose an energetically consistent mixing module dubbed Internal Wave Dissipation, Energy and Mixing (IDEMIX) that describes the generation, interaction, propagation, and dissipation of internal gravity wave energy, dissipation rates, and diapycnal mixing (Figure 4.5).

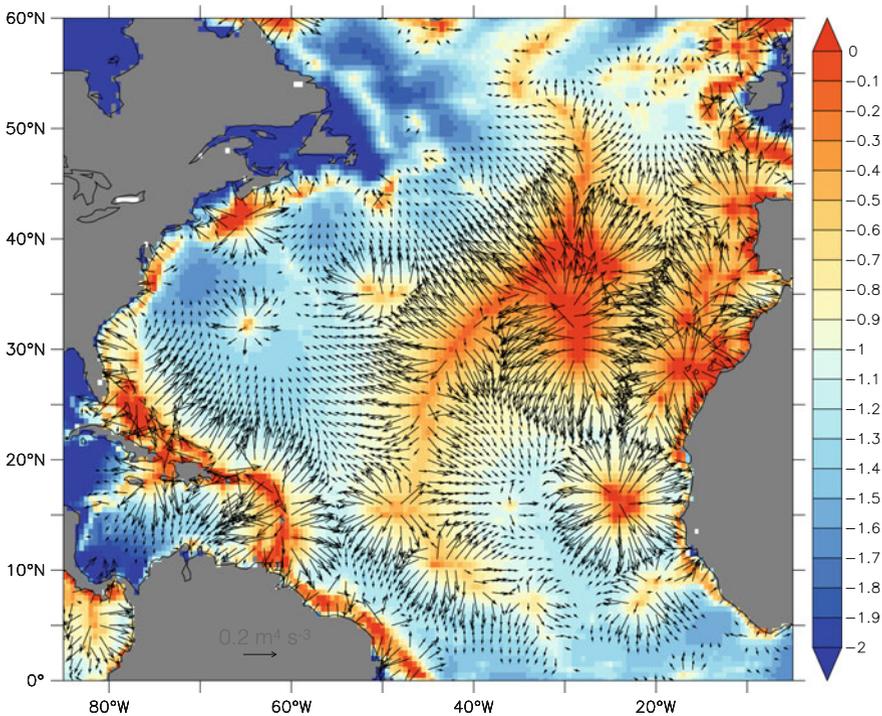


Fig. 4.5 Energy flux of baroclinic M_2 tide ($\text{m}^4 \text{s}^{-3}$) and vertically integrated energy content ($\log E/\text{m}^2 \text{s}^{-2}$) in the North Atlantic from IDEMIX (C. Eden, pers. comm.). The arrow length is limited to $0.2 \text{ m}^4 \text{ s}^{-3}$, and fluxes with amplitudes less than $0.02 \text{ m}^4 \text{ s}^{-3}$ are not shown. The fluxes appear isotropic due to the boundary conditions; directionality is not yet implemented

Observations to study the propagation and dissipation of low-mode internal waves are quite sparse. The main processes are known: non-linear wave–wave interaction (PSI), interaction with the mesoscale shear/vorticity, scattering over deep ocean topography or breaking at continental slopes, but how much of the energy is going into each of the processes is not clear (MacKinnon et al. 2013).

The quantification of generation and propagation of internal waves in the global ocean, the pathways of radiated low-mode internal waves including processes operating along the pathways, and the identification of regions of sources and sinks are challenging tasks. Furthermore, the quantification of the contribution to local dissipation and identification of the involved processes remains difficult. So far, these tasks have been often tackled by either using observations or numerical models. In particular, the observations have not been used to evaluate the estimates based on numerical model simulations. Moreover, it is a common practice to study different mechanisms within different frameworks. For instance, the wind generation has been studied exclusively in isolation from the tidal generation of internal waves. The goal of this project is to unify or to connect different frameworks and to combine or to evaluate model simulations with observations. The former will concentrate on the development of an OGCM that operates not only with high-frequency winds but also with the full tidal potential so that the wave generation through both winds and tides can be studied in a dynamically consistent way. The latter will concentrate on wave energy fluxes with the goal of producing the best estimate of the global distributions of sources and sinks needed for validation and improvement of parameterizations and ultimately to the development of energetically consistent climate models.

Acknowledgements We thank Brian Dushaw (Univ. Washington) for providing estimates of mode 1 internal tides from altimetry and Carsten Eden (Univ. Hamburg) for the energy fluxes from IDEMIX. Janna Köhler received funding from the Deutsche Forschungsgemeinschaft (DFG, grant Ko4814/1-1).

References

- Alford, M.H.: Improved global maps and 54-year history of wind-work on ocean inertial motions. *Geophys. Res. Lett.* **30**(8), 1424 (2003a)
- Alford, M.H.: Redistribution of energy available for ocean mixing by long-range propagation of internal waves. *Nature* **423**, 159–162 (2003b)
- Alford, M.H., Zhao, Z.: Global patterns of low-mode internal-wave propagation. Part I: energy and energy flux. *J. Phys. Oceanogr.* **37**, 1829–1848 (2007)
- Alford, M.H., MacKinnon, J.A., Nash, J.D., Simmons, H., Pickering, A., Klymak, J.M., Pinkel, R., Sun, O., Rainville, L., Mushgrave, R., Beitzel, T., Fu, K.-H., Lu, C.-W.: Energy flux and dissipation in Luzon Strait: two tales of two ridges. *J. Phys. Oceanogr.* **41**, 2211–2222 (2011)
- Alford, M.H., Cronin, M.F., Klymak, J.M.: Annual cycle and depth penetration of wind-generated near-inertial internal waves at ocean station Papa in the Northeast Pacific. *J. Phys. Oceanogr.* **42**, 889–909 (2012)
- Alford, M.H., Peacock, T., MacKinnon, J.A., Nash, J.D., Buijsman, M.C., Centuroni, L.R., Chao, S.-Y., Chang, M.-H., Farmer, D.M., Fringer, O.B., Fu, K.-H., Gallacher, P.C., Graber, H.C., Helfrich, K.R., Jachec, S.M., Jackson, C.R., Klymak, J.M., Ko, D.S., Jan, S., Johnston, T.M.S.,

- Legg, S., Lee, I.-H., Lien, R.-C., Mercier, M.J., Moum, J.N., Musgrave, R., Park, J.-H., Pickering, A.L., Pinkel, R., Rainville, L., Ramp, S.R., Rudnick, D.L., Sarkar, S., Scotti, A., Simmons, H.L., St. Laurent, L. C., Venayagamoorthy, S. K., Wang, Y.-H., Wang, J., Yang, Y. J., Paluszkiwicz, T., Tang, T.-Y.: The formation and fate of internal waves in the South China Sea. *Nature* **521**, 65–69 (2015)
- Arbic, B.K., Wallcraft, A.J., Metzger, E.J.: Concurrent simulation of the eddy general circulation and tides in a global ocean model. *Ocean Model.* **32**, 175–187 (2010)
- Arbic, B.A., Scott, R.B., Chelton, D.B., Richman, J.G., Shriver, J.F.: Effects of stencil width on surface ocean geostrophic velocity and vorticity estimation from gridded satellite altimeter data. *J. Geophys. Res.* **117**, C03029 (2012)
- Bell, T.: Statistical features of sea-floor topography. *Deep-Sea Res.* **22**, 883–892 (1975a)
- Bell, T.: Topographically generated internal waves in the open ocean. *J. Geophys. Res.* **80**(3), 320–327 (1975b)
- Damerell, G.M., Heywood, K.J., Stevens, D.P., Naveira Garabato, A.C.: Temporal variability of diapycnal mixing in Shag Rocks Passage. *J. Phys. Oceanogr.* **42**, 370–385 (2012)
- D’Asaro, E.A., Eriksen, C.C., Levine, M.D., Niiler, P., Paulson, C.A., Meurs, P.V.: Upper-ocean inertial currents forced by a strong storm. Part I: data and comparison with linear theory. *J. Phys. Oceanogr.* **25**, 2909–2936 (1995)
- Dushaw, B.: An empirical model for mode-1 internal tides derived from satellite altimetry: computing accurate tidal predictions at arbitrary points over the world oceans. Technical Memorandum **1–15**, APL-UW (2015)
- Dushaw, B.D., Worcester, P.F., Dzieciuch, M.A.: On the predictability of mode-1 internal tides. *Deep-Sea Res.* **I**(58), 677–698 (2011)
- Eden, C., Czeschel, L., Olbers, D.: Toward energetically consistent ocean models. *J. Phys. Oceanogr.* **44**, 3160–3184 (2014)
- Eden, C., Olbers, D.: An energy compartment model for propagation, non-linear interaction and dissipation of internal gravity waves. *J. Phys. Oceanogr.* **44**, 2093–2106 (2014)
- Egbert, G., Erofeeva, S.: Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Ocean. Technol.* **19**, 183–204 (2002)
- Egbert, G.D., Ray, R.D.: Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data. *Nature* **405**, 775–778 (2000)
- Exarchou, E., von Storch, J.-S., Jungclaus, J.H.: Impact of tidal mixing with different scales of bottom roughness on the general circulation. *Ocean Dyn.* **62**, 1545–1563 (2012)
- Exarchou, E., von Storch, J.-S., Jungclaus, J.H.: Sensitivity of transient climate change to tidal mixing: southern ocean heat uptake in climate change experiments performed with ECHAM5/MPIOM. *Clim. Dyn.* **42**, 1755–1773 (2014)
- Frants, M., Damerell, G.M., Gille, S.T., Heywood, K.J., MacKinnon, J., Sprintall, J.: An assessment of density-based finescale methods for estimating diapycnal diffusivity in the Southern Ocean. *J. Atmos. Ocean. Technol.* **30**, 2647–2661 (2013)
- Furuichi, N., Hibiya, T., Niwa, Y.: Model-predicted distribution of wind-induced internal wave energy in the world’s oceans. *J. Geophys. Res.* **113**, C09034 (2008)
- Gregg, M.C.: Scaling turbulent dissipation in the thermocline. *J. Geophys. Res.* **94**(C7), 9686–9698 (1989)
- Gregg, M.C., Sanford, T.B., Winkel, D.P.: Reduced mixing from the breaking of internal waves in equatorial waters. *Nature* **422**, 513–515 (2003)
- Gustafsson, K.E.: Computations of the energy flux to mixing processes via baroclinic wave drag on barotropic tides. *Deep-Sea Res.* **I**(48), 2283–2295 (2001)
- Hallberg, R.: Stable split time stepping schemes for large-scale ocean modeling. *J. Comput. Sci.* **135**, 54–65 (1997)
- Hibiya, T., Nagasawa, M.: Latitudinal dependence of diapycnal diffusivity in the thermocline estimated using a finescale parameterization. *Geophys. Res. Lett.* **31**(L01301) (2004). <https://doi.org/10.1029/2003GL017998>

- Jiang, J., Lu, Y., Perrie, W.: Estimating the energy flux from the wind to ocean inertial motions: the sensitivity to surface wind fields. *Geophys. Res. Lett.* **32**, L15610 (2005)
- Johnston, T.M.S., Merrifield, M.A., Holloway, P.E.: Internal tide scattering at the line islands ridge. *J. Geophys. Res.* **108**(C11), 3365 (2003)
- Kang, D., Fringer, O.: Energetics of barotropic and baroclinic tides in the monterey bay area. *J. Phys. Oceanogr.* **42**, 272–290 (2012)
- Kelly, S.M., Nash, J.D., Martini, K.I., Alford, M.H., Kunze, E.: The cascade of tidal energy from low to high modes on a continental slope. *J. Phys. Oceanogr.* **42**, 1217–1232 (2012)
- Kelly, S.M., Jones, N.L., Nash, J.D., Waterhouse, A.F.: The geography of semidiurnal mode-1 internal-tide energy loss. *Geophys. Res. Lett.* **40**, 4689–4693 (2013)
- Klymak, J.M., Pinkel, R., Rainville, L.: Direct breaking of the internal tide near topography: Kaena Ridge, Hawaii. *J. Phys. Oceanogr.* **38**, 380–399 (2008)
- Kunze, E., Firing, E., Hummon, J.M., Chereskin, T.K., Thurnherr, A.M.: Global abyssal mixing inferred from lowered ADCP shear and CTD strain profiles. *J. Phys. Oceanogr.* **36**(8), 1553–1576 (2006)
- Li, Z., von Storch, J.-S., Müller, M.: The M_2 internal tide simulated by a $1/10^\circ$ OGCM. *J. Phys. Oceanogr.* **45**, 3119–3135 (2015)
- Llewellyn Smith, S.G., Young, W.R.: Conversion of the barotropic tide. *J. Phys. Oceanogr.* **32**(5), 1554–1566 (2002)
- Lvov, Y.V., Polzin, K.L., Tabak, E.G., Yokoyama, N.: Oceanic internal-wave field: theory of scale-invariant spectra. *J. Phys. Oceanogr.* **40**, 2605–2623 (2010)
- MacKinnon, J., St. Laurent, L., Naveira Garabato, A.C.: Diapycnal mixing processes in the ocean interior. *Int. Geophys.* **103**, 159–183 (chapter 7). Elsevier Ltd (2013)
- Mater, B.D., Venayagamoorthy, S.K., St. Laurent, L., Moum, J.N.: Biases in Thorpe-scale estimates of turbulent dissipation. Part I: assessments from large-scale overturns in oceanographic data. *J. Phys. Oceanogr.* **45**, 2497–2521 (2015)
- McComas, C.H.: Equilibrium mechanics within the oceanic internal wave field. *J. Phys. Oceanogr.* **7**, 836–845 (1977)
- McComas, C.H., Bretherton, F.P.: Resonant interaction of oceanic internal waves. *J. Geophys. Res.* **82**, 1397–1412 (1977)
- Melet, A., Hallberg, R., Legg, S., Polzin, K.: Sensitivity of the ocean state to the vertical distribution of internal-tide-driven mixing. *J. Phys. Oceanogr.* **43**, 602–615 (2013)
- Melet, A., Hallberg, R., Legg, S., Nikurashin, M.: Sensitivity of the ocean state to lee wave-driven mixing. *J. Phys. Oceanogr.* **44**, 900–921 (2014)
- Mellor, G.L.: User Guide for a Three-Dimensional, Primitive Equation. Numerical Ocean Model. Princeton University, Princeton, N.J (2003)
- Müller, M.: On the space- and time-dependence of barotropic-to-baroclinic tidal energy conversion. *Ocean Model.* **72**, 242–252 (2013)
- Müller, M., Cherniawsky, J.Y., Foreman, M.G.G., von Storch, J.-S.: Global M_2 internal tide and its seasonal variability from high resolution ocean circulation and tide modeling. *Geophys. Res. Lett.* **39**, L19607 (2012)
- Munk, W., Wunsch, C.: Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res.* **I**(45), 1977–2010 (1998)
- Nash, J.D., Alford, M.H., Kunze, E.: Estimating internal wave energy fluxes in the ocean. *J. Atmos. Ocean. Technol.* **22**, 1551–1570 (2005)
- Nikurashin, M., Ferrari, R.: Global energy conversion rate from geostrophic flows into internal lee waves in the deep ocean. *Geophys. Res. Lett.* **38**, L08610 (2011)
- Nikurashin, M., Ferrari, R.: Overturning circulation driven by breaking internal waves in the deep ocean. *Geophys. Res. Lett.* **40**, 3133–3137 (2013)
- Nycander, J.: Generation of internal waves in the deep ocean by tides. *J. Geophys. Res.* **110**, C10028 (2005)
- Olbers, D., Eden, C.: A global model for the diapycnal diffusivity induced by internal gravity waves. *J. Phys. Oceanogr.* **43**, 1759–1779 (2013)

- Olbers, D.J.: Nonlinear energy transfer and the energy balance of the internal wave field in the deep ocean. *J. Fluid Mech.* **74**, 375–399 (1976)
- Pinkel, R., Rudnick, D.: Editorial: Hawaii ocean mixing experiment (HOME). *J. Phys. Oceanogr.* **36**, 965–966 (2006)
- Plueddemann, A.J., Farrar, J.T.: Observations and models of the energy flux from the wind to mixed-layer inertial currents. *Deep-Sea Res.* **I**(53), 5–30 (2006)
- Pollard, R.T., Millard Jr., R.C.: Comparison between observed and simulated wind-generated intertidal oscillations. *Deep-Sea Res.* **17**, 813–821 (1970)
- Polzin, K.L., Lvov, Y.V.: Toward regional characterizations of the oceanic internal wavefield. *Rev. Geophys.* **49**, RG4003 (2011)
- Polzin, K.L., Naveira Garabato, A.C., Huussen, T.N., Sloyan, B.M., Waterman, S.: Finescale parameterizations of turbulent dissipation. *J. Geophys. Res.* **119**, 1383–1419 (2014)
- Ray, R.D., Zaron, E.D.: M_2 internal tides and their observed wavenumber spectra from satellite altimetry. *J. Phys. Oceanogr.* **46**, 3–22 (2016)
- Rimac, A., von Storch, J.-S., Eden, C., Haak, H.: The influence of high-resolution wind stress field on the power input to near-inertial motions in the ocean. *Geophys. Res. Lett.* **40**, 4882–4886 (2013)
- Rimac, A., von Storch, J.-S., Eden, C.: The total energy flux leaving the ocean's mixed layer. *J. Phys. Oceanogr.* **46**, 1885–1900 (2016)
- Simmons, H.L., Alford, M.H.: Simulating the long-range swell of internal waves generated by ocean storms. *Oceanography* **25**, 30–41 (2012)
- Simmons, H.L., Hallberg, R.W., Arbic, B.K.: Internal wave generation in a global baroclinic tide model. *Deep-Sea Res.* **II**(51), 3043–3068 (2004)
- Sjöberg, B., Stigebrandt, A.: Computation of the geographical distribution of the energy flux to mixing processes via internal tides and the associated vertical circulation in the ocean. *Deep-Sea Res.* **I**(39), 269–291 (1992)
- St. Laurent, L.C., Simmons, H.L., Jayne, S.R.: Estimating tidally driven mixing in the deep ocean. *Geophys. Res. Lett.* **23**(2106) (2002). <https://doi.org/10.1029/2002GL015633>
- Thorpe, S.A.: Turbulence and mixing in a scottish loch. *Phil. Trans. R. Soc. Lond.* **286**(1334), 125–181 (1977)
- Tian, J., Zhou, L., Zhang, X.: Latitudinal distribution of mixing rate caused by the m_2 internal tide. *J. Phys. Oceanogr.* **36**, 35–42 (2006)
- Walter, M., Mertens, C.: Mid-depth mixing linked to North Atlantic current variability. *Geophys. Res. Lett.* **40**, 4869–4875 (2013)
- Walter, M., Mertens, C., Rhein, M.: Mixing estimates from a large-scale hydrographic survey in the North Atlantic. *Geophys. Res. Lett.* **32**, L13605 (2005)
- Watanabe, M., Hibiya, T.: Global estimates of the wind-induced energy flux to inertial motions in the surface mixed layer. *Geophys. Res. Lett.* **29**(8), 1239 (2002)
- Waterhouse, A.F., MacKinnon, J.A., Nash, J.D., Alford, M.H., Kunze, E., Simmons, H.L., Polzin, K.L., St. Laurent, L., Sun, O.M., Pinkel, R., Talley, L.D., Whalen, C.B., Huussen, T.N., Carter, G.S., Fer, I., Waterman, S., Naveira Garabato, A.C., Sanford, T.B., Lee, C.M.: Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate. *J. Phys. Oceanogr.* **44**, 1854–1872 (2014)
- Waterman, S., Naveira Garabato, A.C., Polzin, K.L.: Internal waves and turbulence in the Antarctic circumpolar current. *J. Phys. Oceanogr.* **43**, 259–282 (2013)
- Whalen, C.B., Talley, L.D., MacKinnon, J.A.: Spatial and temporal variability of global ocean mixing inferred from argo profiles. *Geophys. Res. Lett.* **39**, L18612 (2012)
- Wunsch, C.: Internal tides in the ocean. *Rev. Geophys. Space Phys.* **13**, 167–1982 (1975)
- Wunsch, C., Ferrari, R.: Vertical mixing, energy, and the general circulation of the oceans. *Annu. Rev. Fluid Mech.* **36**, 281–314 (2004)
- Zhao, Z., Alford, M.H.: New altimetric estimates of mode-1 M_2 internal tides in the central north Pacific Ocean. *J. Phys. Oceanogr.* **39**, 1669–1684 (2009)

- Zhao, Z., Alford, M.H., MacKinnon, J.A., Pinkel, R.: Long-range propagation of the semidiurnal tide from the Hawaiian Ridge. *J. Phys. Oceanogr.* **40**, 713–736 (2010)
- Zhao, Z., Alford, M.H., Girton, J., Johnston, T.M.S., Carter, G.: Internal tides around the Hawaiian Ridge estimated from multisatellite altimetry. *J. Geophys. Res.* **116**, C12039 (2011)
- Zhao, Z., Alford, M.H., Girton, J.B., Rainville, L., Simmons, H.L.: Global observations of open-ocean mode-1 M_2 internal tides. *J. Phys. Oceanogr.* **46**, 1657–1684 (2016)

Chapter 5

Toward Consistent Subgrid Momentum Closures in Ocean Models



Sergey Danilov, Stephan Juricke, Anton Kutsenko and Marcel Oliver

Abstract State-of-the-art global ocean circulation models used in climate studies are only passing the edge of becoming “eddy-permitting” or barely eddy-resolving. Such models commonly suffer from overdissipation of mesoscale eddies by routinely used subgrid dissipation (viscosity) operators and a resulting depletion of energy in the large-scale structures which are crucial for draining available potential energy into kinetic energy. More broadly, subgrid momentum closures may lead to both overdissipation or pileup of eddy kinetic energy and enstrophy of the smallest resolvable scales. The aim of this chapter is twofold. First, it reviews the theory of two-dimensional and geostrophic turbulence. To a large part, this is textbook material with particular emphasis, however, on issues relevant to modeling the global ocean in the eddy-permitting regime. Second, we discuss several recent parameterizations of subgrid dynamics, including simplified backscatter schemes by Jansen and Held, stochastic superparameterizations by Grooms and Majda, and an empirical backscatter scheme by Mana and Zanna.

5.1 Introduction

Climate studies require model simulations over periods from centuries to millenia, which are only affordable if ocean models are kept relatively coarse. Many of them stay at a resolution of about one degree and need to parameterize the effect of unresolved mesoscale eddies and smaller-scale motions. The issue of mesoscale eddy

S. Danilov

Alfred Wegener Institute for Polar and Marine Research (AWI), Bremerhaven, Germany
e-mail: sergey.danilov@awi.de

S. Danilov · S. Juricke · A. Kutsenko · M. Oliver (✉)

Jacobs University, Bremen, Germany
e-mail: oliver@member.ams.org; m.oliver@jacobs-university.de

S. Juricke

e-mail: s.juricke@jacobs-university.de

A. Kutsenko

e-mail: a.kutsenko@jacobs-university.de

© Springer Nature Switzerland AG 2019

C. Eden and A. Iske (eds.), *Energy Transfers in Atmosphere and Ocean*,

Mathematics of Planet Earth 1, https://doi.org/10.1007/978-3-030-05704-6_5

parameterization attracts continuing interest as exemplified by recent studies on eddy potential vorticity fluxes (Marshall and Adcroft 2010; Eden 2010; Ringler and Gent 2011; Marshall et al. 2012). With increasing computational power, eddy-permitting (barely resolving the first baroclinic Rossby radius) or eddy-resolving models are becoming feasible for climate studies, too, so that mesoscale dynamics will gradually be resolved. Nonetheless, as the first baroclinic Rossby radius varies widely (with values below 10 km in high latitudes), even eddy-resolving models will not necessarily represent eddy dynamics with the same skill everywhere unless their resolution is on the scale of a few kilometers. Combining resolved mesoscale dynamics in some parts of the ocean with parameterized dynamics in the other part is an interesting possibility, but cautionary results by Hallberg (2013) indicate that the transition from parameterized eddies to resolved eddies can introduce problems of its own.

Even though state-of-the-art eddy-permitting or eddy-resolving models simulate the mesoscale dynamics with a certain skill, they still use some form of explicit and/or implicit viscosity, thought to represent the effect of unresolved small-scale subgrid dynamics. The motivation is based on the picture of quasigeostrophic turbulence (Charney 1971), which indicates that the direct cascade of enstrophy has to be removed at the grid scale to prevent the enstrophy from piling up, causing code instability. Fox-Kemper and Menemenlis (2008) discuss common approaches used in oceanographic practice, in particular the Smagorinsky or Leith parameterizations in either harmonic or biharmonic implementation. While these ideas appear plausible, there are no solid theoretical arguments, especially outside the limits of applicability of quasigeostrophic theory which is questionable at grid scale.

The detailed form of the subgrid operators (e.g., Laplacian vs. biharmonic viscosity), however, is known to impact the large-scale dynamics such as the path and separation of the Gulf Stream (Hecht et al. 2008b). Moreover, removal of enstrophy at the grid scale is accompanied by energy dissipation. For example, Danilov (2005) has shown that a direct enstrophy cascade in two-dimensional turbulence is always associated with a noticeable direct energy cascade, resulting in dissipation at finite resolution. Jansen and Held (2014) point out that the popular biharmonic viscosity operator suppresses resolved eddy motion in models where the separation between the mesoscale and the grid scale is insufficient. This reduces the ability of the flow to drain eddy kinetic energy from the available potential energy (APE), thereby distorting the entire energy cycle. The effect is most pronounced for eddy-permitting models where the grid scale and the scale of APE release are not well separated. It is also important for eddy-resolving models as the first baroclinic Rossby radius may locally drop below grid scale. The remedies are less immediate and open for investigation; searching for them is the main aim of ongoing studies. There is growing interest in this topic in the community as eddy-permitting models are now beginning to be used in climate research, so that the question of how to make them more realistic becomes pressing; see, e.g., Jansen et al. (2015), Berloff (2015), Cooper and Zanna (2015), and Cooper (2017).

Analyzing the effects of spectral pileup and backscatter of eddy energy in response to common subgrid parameterizations is rather straightforward for simple two-dimensional flows with a prescribed kinetic energy production rate (Graham and

Ringler 2013), but the question remains open in the context of more realistic dynamics which includes the effects of baroclinicity and where the geometry of boundaries and topography makes spectral analysis only locally applicable. Moreover, in real flows the balance between energy production and dissipation ceases to be local, which further complicates the situation. It is not clear how the subgrid operators affect the energy exchange between balanced (quasigeostrophic) and non-balanced motions as the resolution is increasing. More broadly, the mathematical side of subgrid parameterization as used in oceanographic tasks needs a more firm basis which would dictate a scale- and frame-invariant structure for admissible parameterizations.

It is important to note that the dynamics on scales close to the grid scale is affected not only by explicit subgrid parameterizations, but also by details of the discretization of momentum advection (see Figure 5.1). For example, high-order upwind transport algorithms based on the flux form of the advection operator have implicit numerical dissipation of the same order of magnitude as typical explicit dissipation (see, e.g., Mohammadi-Aragh et al. 2015). Further, there is evidence for a numerical (Hollingsworth) instability associated with the vector-invariant form of momentum advection which creates noise in the vertical velocity field and thus influences the APE to kinetic energy conversion; see the discussion in Gassmann (2013) and Danilov and Wang (2015). Understanding the effects induced by these or

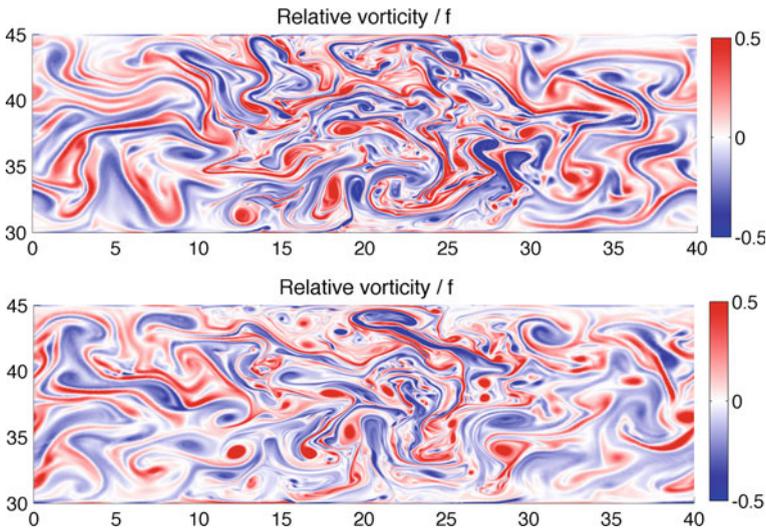


Fig. 5.1 Effect of momentum advection discretization on the relative vorticity field in a baroclinically unstable channel flow (top: vector-invariant form, bottom: flux form; near-surface snapshots are shown). Mesh resolution varies ($1/36$ degree in the central part and coarser elsewhere). Observed scales and amplitude of small eddies in the central part differ substantially between the two schemes due to the difference in implicit dissipation and discretization residual. The variance of vertical velocity (not shown) is substantially lower for the flux form, modifying the APE to eddy KE conversion rate. The figure is based on simulations reported in Danilov and Wang (2015)

other numerical details on the energy balance and accounting for their interplay with subgrid parameterizations is a necessary element on the road to rigorous analysis.

The need to explore the interplay between resolution, parameterized subgrid, and spurious numerical dissipation is particularly important for future earth system models employing multi-resolution technology, for example models based on FESOM (Wang et al. 2014) or ICON (Korn 2017). Recent results point to the retardation of eddy saturation when the upstream resolution is coarse (Danilov and Wang 2015). In multi-resolution models on unstructured meshes, subgrid momentum closures are also needed to stabilize against spurious numerical modes appearing on staggered triangular meshes (Danilov 2013), which adds numerical complexity.

For all the reasons mentioned, the question of how to return the overdissipated energy to the resolved scales is of central importance when working at eddy-permitting resolutions. This is known as the *energy backscatter* problem. On coarser meshes, one needs to additionally parameterize the contribution from mesoscale eddies. In both cases, there is growing interest in stochastic parameterizations. Stochastic parameterizations have been successfully used to maintain sufficient variance in ensemble forecasts (Palmer et al. 2009). However, energy and momentum consistency especially over long simulation timescales have not received as much attention (cf. the discussion in Franzke et al. 2015).

For the momentum closure problem in the ocean, stochastic parameterizations hold promise far beyond the idea of pure dissipation pursued by traditional deterministic subgrid parameterizations and also beyond downgradient parameterizations for unresolved mesoscale eddies. In the ocean context, systematic work on stochastic parameterizations is rather recent: Duan and Nadiga (2007), Mana and Zanna (2014), Jansen and Held (2014), Grooms et al. (2015b), Cooper and Zanna (2015), Cooper (2017), and Berloff (2015) all implement backscatter as stochastic forcing acting on the resolved flow, showing the potential of the approach, but also raising questions about the structure of this forcing and the choice of parameters. Stochastic backscatter can be implemented in a purely statistical way; more sophisticated approaches seek to include dynamical information, for example by shaping the backscatter forcing according to the nonlinear self-interaction derived from elementary solutions to the tangent linear equation (Berloff 2015, 2016).

Further open questions pertain to finding a more general mathematical framework and generalizations away from a quasigeostrophic setting toward the full primitive equations; work in this direction is at the very beginning. Related work on stochastic LES closures for the Navier–Stokes equations was done by San (2014) for the two-dimensional problem in vorticity form and by Xun and Wang (2014) for channel flow in three dimensions. Jansen and Held (2014) show that backscatter for the two-layer quasigeostrophic equations can be parameterized both stochastically and deterministically, with very similar results. Their approach is generalized to a simplified primitive equation isopycnal model in the presence of topography in Jansen et al. (2015). This is a very valuable step, providing one promising starting point and baseline benchmark.

Concerning subgrid dynamics, in most of the approaches cited, the model is either entirely local or is based on global energy constraints and thus couples the energy

budget over the entire domain. There are, however, early attempts at “second-order closures” by Daly and Harlow (1970) and Deardorff (1973), where the Reynolds tensor is treated as a prognostic variable and closure conditions have to be supplied for the higher-order moments. Schumann (1975) suggested a model with a single scalar transport equation for the subgrid energy. However, the algebraic closure relations for the diagnostic subgrid contributions are complicated and subject to solvability constraints; see Schmidt and Schumann (1989) and Schumann (1991). These ideas have been revisited subsequently by Schiestel and Dejoan (2005) and Chaouat (2012), but the problem remains open.

To summarize, the main open questions are:

- Find a suitable mathematical framework for subgrid momentum parameterization with minimized spurious energy dissipation.
- Implement practical backscatter algorithms in primitive equation ocean circulation models.

Any progress will have substantial impact on the energetic consistency of existing and future climate models.

This chapter aims at an elementary introduction to the circle of questions outlined above. Our intent is not to give ready and complete answers, but to highlight the issues and survey some of the emerging approaches.

We begin with a brief summary of the concept of subgrid momentum closures in Section 5.2. In Section 5.3, we review theoretical ideas on quasigeostrophic turbulence, with a brief summary on ocean mesoscale and submesoscale turbulence. Our main goal is to emphasize that the notion of “subgrid” scale, as related to ocean modeling, depends on the resolution, which complicates the question on subgrid closures.

In Sections 5.4 and 5.5, we review several proposed parameterizations. The first is the approach by Jansen and Held (2014) which is based on a local subgrid energy budget and an essentially empirical backscatter term which may be either deterministic or stochastic. The second, more sophisticated, but also more expensive and less easily generalized approach is due to Grooms and Majda (2013, 2014) who replace the Reynolds stress term with a stochastic process and explicitly evolve the local subgrid statistics in a local micro-cell attached to each grid box. The last emerging closure scheme is due to Mana and Zanna (2014); it was initially introduced semi-empirically, but later justified under precise assumptions by Grooms and Zanna (2017). The section closes with a brief review on α -models which provide a framework for regularizing fluid equations without adding dissipation which may possibly be interpreted as a nonlinear remapping of wavenumbers.

Section 5.6 offers concluding remarks and some very brief pointers to the literature for further directions beyond those covered so far.

5.2 Subgrid Momentum Closures

To fix concepts, let us focus on the momentum equations for a homogeneous incompressible or Boussinesq rotating ideal fluid,

$$\partial_t \mathbf{u} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) + 2\boldsymbol{\Omega} \times \mathbf{u} + \rho^{-1} \nabla p = \mathbf{F} + D\mathbf{u}, \quad (5.1a)$$

$$\nabla \cdot \mathbf{u} = 0, \quad (5.1b)$$

where \mathbf{u} is the three-dimensional velocity field, $\boldsymbol{\Omega}$ the rotation vector, ρ the constant density, and p the pressure. All force terms are subsumed into \mathbf{F} . In particular, the system can represent the Boussinesq equations when augmented by thermodynamic equation(s) and with \mathbf{F} representing all other forces including buoyancy. The operator D represents dissipation through physical processes or prior modeling. This equation may be read either as a partial differential equation (PDE) or as a fine-scale numerical approximation thereof.

In analogy with classical large eddy simulation, we introduce a coarsened velocity field $\bar{\mathbf{u}}$. We assume very little about the coarsening process other than that it is linear and commutes with time differentiation. In the classical PDE setting, $\bar{\mathbf{u}}$ may be obtained from \mathbf{u} by convolution with a filter kernel. However, the more interesting point of view is that $\bar{\mathbf{u}}$ represents the solution of a modified numerical model at lower resolution. Then, $\bar{\mathbf{u}}$ satisfies the equation

$$\partial_t \bar{\mathbf{u}} + \bar{\nabla} \cdot (\bar{\mathbf{u}} \otimes \bar{\mathbf{u}}) + 2\boldsymbol{\Omega} \times \bar{\mathbf{u}} + \rho^{-1} \bar{\nabla} \bar{p} = \mathbf{R}(\mathbf{u}) + \bar{\mathbf{F}} + \bar{D}\bar{\mathbf{u}}, \quad (5.2a)$$

$$\bar{\nabla} \cdot \bar{\mathbf{u}} = 0, \quad (5.2b)$$

with *eddy source term*

$$\mathbf{R}(\mathbf{u}) = \bar{\nabla} \cdot (\bar{\mathbf{u}} \otimes \bar{\mathbf{u}}) - \overline{\nabla \cdot (\mathbf{u} \otimes \mathbf{u})} + \overline{D\mathbf{u}} - \bar{D}\bar{\mathbf{u}} \quad (5.3)$$

where $\bar{\nabla}$ and \bar{D} denote the coarsened gradient or divergence operator and coarsened dissipation operator, respectively. Thinking of coarsening as a change of numerical resolution, we do not assume that coarsening commutes with the fine-scale operators even though this is often true for convolution coarsening on the continuum. However, we have made two minor simplifying assumptions: First, we have commuted the coarsening operation into the Coriolis term, which is exactly true on the f -plane and approximately true for a slowly varying Coriolis parameter, and second, we are assuming that the flow is incompressible at the coarse level with \bar{p} denoting the implied coarsened pressure. (So, \bar{p} is not obtained by convolution of p with the filter, but is chosen to enforce incompressibility of the coarse velocities.)

When considering the full Boussinesq system, the transport equations for potential temperature and other thermodynamic quantities need to be coarsened similarly (Aluie and Kurien 2011). For fully compressible flows, it is more natural to coarse-grain the product $\rho\mathbf{u}$, thus modifying the expression for $\mathbf{R}(\mathbf{u})$ above;

see, e.g., Aluie (2013). Additional complications arise with nonlinear equations of state; see, e.g., Eden (2016).

The modeling task is now the following: Find a *closure* or *subscale model* $\overline{\mathbf{R}}(\overline{\mathbf{u}})$ which correlates highly with the true $\mathbf{R}(\mathbf{u})$. The closure may be deterministic or include stochastic terms to reduce bias; it may also include infinitesimal or finite memory. If the momentum equation is coupled to thermodynamics, the same considerations apply to each of the prognostic equations.

One of the most elementary concerns is the distribution and flux of energy across scales, as the statistical behavior of the solution depends on it. In the next section, we illustrate the issues relevant to stratified turbulence in the ocean, concentrating mainly on the quasigeostrophic equations. They allow us to explore the main features of large-scale rotating stratified flow and are also used in most of the recent theoretical studies in the field.

5.3 Quasigeostrophic Turbulence and Ocean Eddies

In this section, we review the energetics of large-scale quasigeostrophic (QG) turbulence. In contrast to three-dimensional turbulence where energy cascades to small scales, QG turbulence is distinguished by an inverse cascade of barotropic kinetic energy to large scales and a cascade of enstrophy to small scales. Thus, it is often said that numerical schemes are required to provide a sink for QG enstrophy at grid scale without dissipating energy. In the following, we explain how and under which conditions this picture arises, but also point out its limitations when used in the context of ocean circulation models.

We emphasize that the classical picture of QG turbulence is strictly valid only under the assumption that QG enstrophy is dissipated at scales much smaller than the forcing scale and that turbulence remains geostrophic across all scales. In stratified flow, the main source of barotropic kinetic energy is the conversion of available potential energy via baroclinic instability. The most unstable baroclinic modes occur close to the first internal Rossby radius of deformation L_d . As we move to even smaller scales in a full model, the ageostrophic or non-balanced component of the flow increases and the quasigeostrophic approximation becomes inaccurate. Thus, there is only a finite small range of scales between L_d and the scale where ageostrophy starts to be important; at smaller scales, the energy cascade is direct. Moreover, since this direct cascade acts as an energy sink, there must also be some downscale flux of energy across the geostrophic range to feed it.

All of the quasigeostrophic models are able to capture only the part of the dynamics that stays close to geostrophic balance, often referred to as *mesoscale eddies*. In the real ocean, the scales at which ageostrophic effects are becoming important are rather close to L_d (see, e.g., Callies and Ferrari 2013) so that the presence of ageostrophic motions, often accompanying smaller-size *submesoscale eddies*, may have a significant impact on the direction of the energy transfer across scales. An additional complication arises from the fact that there is also a direct cascade of

available potential energy which implies that the forcing of the barotropic cascade does not only take place near the most baroclinically unstable mode, but is distributed across a wider range of scales.

In the following, we begin with the simplest concepts in a purely two-dimensional setting, then move to a two-layer model, and finally discuss the continuously stratified quasigeostrophic equations.

5.3.1 Two-Dimensional Turbulence

The very basic notions of rotating turbulence and ocean mesoscale eddies can be introduced in the framework of two-dimensional quasigeostrophic dynamics.

The barotropic quasigeostrophic equations are reviewed, e.g., in Franzke et al. (2019). In the beta-plane approximation and written in terms of the relative vorticity ζ , they read

$$\partial_t \zeta + [\psi, \zeta] + \beta \partial_x \psi = F + D\zeta, \quad (5.4a)$$

$$\zeta = \Delta\psi, \quad (5.4b)$$

where brackets denote the Jacobian operator $[\psi, \zeta] = \nabla^\perp \psi \cdot \nabla \zeta$ with $\nabla^\perp = (-\partial_y, \partial_x)$, ψ denotes the stream function, β is the beta-parameter, D is a dissipation operator to be specified below, and F is the forcing. The forcing F is maintained by baroclinic or barotropic instabilities evolving at some intermediate scales.¹

To begin, we set $\beta = 0$. Further, to simplify the discussion, we non-dimensionalize the horizontal length scale and consider (5.4) on the doubly periodic domain $\mathbb{T}^2 = [0, 2\pi]^2$, so that we can pass to the Fourier representation² where

$$\zeta_k = \frac{1}{2\pi} \int_{\mathbb{T}^2} e^{-ik \cdot x} \zeta(x) dx \quad (5.5)$$

for $k \in \mathbb{Z}^2$. It is useful to separate the dissipation operator D into ‘‘infrared’’ and ‘‘ultraviolet’’ parts that effectively act on large (D_i) and small (D_u) scales. For simplicity, we assume that these operators are diagonal in Fourier space, so that the transformed vorticity equation (5.4a) takes the form

$$\partial_t \zeta_k + J_k = D_i(k) \zeta_k + D_u(k) \zeta_k + F_k \quad (5.6)$$

¹For simplicity, our definition of vorticity (5.4b) does not include a barotropic stretching term modeling the effect of free surface elevation. In a geophysical context, this means that (5.4) is restricted to scales smaller than the external Rossby radius of deformation $L_e = c/f$, where c is the speed of surface gravity waves and f is the Coriolis parameter. This is a reasonable assumption for the ocean where $L_e \approx 2000$ km with $c = 200$ m/s and $f = 0.0001$ s⁻¹, but is more questionable for the atmosphere.

²For the purpose of this exposition, we are using the symmetric definition of the Fourier transform so that the Parseval identity holds with constant one.

where, writing $p = |\mathbf{p}|$ and likewise for the other wavenumber vectors, the Jacobian term is described by

$$J_{\mathbf{k}} = \frac{1}{2\pi} \sum_{\mathbf{k}=\mathbf{p}+\mathbf{q}} \frac{\mathbf{p}^\perp \cdot \mathbf{q}}{p^2} \zeta_{\mathbf{p}} \zeta_{\mathbf{q}}. \quad (5.7)$$

In the absence of dissipation and forcing, equation (5.6) conserves *energy*

$$E = \sum_{\mathbf{k} \in \mathbb{Z}^2} E_{\mathbf{k}} = -\frac{1}{2} \sum_{\mathbf{k} \in \mathbb{Z}^2} \psi_{\mathbf{k}}^* \zeta_{\mathbf{k}} \quad (5.8)$$

and *enstrophy*

$$Z = \sum_{\mathbf{k} \in \mathbb{Z}^2} Z_{\mathbf{k}} = \frac{1}{2} \sum_{\mathbf{k} \in \mathbb{Z}^2} \zeta_{\mathbf{k}}^* \zeta_{\mathbf{k}}, \quad (5.9)$$

with star denoting the complex conjugate.

The presence of two integrals imposes constraints on how energy and enstrophy are transferred in spectral space. The energy balance in each mode \mathbf{k} is obtained by multiplying (5.6) by $\psi_{\mathbf{k}}^*$ and taking the real part, so that

$$\partial_t E_{\mathbf{k}} = T_{\mathbf{k}} + 2 D_i(\mathbf{k}) E_{\mathbf{k}} + 2 D_u(\mathbf{k}) E_{\mathbf{k}} + P_{\mathbf{k}}, \quad (5.10)$$

where $P_{\mathbf{k}} = -\Re[\psi_{\mathbf{k}}^* F_{\mathbf{k}}]$ is the rate of energy pumping and $T_{\mathbf{k}} = \Re[\psi_{\mathbf{k}}^* J_{\mathbf{k}}]$ is the rate of nonlinear energy transfer into mode \mathbf{k} .³ Using (5.7), we can write

$$T_{\mathbf{k}} = \sum_{\{\mathbf{p}, \mathbf{q}\}: \mathbf{k}+\mathbf{p}+\mathbf{q}=0} T(\mathbf{k}|\mathbf{p}\mathbf{q}), \quad (5.11)$$

where

$$T(\mathbf{k}|\mathbf{p}\mathbf{q}) = \frac{1}{2\pi} \mathbf{p}^\perp \cdot \mathbf{q} (q^2 - p^2) \Re[\psi_{\mathbf{k}} \psi_{\mathbf{p}} \psi_{\mathbf{q}}] \quad (5.12)$$

denotes the rate of energy transfer into mode \mathbf{k} from modes $\{\mathbf{p}, \mathbf{q}\}$ and the sum in (5.11) is taken over un-ordered sets $\{\mathbf{p}, \mathbf{q}\}$.

Summing up all the $T_{\mathbf{k}}$, we obtain the overall rate of nonlinear energy transfer T . Clearly, $T = 0$ as the rates of sending and receiving energy must balance across all modes. Following (Fjrtoft 1953), we sort the terms in this sum according to membership in resonant triads of modes

$$\mathcal{S} = \{\{\mathbf{k}, \mathbf{p}, \mathbf{q}\}: \mathbf{k} + \mathbf{p} + \mathbf{q} = 0\}, \quad (5.13)$$

³Expression (5.11) shows that the *instantaneous* rate of energy transfer into mode \mathbf{k} can only be nonzero provided $\psi_{\mathbf{k}}$ is nonzero. This, however, does not imply that a mode which is initially zero remains zero for $t > 0$ as (5.6) allows a tendency for $\zeta_{\mathbf{k}}$. See, e.g., the discussion in Moffatt (2014).

so that

$$T = \sum_{\{\mathbf{k}, \mathbf{p}, \mathbf{q}\} \in \mathcal{S}} (T(\mathbf{k}|\mathbf{p}\mathbf{q}) + T(\mathbf{p}|\mathbf{k}\mathbf{q}) + T(\mathbf{q}|\mathbf{k}\mathbf{p})). \quad (5.14)$$

Within each triad, $\mathbf{k}^\perp \cdot \mathbf{q} = -\mathbf{p}^\perp \cdot \mathbf{q}$. This directly implies⁴ that

$$T(\mathbf{p}|\mathbf{k}\mathbf{q}) = -\frac{q^2 - k^2}{q^2 - p^2} T(\mathbf{k}|\mathbf{p}\mathbf{q}) \quad (5.15a)$$

and

$$T(\mathbf{q}|\mathbf{k}\mathbf{p}) = -\frac{k^2 - p^2}{q^2 - p^2} T(\mathbf{k}|\mathbf{p}\mathbf{q}). \quad (5.15b)$$

These identities constrain the transfer of energy within the triad: If $p < k < q$ and mode \mathbf{k} loses energy by interacting with modes \mathbf{p} and \mathbf{q} , the two other modes gain energy; vice versa, if mode \mathbf{k} gains energy in this triad interaction, then modes \mathbf{p} and \mathbf{q} lose energy. The same holds true for enstrophy. In other words, nonlinear interactions between three modes always transfer energy and enstrophy either from or to the central component.

The total transfer $T_{\mathbf{k}}$ involves all triads this mode participates in and cannot be predicted without additional arguments. Consider first the case without forcing and dissipation, and define the *energy wavenumber* k_e as the centroid of the spectral energy density $E(k)$:

$$k_e = \frac{1}{E} \sum_k k E(k), \quad (5.16)$$

where we assume that the distribution of energy is isotropic in wavenumber space with $E(k)$ denoting the energy in the shell $k = |\mathbf{k}|$. The second moment

$$I = \sum_k (k - k_e)^2 E(k) = Z - k_e^2 E \quad (5.17)$$

is expected to increase with time if energy spreads over wavenumbers. This is natural to expect for any energy spectrum that is initially spectrally localized. Conservation of energy and enstrophy implies that

⁴An independent, conceptual proof goes as follows. Consider a flow in the absence of dissipation and forcing where the Fourier coefficients of only a single resonant triad are nonzero. Such a flow preserves energy and enstrophy, so that $\partial_t (E_{\mathbf{k}} + E_{\mathbf{p}} + E_{\mathbf{q}}) = 0$ and $\partial_t (k^2 E_{\mathbf{k}} + p^2 E_{\mathbf{p}} + q^2 E_{\mathbf{q}}) = 0$. But with only a single active triad, we can factor out $\mathfrak{R}[\psi_{\mathbf{k}} \psi_{\mathbf{p}} \psi_{\mathbf{q}}]$ from the expressions for the rate of energy and enstrophy transfer. Thus, the algebraic prefactors must already cancel, implying triadwise conservation of energy and enstrophy also for general flows; (5.15) is a direct consequence of these conservation laws.

$$\frac{dI}{dt} = -E \frac{dk_e^2}{dt}, \quad (5.18)$$

so when energy spreads over wavenumbers, k_e decreases; i.e., energy moves to larger scales. Similarly, it can be shown that the enstrophy centroid moves downscale if and only if the second moment of enstrophy indicates a spread of the enstrophy distribution; see Vallis (2006) for details. This consideration indicates that if two-dimensional freely evolving turbulence develops cascades, we should expect an inverse energy cascade and a direct enstrophy cascade. It does not mean that there is no energy transfer to small scales or enstrophy transfer to large scales, it only means that on average energy tends to go upscale and enstrophy tends to go downscale.

In practice, turbulent flows are forced–dissipative systems.⁵ They can reach a statistically steady state if dissipation is present at both spectral ends, as is envisioned in (5.6). Although the cause of infrared dissipation is not immediately apparent, in many cases its role can be efficiently played by bottom friction.⁶ We thus return to the forced–dissipative case and consider the idealized situation when the forcing F is spectrally localized to a small interval around a forcing wavenumber k_f , infrared dissipation is localized to wavenumbers $k < k_i$, and ultraviolet dissipation is localized to $k > k_u$, with $k_i < k_f < k_u$.⁷ Assuming statistical stationarity, the mean rate of energy injection ε is balanced by the mean rate of energy dissipation in the infrared ε_i and the mean rate of energy dissipation in the ultraviolet ε_u , i.e.,

$$\varepsilon = \varepsilon_i + \varepsilon_u. \quad (5.19)$$

Likewise, writing η to denote the mean rate of enstrophy injection near wavenumber k_f , we balance with the mean enstrophy dissipation rates η_i and η_u in the respective dissipation ranges, so that

$$\eta = k_f^2 \varepsilon = \eta_i + \eta_u. \quad (5.20)$$

⁵The real ocean is close to this idealization, but not fully, as some eddies leave their regime of creation and evolve freely before being dissipated. Thus, temporal averages of forcing and dissipation may not balance locally in space.

⁶Physically, bottom or surface drag is due to small-scale turbulence, yet modeled in the equations without horizontal differential operators. Thus, the bulk of energy dissipation will happen at the scales where energy is residing; i.e., it will be infrared.

⁷In practice, this assumption is well satisfied at the infrared end because we commonly get energy spectra that are sufficiently red (with theoretical slopes of $-5/3$ or -3), so that most energy is residing at small wavenumbers. Thus, linear (Rayleigh) friction modeling bottom drag with $D_i = -\lambda$ is concentrated where the kinetic energy is concentrated—at small wavenumbers. The same holds true for quadratic friction proportional to $\mathbf{u}|\mathbf{u}|$. The situation is more subtle at the ultraviolet end. Laplacian viscosity has $D_u(k) = -\nu k^2$; i.e., it is concentrated where the enstrophy is concentrated. Here, we have a formal problem, for if the spectral slope is -3 , $D_u(k) E(k)$ will have a maximum in the forcing range. The slope is frequently even steeper (when coherent vortices are formed), so Laplacian viscosity will affect the forcing range. Thus, for true ultraviolet dissipation one needs hyperviscosity (a biharmonic operator at least, but even that is insufficient if the spectra are steeper than -4 which sometimes occurs).

Noting that $\eta_i \leq k_i^2 \varepsilon_i$, we estimate

$$\eta \geq \eta_u = k_f^2 \varepsilon - \eta_i \geq k_f^2 \varepsilon - k_i^2 \varepsilon_i \geq (k_f^2 - k_i^2) \varepsilon = (1 - k_i^2/k_f^2) \eta. \quad (5.21)$$

Thus, $\eta_u \rightarrow \eta$ in the limit $k_i/k_f \rightarrow 0$. Similarly, noting that $\eta_u \geq k_u^2 \varepsilon_u$, we can show that $\varepsilon_i \rightarrow \varepsilon$ when $k_f/k_u \rightarrow 0$.

Thus, in the asymptotic limit $k_i \ll k_f \ll k_u$, there is only upscale energy transfer for $k < k_f$ and only downscale enstrophy transfer for $k > k_f$. As a result, these regimes are called (*inverse*) *energy range* and (*direct*) *enstrophy range*, respectively. The mean energy spectral density $\langle E(k) \rangle$ should be such that a constant spectral energy flux is carried across each range.⁸ Assuming that the mean energy transfer is spectrally local, as well as spatially homogeneous and isotropic, one expects that, in the energy range, $\langle E(k) \rangle$ depends only on the infrared energy dissipation rate $\varepsilon_i = \varepsilon$ and on k . This lead Kraichnan (1967), Leith (1968), and Batchelor (1969), hereafter KLB (following earlier arguments by Kolmogorov for classical turbulence) to conclude that the only dimensionally consistent scaling law is

$$\langle E(k) \rangle = C_E \varepsilon^{\frac{2}{3}} k^{-\frac{5}{3}}. \quad (5.22)$$

Likewise, in the enstrophy range, one expects that $\langle E(k) \rangle$ will depend only on the ultraviolet enstrophy dissipation rate $\eta_u = \eta$ and on k , leading to the scaling law

$$\langle E(k) \rangle = C_Z \eta^{\frac{2}{3}} k^{-3}. \quad (5.23)$$

The picture outlined above has two important limitations. The first one relates to the KLB assumption that only local triad interactions (triads where p , k , and q are of the same magnitude) contribute to the mean transfer of energy. For an individual wave number \mathbf{k} in the energy or the enstrophy range where forcing and dissipation are absent, we expect that the mean transfer rate $\langle T_k \rangle$ is zero, which only means that some triads carry energy to \mathbf{k} and some from it. How they do this, however, does not really agree with the KLB picture—in real forced–dissipative two-dimensional turbulence, the contribution from non-local triads is indispensable. For a mode \mathbf{p} in the energy range, one cannot neglect triads with two long legs \mathbf{k} and \mathbf{q} in the forcing or enstrophy range. The local triads with legs being about \mathbf{p} dominate locally, but

⁸To be mathematically precise, the averaging operation $\langle \cdot \rangle$ must be seen as an ensemble average, even though we have no *a priori* knowledge of the probability measure. In practice, the necessary size of the ensemble is also an issue, so one may want to resort to averaging in time: A configuration is “practically statistically stationary” if the difference of averages over an interval $[t, t + T]$ and $[t, t + 2T]$ is less than some prescribed tolerance. One needs, however, to recognize that the required time interval T depends on the quantities we work with. It is relatively small for $E(k)$ and the mean spectral energy flux $\Pi(k)$, it is relatively large for E_k or T_k , and it is even larger for partial transfers $T(\mathbf{k}|\mathbf{p}\mathbf{q})$.

contrary to expectations their average effect is not leading to the inverse energy transfer. Similarly, for mode \mathbf{k} in the enstrophy range, one cannot neglect triads with one short leg \mathbf{p} in the forcing or energy range. The first such analysis is due to Maltrud and Vallis (1993) and was corroborated by Danilov and Gurarie (2001). This means that vortices near the forcing scale are strong enough to stir small-scale vortices in the enstrophy range, which is precisely the way these smaller vortices are formed. The presence of non-locality violates the KLB argument, for it is explicitly assumed that $\langle E(k) \rangle$ in the energy range may only depend on the mean energy flux ε and on k (and not, e.g., on the forcing range), and similarly for the enstrophy range.

The second limitation of the classical KLB picture is that in real systems, the energy and enstrophy ranges are finite. If forcing pumps energy at intermediate scale k_f with finite separation from k_i and k_u , both energy *and* enstrophy are transferred to large *and* small scales through nonlinear interactions. If the wavenumber intervals separating forcing from dissipation are sufficiently broad, most of the energy is transferred upscale and most of the enstrophy downscale. However, these intervals are never broad enough in the ocean, and the question of the amplitude of the direct energy cascade relevant to the ocean is open.

However, even on finite ranges, the classical picture is not entirely lost. The following argument due to Gkioulekas and Tung (2007) provides integral bounds on energy and enstrophy fluxes which do not depend on infinite scale separation. To ease notation, we assume a wavenumber continuum (i.e., an unbounded domain in physical space) and, as before, consider energy densities and energy transfer rate densities as a function of the wavenumber modulus k . The argument, in essence, does not depend on this assumption. In a statistically stationary state, the average $\langle \partial_t E_k \rangle = 0$, so that, taking the time or ensemble mean of (5.10) and averaging over the shell $|\mathbf{k}| = k$, we obtain

$$\langle T(k) \rangle = -D(k) \langle E(k) \rangle - \langle P(k) \rangle \quad (5.24)$$

where we assume that the dissipation operators depend only on k , so that we can write $D(k) = 2D_i(k) + 2D_u(k)$. Retaining the assumption that D_i is dissipating at wavenumbers smaller than k_i and D_u is dissipating at wavenumbers larger than k_u , the mean spectral energy flux

$$\Pi(k) = \int_k^\infty \langle T(\kappa) \rangle d\kappa = - \int_0^k \langle T(\kappa) \rangle d\kappa \quad (5.25)$$

is necessarily negative for $k < k_i$ and positive for $k > k_u$. The equality between the two integrals in (5.25) holds for each realization pointwise in time due to conservation of energy in the inviscid unforced system. Analogous statements hold true for the spectral enstrophy flux. Then,

$$\begin{aligned}
\int_0^k 2\xi \Pi(\xi) d\xi &= \int_0^k (\kappa^2 - k^2) \langle T(\kappa) \rangle d\kappa \\
&= - \int_k^\infty (\kappa^2 - k^2) \langle T(\kappa) \rangle d\kappa \\
&= \int_k^\infty (\kappa^2 - k^2) (D(\kappa) \langle E(\kappa) \rangle - \langle P(\kappa) \rangle) d\kappa, \quad (5.26)
\end{aligned}$$

where the first identity is obtained by exchanging the order of integration, the second identity is once again based on the conservation of energy and enstrophy in the inviscid unforced case, and the last step uses the statistical stationarity relation (5.24). When $k > k_u$, the rate of energy pumping $P(\kappa)$ appearing in the integrand of (5.26) is zero, while the contribution from dissipation is negative. Thus, for every $k > k_u$ and, trivially, for every $k < k_i$,

$$\int_0^k \xi \Pi(\xi) d\xi < 0. \quad (5.27)$$

Due to the weight in this integral, we see that the upscale flux of energy for $k < k_i$ must be typically larger than the downscale flux of energy for $k > k_u$. A similar inequality shows that the enstrophy flux is predominantly downscale (Gkioulekas and Tung 2007).

Whether or not inertial ranges can be observed depends on the spectral loci of dissipation and forcing. In particular, when an inverse cascade is observed, it only means that some energy dissipation is located at smaller wavenumbers than energy forcing. As a rule, dissipation and forcing are spread over wavenumbers and may even intersect. Moreover, when forcing extends up to the spectral cutoff k_{\max} of a simulation, the small direct cascade of energy may be partially hidden. Thus, in most cases, clean inertial ranges are absent. And even when inertial ranges void of dissipation and forcing exist, the observed spectra may deviate from the KLB predictions because non-local triad interactions are always present and may be significant for finite ranges; see the discussion and examples in Danilov (2005). We conclude that spectral slopes alone tell very little about the nature of the underlying dynamics, and one must turn to exploring the distribution of forcing and dissipation over scales.

Let us comment briefly on the case when $\beta \neq 0$. In this situation, energy is channeled into zonal modes and large-scale features become highly anisotropic: Jets appear near the Rhines scale $L_{\text{Rh}} = E^{1/4}/\beta^{1/2}$ which is several hundred km for ocean conditions (see Rhines 1975 and the discussion in Danilov and Gurarie 2004). Smaller scales are largely unaffected.

As far as subgrid closures are concerned, the framework of two-dimensional incompressible turbulence implies that the closures should be consistent with the k^{-3} power law in the enstrophy range. The extent to which this is possible with traditional closures is explored by Graham and Ringler (2013). The degree to which this is relevant to the dynamics of the real ocean remains an open question, for

dynamics at scales smaller than the internal Rossby radius develop an ageostrophic component.

5.3.2 Two-Layer Geostrophic Flows

In the presence of stratification, the situation becomes more complex. The general picture as presented in textbooks (see Salmon 1998; Vallis 2006) is as follows. On scales larger than the first internal Rossby radius L_d , there is a direct cascade of baroclinic (available potential plus kinetic) energy and an inverse cascade of barotropic (kinetic) energy. The baroclinic cascade is maintained through instabilities that release the available potential energy from an existing pool. It feeds the barotropic cascade at scales around L_d via the mechanism of baroclinic instability. This energy is transferred upscale where it is dissipated. On scales smaller than L_d , the layers interact only weakly and behave similarly to two-dimensional turbulence discussed above. In this regime, the dynamics are governed by the direct enstrophy cascade, implying the scaling exponent -3 for the modal or layer kinetic energy spectra. We note that this implies the presence of a direct energy cascade at these scales.

In this section, we discuss these concepts in the simplest possible setting, the two-layer quasigeostrophic model. It is essential that the two-layer model allows for a coupling between the eddy potential energy dynamics and the eddy kinetic energy. In this sense, it represents a minimum model for the real dynamics in ocean and atmosphere.

The two-layer QG model introduces important corrections to the single-layer situation explained in Section 5.3.1 above. First, it shows that the concept of spectrally localized forcing does not work, for the energy is supplied to the system over a broad range of scales, with the maximum spectral density of pumping shifted toward the scale of the energy spectrum maximum. Second, the notion of cascade has to be adjusted, for predictions are made for the baroclinic and barotropic energies, not for the layers.

For simplicity, we assume the layer depths are equal. The two-layer system can then be written as

$$\partial_t q_i + [\psi_i, q_i] = F_i + \delta_{2i} D_i \psi_i + D_u \psi_i + (-1)^{i+1} \kappa (\psi_1 - \psi_2), \quad (5.28a)$$

$$q_i = f_0 + \beta y + \Delta \psi_i + (-1)^i k_d^2 (\psi_1 - \psi_2)/2, \quad (5.28b)$$

where $i \in \{1, 2\}$,

$$k_d = \frac{1}{L_d} = \frac{\sqrt{8} f_0}{N_0 H} \quad (5.29)$$

is the inverse of the baroclinic Rossby radius, f_0 is the Coriolis frequency, N_0 is the typical Brunt–Väisälä frequency, and H is the total fluid depth; see, e.g., Franzke et al. (2019) for details.

We remark that when diagnosed using the leading-order per-layer geostrophic balance relation, the difference in layer stream functions, $\psi_1 - \psi_2$, is proportional to the displacement of the interface between the layers. Thus, the last term in (5.28b) can be interpreted as the contribution to potential vorticity perturbations from the layer interface and is referred to as the stretching term.

We think of infrared dissipation D_i acting as bottom drag only on the lower layer. Then, $D_i = -\lambda \Delta$ with λ the bottom drag coefficient. Ultraviolet dissipation is typically modeled by hyperviscosity of some order $n \geq 2$, so that $D_u = \nu(-\Delta)^n$ with hyperviscosity coefficient ν . The last term in (5.28a) models thermal relaxation of the layer interface, with $2\kappa/k_d^2$ the inverse timescale. It restores interface displacement and thus enters the layer equations with the opposite sign.

Although the ocean is mainly driven by wind stress applied to the upper layer, a theoretically simpler situation occurs when the interface between layers is relaxed toward a position with a uniform slope, i.e., taking $F_i = -(-1)^i \kappa U y$, with y the meridional coordinate. Equation (5.28a) in this case has an equilibrium solution $\psi_1 - \psi_2 = -Uy$, which in the presence of bottom drag implies $\psi_1 = -Uy$ and $\psi_2 = 0$.⁹ The velocity U defines the vertical shear and interfacial slope in the two-layer QG model. This equilibrium solution corresponds to a pool of available potential energy (APE) and can be baroclinically unstable.

Splitting the stream functions into the equilibrium stream functions and perturbation or “eddy” stream functions ψ_1^{eddy} and ψ_2^{eddy} , we write

$$\psi_1 = -yU + \psi_1^{\text{eddy}} \quad \text{and} \quad \psi_2 = \psi_2^{\text{eddy}}. \quad (5.30)$$

Further, it is useful to rewrite the system in terms of the *eddy barotropic stream function* ψ and the *eddy baroclinic stream function* τ , respectively, defined by

$$\psi = \frac{\psi_1^{\text{eddy}} + \psi_2^{\text{eddy}}}{2} \quad \text{and} \quad \tau = \frac{\psi_1^{\text{eddy}} - \psi_2^{\text{eddy}}}{2}, \quad (5.31)$$

and the corresponding *eddy barotropic potential vorticity* q and *eddy baroclinic potential vorticity* ω defined as

$$q = \Delta\psi \quad \text{and} \quad \omega = \Delta\tau - k_d^2 \tau. \quad (5.32)$$

We note that the stretching term from (5.28b) appears as the second term in the definition of ω .

Substituting (5.30) into (5.28), writing out the sum and the difference of the layer equations, and rewriting all expressions in terms of the modal stream functions (5.31) and their associated potential vorticities, we obtain

⁹To see this, add a small deviation from the linear dependence in the y -direction and then consider the limit when this deviation vanishes.

$$\partial_t q + [\psi, q] + [\tau, \omega] + \frac{U}{2} \partial_x (q + \Delta \tau) + \beta \partial_x \psi = \frac{1}{2} D_i (\psi - \tau) + D_u \psi, \quad (5.33a)$$

$$\partial_t \omega + [\psi, \omega] + [\tau, q] + \frac{U}{2} \partial_x (\omega + q + k_d^2 \psi) + \beta \partial_x \tau = -\frac{1}{2} D_i (\psi - \tau) + D_u \tau + \kappa \tau. \quad (5.33b)$$

In the following, we will endow the perturbation quantities with doubly periodic boundary conditions. This is possible because the forcing terms, which are non-periodic in the y -direction, drop out of the equations for the perturbation quantities. However, the information on forcing is retained in the terms proportional to U .

The barotropic equation (5.33a) contains self-advection (i.e., the advection of barotropic eddy PV by the barotropic velocity field), whereas the baroclinic equation (5.33b) is linear in the baroclinic variables. Thus, barotropic dynamics are similar to two-dimensional vorticity dynamics characterized by an inverse energy cascade, whereas baroclinic dynamics are similar to the advection of a passive tracer which possesses a direct energy cascade.¹⁰

As in Section 5.3.1, we consider the modal energy balances for the barotropic (kinetic) energy

$$E^\psi = \sum_{k \in \mathbb{Z}^2} E_k^\psi = -\frac{1}{2} \sum_{k \in \mathbb{Z}^2} \psi_k^* q_k \quad (5.34)$$

and baroclinic energy

$$E^\tau = \sum_{k \in \mathbb{Z}^2} E_k^\tau = -\frac{1}{2} \sum_{k \in \mathbb{Z}^2} \tau_k^* \omega_k = \frac{1}{2} \sum_{k \in \mathbb{Z}^2} (k^2 + k_d^2) |\tau_k|^2, \quad (5.35)$$

where the contribution prefactored by k^2 is baroclinic kinetic energy and the contribution prefactored by k_d^2 is available potential energy. Taking the Fourier transform of the barotropic and baroclinic equations, multiplying with ψ_k^* and τ_k^* , respectively, and taking the real part, we obtain

$$\partial_t E_k^\psi = T_k^\psi + C_k^\psi + D_k^\psi, \quad (5.36a)$$

$$\partial_t E_k^\tau = T_k^\tau + C_k^\tau + G_k + D_k^\tau. \quad (5.36b)$$

The terms

$$T_k^\psi = \Re[\psi_k^* J_k(\psi, q)], \quad (5.37a)$$

$$T_k^\tau = \Re[\tau_k^* J_k(\tau, q)] - k_d^2 \Re[\tau_k^* J_k(\psi, \tau)] \quad (5.37b)$$

with $\mathbf{k} = (k_x, k_y)$ describe energy transfer within the barotropic and baroclinic modes,

¹⁰Scott and Arbic (2007) show that there will be baroclinic self-interactions for unequal layers, leading to an inverse cascade of baroclinic kinetic energy.

$$C_k^\psi = \Re[\psi_k^* J_k(\tau, \omega)] - \frac{U}{2} k^2 \Re[ik_x \psi_k^* \tau_k], \quad (5.37c)$$

$$C_k^\tau = \Re[\tau_k^* J_k(\psi, \Delta\tau)] - \frac{U}{2} k^2 \Re[ik_x \tau_k^* \psi_k] \quad (5.37d)$$

describe transfer from baroclinic to barotropic modes and vice versa, respectively,

$$G_k = \frac{U}{2} k_d^2 \Re[ik_x \tau_k^* \psi_k] \quad (5.37e)$$

represents the generation of energy, and all dissipative terms are subsumed into D_k^ψ and D_k^τ .

One can readily see that the generation term is proportional to the meridional buoyancy flux which tends to level off the layer interface (for APE has to be released) if the system is baroclinically unstable. In this case, its mean value has to be positive definite in a statistically stationary sense. Note that G_k is defined by the dynamics and is not an external parameter as in 2D barotropic turbulence theory.

Since

$$\sum_{k \in \mathbb{Z}^2} T_k^\psi = \sum_{k \in \mathbb{Z}^2} T_k^\tau = 0, \quad (5.38)$$

these two terms only redistribute energy between scales. Likewise,

$$\sum_{k \in \mathbb{Z}^2} C_k^\psi = - \sum_{k \in \mathbb{Z}^2} C_k^\tau, \quad (5.39)$$

so that these terms only redistribute energy between baroclinic and barotropic modes.

In the traditional view of baroclinic turbulence (Rhines 1977; Salmon 1980), one introduces spectral energy fluxes analogous to (5.25),

$$\Pi^\psi(k) = - \int_0^k T^\psi(\kappa) d\kappa \quad \text{and} \quad \Pi^\tau(k) = - \int_0^k T^\tau(\kappa) d\kappa, \quad (5.40)$$

describing the redistribution of energy between scales. There are numerous publications discussing the behavior of fluxes in this situation (e.g., Scott and Arbic 2007). The barotropic flux Π^ψ can be shown to be negative at $k < k_d$ corresponding to an inverse cascade of barotropic energy, while the baroclinic flux Π^τ is always positive corresponding to a direct cascade of full (i.e., potential and kinetic) baroclinic energy. Although there is an upscale (i.e., toward large scales) transfer of barotropic kinetic energy, there is no inertial range at $k < k_d$ because the transfer of energy from the baroclinic into barotropic mode is spread over all wavenumbers, being stronger at smaller k .

Thus, no spectral law can be predicted for the inverse cascade in this case. In contrast, on scales smaller than L_d the stretching term in the expression for the

quasigeostrophic potential vorticity becomes small compared to the relative vorticity and, as already mentioned, each layer behaves as in two dimensions implying the scaling exponent -3 for the kinetic energy.

This picture relies on the fact that the assumed forcing maintains a pool of available potential energy which is then transferred to eddies through baroclinic instability, which develops into a nonlinear regime of quasistationary balance between the release of potential energy, nonlinear transfer, and dissipation. In general, forcing will drive both barotropic and baroclinic components of the mean flow. But even if forcing is only baroclinic, as is the case here, a mean barotropic flow is created in the presence of friction and/or topography. For uniform shear, the release of APE through baroclinic instability is the main source of energy for the eddies, but the kinetic energy of the mean flow may also be important in general.

The picture described so far is tied to the choice of writing the fields in terms of barotropic and baroclinic modes. Arguments will differ when looking at the transfer of energy between layers or between kinetic and potential energy. In particular, the sum of transfers between modes is zero *only* when integrated over wavenumbers. This explains why the picture of transfers will be modified if considered for layers (there will be transfers between the layers) and for the total energy (when baroclinic and barotropic kinetic energies will be combined, and potential energy split off the baroclinic energy).

The total energy at wave vector \mathbf{k} ,

$$E_{\mathbf{k}} = \frac{1}{4} k^2 (|\psi_1|_{\mathbf{k}}^2 + |\psi_2|_{\mathbf{k}}^2) + \frac{1}{2} k_{\text{d}}^2 |\tau|_{\mathbf{k}}^2 \quad (5.41)$$

has a rate of change

$$\partial_t E_{\mathbf{k}} = T_{\mathbf{k}}^K + T_{\mathbf{k}}^P + G_{\mathbf{k}} + D_{\mathbf{k}}^i + D_{\mathbf{k}}^u \quad (5.42)$$

with transfer rates

$$T_{\mathbf{k}}^K = \frac{1}{2} \Re[\psi_{1\mathbf{k}}^* J_{\mathbf{k}}(\psi_1, \Delta\psi_1) + \psi_{2\mathbf{k}}^* J_{\mathbf{k}}(\psi_2, \Delta\psi_2)], \quad (5.43a)$$

$$T_{\mathbf{k}}^P = -k_{\text{d}}^2 \Re[\tau_{\mathbf{k}}^* J_{\mathbf{k}}(\psi, \tau)], \quad (5.43b)$$

a generation term $G_{\mathbf{k}}$ as before, and rates of frictional (infrared) dissipation $D_{\mathbf{k}}^i$ and viscous (ultraviolet) dissipation $D_{\mathbf{k}}^u$.

Intermodal or interlayer transfers are now included in the kinetic and potential energy transfers. The emerging picture is perhaps the most transparent; see Figure 5.2. It shows that generation is nearly compensated by large-scale dissipation, that the EPE flux is direct, for it takes the generated eddy energy $G_{\mathbf{k}}$ and carries it to larger wavenumbers gradually releasing it to kinetic energy, and that the EKE flux is inverse, for it takes the released potential energy and carries it back to the interval of small wavenumbers where it is dissipated. It is important to note that transfers into EKE and from EPE are centered at k_{d} and occupy at least one octave of wavenumbers more on

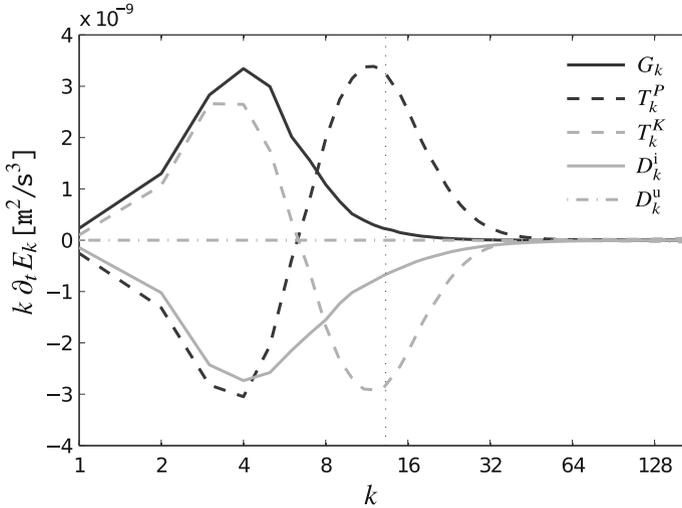


Fig. 5.2 Spectral energy fluxes corresponding to the fluxes in (5.42), integrated over the wavenumber shell $|k| = k$. Figure adapted from Jansen and Held (2014), their Figure 4. Note that the vertical axis shows $k \partial_t E_k$, so that area under the curve in singly logarithmic scaling corresponds to total transfer rates. Note further that the scale of wavenumbers k shown is normalized by $2\pi/L$, where L is the domain size. The deformation scale $k_d = 1/L_d$ is marked by the vertical dotted line

the short-wave side. Simulations by Jansen and Held (2014) demonstrate a spectrum of barotropic EKE close to but steeper than -3 starting from k_d and toward larger wavenumbers. Yet, a substantial part of the interval where this spectrum is observed is where the transfers take place, i.e., where there cannot be an inertial range. In other words, the existence of a well-defined spectral slope is not an indicator of an inertial range, which is frequently forgotten.

Although the theoretical prediction of the inverse cascade is formally made for the barotropic kinetic energy, it is commonly observed for baroclinic kinetic energy and for layer kinetic energies. This behavior is clarified by Scott and Arbic (2007).

We see that if there is a hope for the interval of self-similar behavior in layer QG dynamics, such behavior should be on the side of small wavenumbers and be consistent with the -3 spectral law. However, the two-layer setup indicates very clearly that the transfer from EPE to EKE involves wavenumbers around k_d or larger. For this reason, this spectral law and self-similar behavior of inertial range can only be expected to hold for wavenumbers essentially larger than k_d , which come too close to the scales where ageostrophy is important in the real ocean. The wavelength associated with $k_d = 1/L_d$ is $2\pi L_d$. On meshes with spacing $a = L_d$, this wavelength is well resolved, but this extra resolution is just needed to accommodate for spectral exchanges between EPE and EKE. In practice, in ocean circulation models, the resolution of k_d^{-1} is not always (or not everywhere) achieved. In this case, eddy dynamics may suffer not only from excessive subgrid dissipation but also from the mere fact that the interval where EPE has to feed EKE is too short. The spectral

interval where most of the generation (conversion from the available potential energy to the EPE) takes place tends to be at wavenumbers smaller than k_d . Yet, as shown by Jansen and Held (2014), the generation turns out to be sensitive to dissipation in the vicinity of k_d . We propose that the ability of subgrid closures to least interfere with energy generation presents a convenient guiding principle in these cases.

An important parameterization for relatively coarse, non-eddy-permitting ocean simulations was introduced by Gent and McWilliams (1990); it is now known as the Gent–McWilliams parameterization. Here, we explain the idea in the context of the two-layer model (5.28). On scales larger than L_d , the relative vorticity is expected to be small compared to the stretching term, the last term in (5.28b), which models perturbations of the layer interface. Correspondingly, the dominant nonlinear contribution to (5.28a) is the divergence of the thickness flux

$$(-1)^i k_d^2 [\psi_i^{\text{eddy}}, \frac{1}{2} (\psi_1^{\text{eddy}} - \psi_2^{\text{eddy}})] = (-1)^i k_d^2 [\psi, \tau] = (-1)^i k_d^2 \nabla \cdot (\tau \nabla^\perp \psi). \quad (5.44)$$

This term will be very small if mesoscale eddies are not well resolved. The proposal of Gent and McWilliams (1990) amounts to adding a flux divergence of the form

$$\mathcal{F}_i = (-1)^i k_d^2 \nabla \cdot (\varkappa \nabla \tau) \quad (5.45)$$

to the right-hand sides of the two-layer equations as a parameterization for the effect of unresolved eddies on the resolved flow. The coefficient \varkappa is sometimes taken constant, more frequently, however, selected as a polynomial of the velocity difference $U = |\mathbf{u}_1 - \mathbf{u}_2|$ based on qualitative theory where degree and the coefficients of the polynomial are chosen empirically. In this case, \varkappa is a measure of vertical instability in the system; see, e.g., Stone (1972), Cessi (2008), and Held and Larichev (1996).

By construction, the \mathcal{F}_i model only the subgrid layer thickness flux, not the full potential vorticity flux. They provide a sink for potential energy, thus emulating the effect of baroclinic instability on the potential energy balance in a model that is too coarse to resolve this process directly. This technique prevents the buildup of an unlimited pool of available potential energy, but does not model Reynolds stresses nor does it feed energy back into the pool of resolved eddy kinetic energy.

Note that while (5.45) looks like diffusion, it acts on the *layer thickness*. Whenever the interface between two layers is inclined, thickness diffusion means that it will be leveling the interface. This implies that fluid will move in opposite direction in the layers, showing that the Gent–McWilliams parameterization creates a circulation that tends to flatten isopycnals. Thus, while thickness diffusion proceeds in two dimensions, the generated fluid motion is three-dimensional, and it is advective. See Gent (2011) for a detailed discussion.

5.3.3 Continuously Stratified and Surface QG Dynamics

Even within the quasigeostrophic family of models, the picture presented so far is not the end of the story. First, when allowing for continuous stratification, there are many baroclinic vertical modes. Second, there are surface-trapped motions that can be understood in the framework of surface geostrophic dynamics (SQG); see, e.g., a discussion and further references in von Storch et al. (2019).

For simplicity, we consider the three-dimensional QG equations on a layer of uniform depth H with rigid lid upper boundary condition at $z = 0$. The model then reads

$$\partial_t q + [\psi, q] = 0, \quad (5.46a)$$

$$q = f + \Delta_h \psi + f_0^2 \partial_z \frac{\partial_z \psi}{N^2(z)} \quad (5.46b)$$

where Δ_h denotes the horizontal Laplacian and brackets, as before, the horizontal Jacobian, with boundary conditions for the buoyancy b at $z = 0, -H$:

$$\partial_t b + [\psi, b] = 0, \quad (5.47a)$$

$$b = f_0 \partial_z \psi. \quad (5.47b)$$

According to Wunsch (1997), the bulk of ocean kinetic energy is well captured by the barotropic and first baroclinic modes. For this reason, the major conclusion regarding the spectral slope -3 of the direct enstrophy cascade remains valid for the bulk of the ocean. However, the standard basis for vertical modes, as given by the eigenvalue problem

$$f_0^2 \partial_z \frac{\partial_z \Psi_n(z)}{N^2(z)} + \lambda_n^2 \Psi_n(z) = 0 \quad (5.48)$$

with zero boundary conditions for $\partial_z \Psi$ at $z = 0, -H$, does not take into account surface buoyancy perturbations. Baroclinic instabilities evolving as solutions of (5.46) deal with the modes of the full operator that satisfy the boundary conditions (5.47), and cannot be understood in the frame of the standard basis. Certain textbook instabilities, for example the Eady problem (see, e.g., Vallis 2006), rely entirely on surface-trapped dynamics.

Even though it is possible to reformulate the surface dynamics as δ -sheets of potential vorticity, such solutions cannot be represented in terms of the vertical eigenmode basis. Layered models, however, include the surface dynamics in their upper and lower layer potential vorticity. Therefore, the two-layer model described in Section 5.3.2 above cannot separate surface-driven instabilities from interior instability mechanisms, and the simplest model where this can be explored is the three-layer model studied in Badin (2014).

To separate the surface dynamics from the interior in the continuously stratified QG model, one considers the case where $q = \text{const}$ in an infinitely deep layer, so that only surface dynamics remains. Then, the horizontal Fourier coefficients of the stream function ψ representing the surface buoyancy perturbation satisfy

$$f_0^2 \partial_z \frac{\partial_z \psi_k(z)}{N^2(z)} - k^2 \psi_k(z) = 0 \quad (5.49)$$

with non-homogeneous Neumann conditions at the top and decay toward infinite depth. When $N = \text{const}$, the ψ_k decay with depth as $\exp(kNz/f_0)$; i.e., they decay on a vertical scale

$$H \gtrsim \frac{f_0}{kN}. \quad (5.50)$$

Correspondingly, in a uniformly stratified layer of depth H , only surface perturbations larger in size than the first baroclinic Rossby radius may reach through the fluid depth.

In the absence of forcing and dissipation, surface dynamics will preserve integrals of buoyancy variance and the product ψb . The latter leads to an inverse cascade at large scales with a -1 spectrum of surface kinetic energy, and the former leads to a direct buoyancy variance cascade with a $-5/3$ spectrum (see, e.g., Smith et al. 2002). Note that the prediction concerns surface kinetic energy and is valid for uniform stratification. Since in this case $|\nabla \psi_k| = k |\psi_k| \sim |b_k|$, we expect the same spectral law for kinetic energy

$$E_{\text{EKE}} = \frac{1}{4} \sum_k |\nabla \psi_k|^2 \quad (5.51)$$

and available potential energy

$$E_{\text{APE}} = \frac{1}{4N^2} \sum_k |b_k|^2. \quad (5.52)$$

Small scales do not penetrate deep, and spectra become steeper. Furthermore, they are modified by stratification; examples including exponential stratification and the case of a mixed upper layer are discussed in Callies and Ferrari (2013).

Instabilities in the real ocean project on both deep-ocean modes and surface modes, and depend on the structure of PV of the basic ocean state. Surface dynamics are expected to be an important contributor at locations where the interior PV gradients are weak. Since shallow surface modes are excited as a result of evolving instability, the transfer of available potential energy into eddy kinetic energy is not spectrally local. This implies that the argument in favor of precisely the -1 or $-5/3$ spectral slope at the surface is rather weak. However, it is appropriate to expect that spectral laws for near-surface velocities at small scales are shallower than the -3 prediction for the enstrophy range.

5.3.4 *Ocean Models and Observational Evidence*

To study ocean turbulence beyond the idealized models mentioned before, we must turn to numerical studies of the primitive equations, full ocean circulation models, and observational evidence. In this context, mesoscale or submesoscale structures which become ageostrophic in high-resolution models are of particular interest. In the following, we review a few studies which highlight these issues with the understanding that this selection is far from being complete or representative.

To begin, the recent interest in surface quasigeostrophic (SQG) dynamics was triggered by the observation that energy spectra of surface geostrophic velocities derived from altimetric data are noticeably shallower at many locations than spectra predicted by the theory of QG turbulence (Lapeyre 2009). High-resolution simulations also lend support to the relevance of the SQG concept for understanding the simulated behavior and observations. However, at scales about the first baroclinic Rossby radius and smaller, in real situations as well as in high-resolution simulations with full primitive equations, surface quasigeostrophic dynamics are accompanied by frontal and mixed-layer instabilities which deviate from geostrophy. Klein et al. (2008) and Capet et al. (2008) analyze the near-surface dynamics in ocean simulations performed at the resolution of 2 km and down to 0.75 km, respectively, and demonstrate that there is a close resemblance to SQG dynamics. The spectra of surface kinetic energy have a slope close to -2 from the spectral maximum to the spectral cutoff at large k . This is much shallower than the slope predicted by quasigeostrophic theory. The conceptual difference to SQG is, however, that the Rossby numbers of eddies at these scales are no longer small and a substantial ageostrophic flow component is generated, which modifies the turbulent energy fluxes. The presence of frontal and mixed-layer instabilities implies that the transfer of available potential energy into kinetic energy continues at rather small scales associated with these instabilities. Nevertheless, the near-surface velocities are nearly in geostrophic balance and the ageostrophic components explain only a small fraction of kinetic energy, only visible close to the high-wavenumber spectral end. Despite their smallness with respect to the dominant rotational component (computed via the Helmholtz decomposition), they are responsible for the downscale cascade of the total eddy kinetic energy. The cascade of the dominant rotational component of the velocity behaves differently: It is upscale and of smaller amplitude than the cascade of full velocity. The fact that it is upscale is perhaps not surprising: As there are transfers from the available potential energy to kinetic energy, as in QG or SQG turbulence, the flux of rotational kinetic energy proceeds to larger scales from the scale of forcing.

Callies and Ferrari (2013) discuss existing views and assess two data sets shedding light on the behavior of ocean submesoscales. They consider scales from about 200 to 1 km. For a site in the Gulf Stream, they found steep (-3) spectra of kinetic energy for scales between 200 and 20 km, and shallower spectra at smaller scales consistent with the -2 slope of the internal gravity wave spectrum. For a site in the North Pacific, they report shallower spectra whose behavior with depth, however, does not

agree with the prediction of SQG. It is proposed that the gravity wave continuum and unbalanced motions can contribute to this behavior.

To summarize, the range of submesoscale, where the subgrid scales of eddy-permitting (and eddy-resolving) ocean circulation models are located, combines features of QG and SQG turbulence but also includes ageostrophic (unbalanced) motions, depending on mesh resolution and ocean stratification. Wavenumbers larger than k_d are those of the forward cascade of EKE, but the inverse cascade can be present for the rotational component of EKE at even smaller scales if small-scale instabilities continue to transfer the available potential energy to EKE. No true slope prediction can be made on scales around k_d because of intermodal exchanges and spectrally spread dissipation.

Inertial ranges may emerge on the side of smaller scales on very high-resolution meshes, but even there one should expect a dependence on the depth and a contribution from submesoscale (frontal) instabilities. So even when they emerge, inertial ranges may deviate from the predictions of QG turbulence because of a forward energy cascade. The dominance of the rotational velocities in the energy spectra does not imply their dominance in energy transfers at large wavenumbers. One may try to draw a certain analogy with the $-5/3$ spectrum observed in the atmosphere between 500 and 10 km, which is that of stratified turbulence with forward energy cascade; see the bibliography, discussion, and analysis of high-resolution simulations in Augier and Lindborg (2013). On larger scales, it matches the dynamics predicted by QG theory.

At present, resolutions in ocean circulation models are such that the near-subgrid scales are in a range where self-similar behavior is unlikely. Subgrid closures can therefore not be universal in the range of resolutions about the Rossby radius. Hence, perhaps the guiding principle should be that of minimizing their damping effect on the rate at which energy is released from the pool of APE and the KE of the background state to the EKE at the resolved scales.

5.4 Energy Backscatter

Although most ocean circulation models used for climate research are coarse, the number of eddy-permitting models is increasing and will dominate in the future. Such models simulate eddy dynamics, but cannot resolve it fully for they suffer from overdissipation. Its origin can be explained as follows. In order to remove the variance of velocity and enstrophy at grid scales for numerical stability, they use harmonic or biharmonic dissipative subgrid viscosity operators (Fox-Kemper and Menemenlis 2008). Together with removing the grid-scale noise, these operators also dissipate energy at adjacent scales, which are in this case the scales close to the internal Rossby radius. As we have seen, these scales host exchanges between the potential and kinetic energy compartments and also determine the eddy energy release from the available potential energy. Their overdissipation is the reason why eddy-permitting flows seldom reach the observed levels of eddy kinetic energy.

The problem of overdissipation and, in more broad context, of subgrid closure that takes into account the existence of unresolved scales has been known in atmospheric sciences for a long time. First papers on this issue appeared almost simultaneously with the KLB concept of two-dimensional turbulence; see Leith (1971) and the discussion in Frederiksen and Davies (1997). It may be explained within the spectral picture of triad interactions in two-dimensional turbulence as follows: Since we can only resolve wavenumbers up to some k_{\max} numerically, it is clear that we miss not only spectrally local interactions responsible for the enstrophy transfer through the boundary at $k = k_{\max}$, which lead to a net energy drain and hence behave as a form of dissipation in the ensemble mean, but also non-local triads, having two legs at $k > k_{\max}$ or on both sides of k_{\max} and one leg at large scales $k_{LS} \ll k_{\max}$, which might force the resolved scales. These interactions are termed *backscatter*. It is not represented by the usual dissipative subgrid operators, which is the main cause of overdissipation in conventional models. A fully deterministic representation of backscatter is impossible as the details of the state of the subgrid are in principle not available. Thus, the best we can hope for is some stochastic model of backscatter.

Theoretical developments in this direction assume as a rule QG dynamics, periodic boundary conditions or spherical geometry, ensemble averaging, and “spectral language” to come up with parameterizations. As an example, we mention the work by Kitsios et al. (2013) who derive both stochastic and deterministic closures by comparing truncated and high-resolution dynamics in a two-layer QG setup on the sphere. It is believed that both types can be equally skillful, for, in any case, useful formulas rely on ensemble averaging and thus do not describe realizations.

Although the turbulent dynamics dictates that drain and backscatter should be described as stochastic processes, additional issues such as numerical stability have to be taken into account. For deterministic parameterizations, the resulting expressions contain powers of the Laplacian, sometimes going beyond the biharmonic one. Their study shows that even in the context of two-layer QG turbulence which is statistically homogeneous in the zonal direction, the final parameterizations of drain and backscatter depend not only on k_d and k_{\max} , but also on the extent of the energy-containing range.

It will be much more difficult to propose parameterizations for domains with horizontal boundaries where spectral language and zonal homogeneity are missing. In addition, all backscatter parameterizations raise the question of numerical stability due to the effective negative viscosity of the terms providing backscatter. Finally, in addition to eddy–eddy interactions considered by Kitsios et al. (2013), contributions may come from interactions involving the unresolved mean field (in the sense of time averages) component. These examples show that progress is possible, but we can hardly expect universally valid solutions.

In the following, we review two specific backscatter parameterizations in detail. The first, due to Jansen and Held (2014), is based on a very straightforward scalar model for the subgrid energy. The second, due to Grooms and Majda (2014), uses a more sophisticated linear model for the subgrid dynamics.

5.4.1 Models with Scalar Subgrid Energy Budget

Since, as mentioned above, comprehensive first-principle models are necessarily complex, we think that simplified implementations of energy backscatter proposed by Jansen and Held (2014) and Jansen et al. (2015), who consider kinetic energy backscatter for QG and primitive equations, respectively, deserve attention. These parameterizations do not aim at mimicking missing interactions with subgrid scales, but seek instead to compensate for the overdissipated energy, which is much easier. Importantly, the amount of energy returned through the proposed backscatter parameterization can be controlled, which is a prerequisite for stability of the algorithm.

Jansen and Held (2014) study the two-layer quasigeostrophic equations with the Leith parameterization as nonlinear small-scale dissipation operator. In each layer i ,

$$D_u \psi_i = -\Delta(v_i \Delta^2 \psi_i) \quad \text{with} \quad v_i = C_L a^6 |\Delta^2 \psi_i|, \quad (5.53)$$

where a denotes the grid-spacing and $C_L = 0.005$. The associated overall rate of viscous dissipation at wavenumber \mathbf{k} is

$$V_{\mathbf{k}} = \frac{1}{2} \sum_{i \in \{1,2\}} k^2 (\psi_i)_{\mathbf{k}}^* (v_i \Delta^2 \psi_i)_{\mathbf{k}}. \quad (5.54)$$

(The layers are assumed to be of equal thickness, hence the additional factor of $1/2$ in the expression for $V_{\mathbf{k}}$ and in similar expressions below.) The rate of frictional dissipation in the bottom layer at wavenumber \mathbf{k} is

$$F_{\mathbf{k}} = \frac{1}{2} \lambda k^2 |\psi_2|_{\mathbf{k}}^2. \quad (5.55)$$

Summing over wavenumbers, the total rate of ultraviolet and infrared dissipation is

$$V = \sum_{\mathbf{k}} V_{\mathbf{k}} \quad \text{and} \quad F = \sum_{\mathbf{k}} F_{\mathbf{k}} \quad (5.56)$$

Since transfers are summed to zero, the overall balance of energy is

$$\partial_t E = G - F - V, \quad (5.57)$$

where G is the generation term with Fourier representation (5.37e).

To compensate for the excessive dissipation at small scales, the simplest model is to add an energy source that returns energy at a rate

$$S = (1 - \varepsilon) V \quad (5.58)$$

so that all but a small fraction $\varepsilon \approx 0.1$ representing the physical rate of ultraviolet dissipation is returned. Jansen and Held (2014) tested two different models for this

source, one deterministic and the other stochastic. In the deterministic version, each layer potential vorticity equation is given a source term

$$s_i = -A(t) \Delta^2 \psi_i \quad (5.59)$$

which corresponds to negative Laplacian viscosity in the momentum equations. The amplitude $A(t)$ is set by the condition that the constraint (5.58) is satisfied at every instance in time. Since the Laplacian is less scale-selective than the biharmonic ultraviolet dissipation, energy will be returned at larger scales than those at which it is dissipated.

The second implementation is stochastic, with

$$s_i = A(t)^{1/2} \eta(\mathbf{x}, t), \quad (5.60)$$

where the η is Gaussian noise, δ -correlated in space and time. The forcing is kept barotropic; i.e., the same noise process is used for both layers, to replenish the inverse cascade of barotropic kinetic energy. In this case, the ensemble mean $\langle S \rangle$ will be proportional to $A(t)$ so that the amplitude can be found at each time step from the constraint (5.58). Of course, (5.58) is satisfied only in the ensemble mean. However, it is also approximately satisfied for each realization as the stochastic forcing is distributed over a large number of spatial locations of the computational grid. The rate of backscatter energy pumping at mode \mathbf{k} is given by

$$S_k = \frac{1}{2} \sum_{i \in \{1,2\}} (\psi_i)_k^* (s_i)_k. \quad (5.61)$$

Thus, even when the s_i have a white noise spectrum, energy backscatter is biased toward the scales with already high energy content. In practice, this involves scales larger than those of V_k .

Jansen and Held (2014) conclude that both parameterizations work rather similarly; however, the stochastic implementation returns energy over a broader interval of wavenumbers. The principal question here, namely how much energy has to be returned and where it should be returned, is left without answer and presents a topic for future research. The amplitude of backscatter is selected globally, which is only appropriate if flow energy is distributed uniformly. In the general case, one needs a local criterion.

A small variation of the kinematic backscatter assumption (5.58) is the introduction of a dynamic global subgrid energy budget E_{sg} via

$$\dot{E}_{\text{sg}} = V - S - 2\lambda\gamma E_{\text{sg}}. \quad (5.62)$$

The last term represents dissipation of subgrid energy by bottom friction, where the parameter γ is the fraction of subgrid energy residing in the lower layer and λ the

bottom drag coefficient.¹¹ This form of a global subgrid energy reservoir is suggested by Jansen and Held (2014) as a motivation to justify assumption (5.58), but could also be used computationally by assuming that the amplitude of backscatter $A(t)$ is proportional to the total subgrid energy in the reservoir.

At the next level of complexity, one may use a local subgrid energy budget. Jansen et al. (2015) suggest a budget for the subgrid energy density e of the form

$$\partial_t e = \nu - s - \nabla \cdot \mathbf{F} - d, \quad (5.63)$$

where ν is the rate of viscous dissipation per unit volume of the resolved scales, s is the rate of backscatter per unit volume, \mathbf{F} is the flux redistributing subgrid energy, and d is the rate of dissipation of subgrid energy per unit volume. Each of these terms must be modeled. Jansen et al. (2015) assume biharmonic ultraviolet dissipation

$$\nu = \frac{1}{H} \sum_i h_i \nu_i |\Delta \mathbf{u}_i|^2, \quad (5.64)$$

where H is the total depth, h_i are the layer depths, and ν_i are the layer horizontal biharmonic viscosity coefficients, assumed positive. In this setting, all operators act in the horizontal only. For the backscatter source, one can take harmonic viscosity so that

$$s = -\frac{1}{H} \sum_i h_i \nu_{bs} |\nabla \mathbf{u}_i|^2 \quad (5.65)$$

with *negative* coefficient of viscosity

$$\nu_{bs} = -C_{bs} a (\max\{2e, 0\})^{\frac{1}{2}} \quad (5.66)$$

with C_{bs} an order-one constant. If the energy to be scattered back becomes too large, e becomes negative and backscatter viscosity goes to zero. This controls the amount of energy returned back.

A major point for discussion is the choice of flux \mathbf{F} for the subgrid energy. Jansen et al. (2015) choose the purely diffusive flux

$$\mathbf{F} = -k_{sg} \nabla e, \quad (5.67)$$

where k_{sg} is an appropriately selected constant of diffusivity. This choice is guided by the observation that the transfer from and to the subgrid can be very spatially rough so that a mechanism is needed to regularize the distribution of e horizontally.

¹¹In a more realistic setting, one may wish to also model the dissipation of the fraction of subgrid energy that cascades downscale and turns into fully three-dimensional turbulence. This intention is implicit in the closure condition (5.58).

However, the question arises whether subgrid energy should not perhaps be advected by the resolved flow or be subject to some other non-local mechanism of transfer.

Finally, the dissipation rate d in (5.63) is typically small and may be neglected.

It turns out that the backscatter parameterizations by Jansen and Held (2014) and Jansen et al. (2015) lead to noticeable improvements even in situations where non-trivial bottom topography is present, and allow the mesoscale eddy dynamics in eddy-permitting simulations to approach those of high-resolution runs. On a qualitative level, the success of these simple implementations of backscatter rests on the idea that energy needs to be scattered back only in places where it is strongly dissipated. Although the vertically averaged or basin-averaged subgrid kinetic energy balance used to assess the backscatter viscosity presents an oversimplification, the energy scattered back is nevertheless modulated by the distribution of resolved energy. This also implies that the parameterization may bring improvements only in situations when an eddy-permitting model already correctly simulates the pattern of kinetic energy distribution but lacks amplitude. In realistic applications resolving the vertical structure with many more layers, the vertical distribution of backscatter viscosity may matter, since surface-trapped modes may exhibit more vertical structure, but this remains to be seen. A theory of where to return the energy scattered back presents an interesting question for further research too. Clearly, with only the harmonic operator at one's disposal, the deterministic backscatter parameterization has limited capabilities so that stochastic closures may still be needed. Furthermore, a missing point is the cascade of EPE which is dissipated too by subgrid diffusive closures or through upwind transport algorithms. Too diffusive transport schemes may result in the reduced transfer between EPE and EKE, so that the role of subgrid closures in the tracer equations should be explored. Conversely, Ilıcak et al. (2012) show that spurious mixing of transport algorithms depends on velocity variance at grid scales, so that energizing these scales above a certain level is not recommended. This issue is further explored in Klingbeil et al. (2019). This set of questions shows that even in the context of energy backscatter, the problem is far from being resolved and new ideas are required.

To apply these ideas to the full primitive equations, we face further questions. To our knowledge, this has not been pursued exhaustively, and we can only sketch a direction; more theoretical analysis and numerical experiments are needed here. To start, we may localize even further, treating the subgrid energy density e as a full three-dimensional field, so that the evolution equation (5.63) now takes the form

$$\partial_t e = \nu - s - \nabla_h \cdot \mathbf{F}_h - \partial_z F_z - d, \quad (5.68)$$

where \mathbf{F} is the flux redistributing the subgrid energy, taken as $\mathbf{F}_h = -K_h \nabla_h e$ and $F_z = -K_z \partial_z e$, where K_h and K_z are appropriately selected horizontal and vertical diffusion coefficients. As before, d is the rate of dissipation of subgrid energy; it is small and may be neglected. This approach is more expensive, for now the evolution equation has to be integrated in three spatial dimensions.

The contribution from ultraviolet dissipation now takes the form

$$\nu = -\nu_{\text{visc}} |\Delta_h \mathbf{u}|^2, \quad (5.69)$$

where ν_{visc} is the coefficient of horizontal biharmonic viscosity. Vertical viscosity in the momentum equation would generally be provided by a vertical mixing parameterization which relies on some physics and empirical data, e.g., using a KPP closure (Large et al. 1994) or k - ε closure (Umlauf and Burchard 2003). The corresponding backscatter source term reads

$$s = -\nu_{\text{bs}} |\nabla \mathbf{u}|^2 \quad (5.70)$$

where ν_{bs} is again given by an expression of the form (5.66).

One may consider stochastic implementation options for the backscatter source. A caveat here is that for the primitive equations, the source must respect the divergence condition. This could be done by a simple projection. Another possibility is to write the horizontal source in the form

$$s = \nabla \times (\Psi \mathbf{k}), \quad (5.71)$$

with $\Psi(\mathbf{x}, t) = P(\mathbf{x}) \eta(\mathbf{x}, t) A(t)$. Here, P is a spatial pattern of eddy kinetic energy (which may be modeled, inferred from high-resolution simulations, or taken from observations), $A(t)$ is the amplitude (selected to ensure subgrid energy balance), and η is a random field generated, for example, by a Markov process. Despite the presence of a differential operator, one has to introduce correlations in time and space to ensure that the resulting forcing is smooth.

Another issue is that, for the primitive equations, the concept of backscatter relates in equal measure to the momentum and to the tracer equations. Compensation for kinetic energy overdissipation is not necessarily sufficient if tracer variance is overdissipated. In principle, an approach resembling the one applied to the quasi-geostrophic potential vorticity equations can be proposed. However, there are some technical difficulties. First, in many cases dissipation is already built into the implementation of the transport operators and cannot be easily accessed. Second, even if it is not, biharmonic operators are not always available for tracers.¹² Finally, the production of tracer variance and the production of kinetic energy are linked, so that additional theoretical analysis is required.

5.4.2 Stochastic Superparameterizations

While the backscatter approximation by Jansen and Held (2014) and Jansen et al. (2015) seeks to return excessively dissipated energy back to the main flow, it needs

¹²For tracers, one commonly uses rotated operators that mix along isopycnals. Their biharmonic versions are more expensive and less stable. Furthermore, such operators do not directly dissipate potential energy.

an eddy-permitting model that is able to simulate a correct pattern of eddy variability. Their subgrid representation only captures the subgrid energy e and does not attempt to represent the parameterized action of the Reynolds stress. In models that are not fully eddy-permitting, this approach will not work and one needs a more sophisticated model of the subgrid. We will discuss so-called *stochastic superparameterizations* as proposed by Grooms and Majda (2013, 2014) and Grooms et al. (2015b) in the context of quasigeostrophic two-layer models.

The main difference between the stochastic parameterization (SP) and the stochastic superparameterization (SSP) is that the latter involves a prognostic fine grid equation which is motivated by the underlying physical evolution equation and involves coarse mesh quantities as well as a stochastic source term. We first explain the idea in the context of the simple single-layer model (5.4), where the essential features of the method are already visible with less notational effort.

Let us decompose the stream function ψ into a coarse mesh stream function ψ_c and a fine mesh stream function ψ' , and likewise define the corresponding vorticities, so that

$$\psi = \psi_c + \psi' \quad \text{and} \quad \zeta = \zeta_c + \zeta', \quad (5.72)$$

where it is understood that $\zeta_c = \Delta\psi_c$ and $\zeta' = \Delta\psi'$. We also split the forcing into a deterministic physical forcing F_c on the coarse mesh and a stochastic forcing F' on the fine mesh. Inserting this ansatz into (5.4), we obtain

$$\begin{aligned} \partial_t \zeta_c + \partial_t \zeta' + [\psi_c, \zeta_c] + [\psi_c, \zeta'] + [\psi', \zeta_c] + [\psi', \zeta'] \\ + \beta \partial_x \psi_c + \beta \partial_x \psi' = F_c + F' + D\zeta_c + D\zeta'. \end{aligned} \quad (5.73)$$

We now split this equation into an evolution equation for the coarse variables and an evolution equation for the fine mesh variables. This procedure is non-rigorous, so there is some freedom of choice. However, the fine system should be linear with constant coefficients so that it can be solved explicitly, for otherwise the combined computational cost would be higher than the cost of simulating the entire system on the fine grid.

Following Grooms and Majda (2014), we decompose the domain Ω into disjoint subdomains Ω_i . Each subdomain contains exactly one grid point of the coarse mesh, and the coarse mesh variables are assumed constant on each adjacent subdomain. The fine systems are then solved independently for one coarse time step with periodic boundary conditions on each subdomain.

The coarse system should contain all coarse terms and the divergence of the eddy potential vorticity flux¹³

$$\mathbf{F}_{\text{epv}} = \overline{\zeta' \nabla^\perp \psi'}. \quad (5.74)$$

¹³The overbar operation denotes averaging over one fine grid cell. Note that it is not possible to simply take the fine grid Jacobian and average over a coarse cell, as this expression would be identically zero. Instead, one uses the average flux and practically operates on it with the coarse grid divergence.

All the remaining terms go into the fine mesh system which is solved independently on each coarse mesh cell. The coarse system then takes the form

$$\partial_t \zeta_c + [\psi_c, \zeta_c] + \nabla \cdot \mathbf{F}_{\text{epv}} + \beta \partial_x \psi_c = F_c + D\zeta_c, \quad (5.75)$$

and the fine system reads

$$\partial_t \zeta' + [\psi_c, \zeta'] + [\psi', \zeta_c] + \beta \partial_x \psi' = S + F' + D\zeta', \quad (5.76)$$

where the nonlinear eddy–eddy interactions are represented by

$$S = \nabla \cdot \mathbf{F}_{\text{epv}} - [\psi', \zeta']. \quad (5.77)$$

The coarse system is solved on the coarse grid. The fine system is linear except for the eddy–eddy term which must be modeled. Grooms and Majda (2014) suggest to replace each Fourier mode S_k by the right-hand side of an Ornstein–Uhlenbeck stochastic process of the form

$$d\zeta = -\gamma \zeta dt + \sigma dW, \quad (5.78)$$

where W is a standard Wiener process.¹⁴ The Ornstein–Uhlenbeck process is controlled by two parameters, the inverse correlation time γ and the variance $\sigma^2/(2\gamma)$ which will later be chosen differently for different wavenumbers. They further assume that the coarse grid fields can be held constant in each fine cell and that there is no forcing on the fine scale. Then, the full fine-scale model in the Fourier representation reads

$$d\zeta'_k = (\ell_k - \gamma_k) \zeta'_k dt + \sigma_k dW_k, \quad (5.79)$$

where ℓ_k is the Fourier symbol of all linear terms in (5.76) and the Wiener processes W_k are mutually independent.

The crucial observation is that S is quadratic in fine-scale quantities, so that a space average corresponds, by the Plancharel theorem, to an integral over $|\psi_k|^2$. Averaging further over the stochastic ensemble, it is clear that it suffices to compute the evolution of $\mathbb{E}[|\psi_k|^2]$. By the Itô Lemma, it is easy to derive a deterministic linear ordinary differential equation for this quantity, which can be solved explicitly and independently for each wavenumber.

The coefficients γ_k are tuned so that the equilibrium distribution without the interaction terms ℓ_k matches a given power spectrum. Later, when initializing the second moment equation, the initial value is taken to be the equilibrium value, again without interactions. Thus, the effect of the interaction with the coarse grid quantities,

¹⁴The stochastic process can also be viewed as accounting for all other approximations which are implicitly made: the replacement of exact operators by coarse grid approximations in (5.75), the use of periodic boundary conditions for the fine cell dynamics, and the uncertainty in the initialization and re-initialization of the fine grid dynamics.

which are encoded in ℓ_k , is to color the fine grid statistics consistent with the coarse grid flow. In particular, when applying this method to stratified models, $\ell_k > 0$ at scales where the flow is baroclinically unstable, resulting in growth of the primed quantities.

At this point, the effective subgrid dynamics as seen from the coarse grid is purely deterministic. Grooms and Majda (2014) found that it is necessary to take a large number of modes in the subgrid to match the correct spectral decay. To keep the computational cost low, and to account for the observation that real ocean eddies have anisotropies that vary in space and time, they select at random a direction in each subgrid cell, independent for each point in coarse space–time, and choose a one-dimensional spectral decomposition in this cell.

Grooms et al. (2015b) perform a detailed computational study of their stochastic superparameterization in a two-layer zonally reentrant channel mimicking the Antarctic Circumpolar Current. They compare the model with a deterministic Gent–McWilliams (GM) parameterization in a regime where mesoscale eddies are not resolved on the coarse grid and with an eddy-resolving high-resolution simulation. Their setting is similar to the two-layer quasigeostrophic model discussed in Section 5.3.2, with vorticity equations

$$\partial_t q_1 + [\psi_1, q_1] = -\frac{2}{\rho_0 H} \partial_y F(y) + \nu_2 \Delta^2 \psi_1, \quad (5.80a)$$

$$\partial_t q_2 + [\psi_2, q_2] = -r \Delta \psi_2 + \nu_2 \Delta^2 \psi_2, \quad (5.80b)$$

where ν_2 is Newtonian viscosity, r is the Ekman drag coefficient, and the layer potential vorticities are given by

$$q_1 = f_0 + \beta y + \Delta \psi_1 + \frac{1}{2} k_d^2 (\psi_2 - \psi_1) - k_e^2 \psi_1, \quad (5.81a)$$

$$q_2 = f_0 + \beta y + \Delta \psi_2 + \frac{1}{2} k_d^2 (\psi_1 - \psi_2) + \frac{2f_0}{H} h_b. \quad (5.81b)$$

The lateral boundary conditions are periodic in the zonal and stress-free in the meridional direction. This system is different from the two-layer equation (5.28) in the following way: The flow here is driven by a steady sinusoidal wind forcing $F(y)$.¹⁵ It includes explicit bottom topography h_b to avoid unrealistic spin-up of the mean current. A stretching term with coefficient $k_e = 1/L_e = f/\sqrt{gH}$ is included in the upper layer potential vorticity so that the model is formally valid to scales up to the external Rossby radius of deformation L_e . Finally, dissipation is second order as is common for relatively coarse resolution models; the eddy-resolving simulation and the fine grid dynamics, however, are set up with fourth-order dissipation.

In the coarse model, the divergence of the subgrid potential vorticity fluxes is introduced layerwise as explained in the single-layer setting; for details, see (Grooms and

¹⁵The steady wind stress tilts the layer interface through Ekman pumping. When it is sufficiently tilted, the flow becomes baroclinically unstable. So here, as before, the mean forcing maintains a pool of mean kinetic and available potential energies.

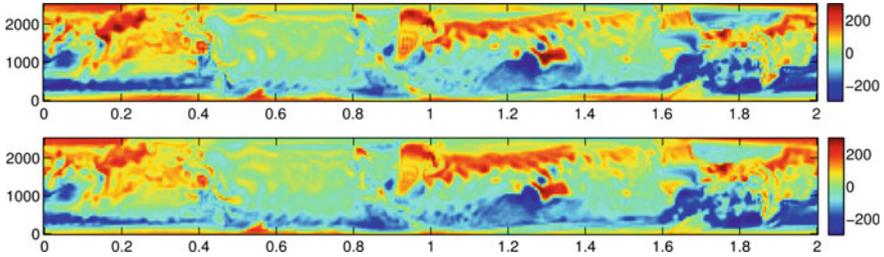


Fig. 5.3 Time mean bias of interface elevation in meters of the Gent–McWilliams parameterization (top) vs. the stochastic superparameterization (bottom). The zonal direction is shown in the horizontal in units of 10 000 km, the meridional direction in the vertical in units of kilometers. Graphs are adapted from Grooms et al. (2015b)

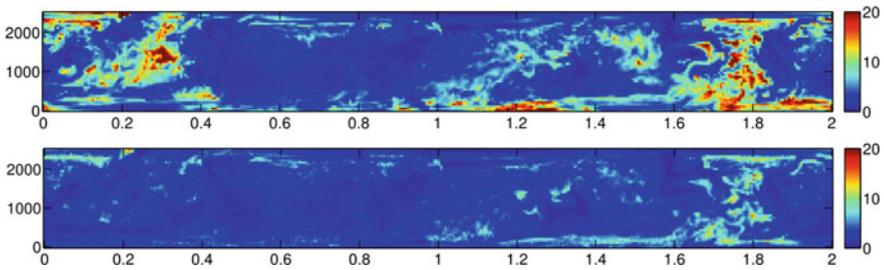


Fig. 5.4 Ratio of interface elevation time variance of the reference simulation over time variance corresponding to the data shown in Figure 5.3. The variance deficiency of the Gent–McWilliams parameterization (top) is significantly larger than the variance deficiency of the stochastic superparameterization (bottom). Graphs are adapted from Grooms et al. (2015b)

Majda 2014). In order to compare the performance of GM and SPP, the authors analyze the temporal statistics of the surface elevation at each fixed point; in particular, they compute the bias of the mean and the bias of the variance relative to the highly resolved reference simulation. While the time mean biases of both parameterizations are similar in magnitude and spatial pattern (Figure 5.3), the time variance of the stochastic superparameterization run is significantly closer to the reference simulation, even though both models are variance deficient (Figure 5.4).

While this model achieves a spectrally consistent proxy for the missing potential vorticity flux, it imposes locality and spatial uniformity of the subgrid model. In reality, however, as follows from the pattern of eddy kinetic energy in Grooms et al. (2015b), the eddy kinetic energy is (i) essentially non-uniform and (ii) does not correlate with the places of maximum instability—indeed, EKE spots are always downstream of places with maximal baroclinic instability. So, the main conceptual question is how to introduce non-trivial advection of subgrid quantities by the coarse flow.

Another question is how to make this approach practical. In essence, we need the dependence of mean subgrid fluxes as a function of velocities and quasigeostrophic

PV gradients. For a two-layer quasigeostrophic model, this can still be done. However, for continuous stratification and for the primitive equations, there is the vertical dependence which will soon require more computational effort than directly applying an eddy-resolving mesh. Another question is how to avoid overexciting gravity waves in a gravity wave-permitting model.

There may be possible simplifications. For example, since the fine grid dynamics depends only on the magnitude of resolved velocities, PV gradients, and the angle between the velocities and gradients, one may only consider a finite set of velocity values and PV gradients, and interpolate the results for the eddy flux divergence between the simulated patterns computed for this set. Such a lookup table may considerably reduce computational cost. Further, for primitive equation models, the subgrid may still be treated in quasigeostrophic approximation. Thus, it will be possible to represent PV gradients on the subgrid which is seen essential for providing a proper proxy for baroclinic instability, and cannot be easily done if the superparameterization is formulated in terms of primitive variables where information on PV gradients would be lost when going to the fine grid.

5.5 Other Closures

5.5.1 *The Mana–Zanna Parameterization of Ocean Mesoscale Eddies*

Mana and Zanna (2014) study the correlation of different functional forms for the eddy source term with a highly resolved direct numerical simulation, select the best candidate function, and match the remaining coefficients with the empirical data. More detailed tests in a double gyre configuration are reported in Zanna et al. (2017). In these two papers, the authors work in a 3-layer quasigeostrophic setting; possible extensions to primitive equation models are discussed and tested in Anstey and Zanna (2017). In the following, we describe the Mana–Zanna parameterization following the concise derivation later given by Grooms and Zanna (2017). We will present a slightly more general view which raises interesting possibilities for further optimization of the closure.

For simplicity, we restrict the discussion to the barotropic single-layer QG equations without β -effect. Working exclusively in the continuum setting on the plane, we define a time-independent coarsening operation via convolution with a filter kernel, i.e.,

$$\bar{\zeta}(\mathbf{x}) = \int_{\mathbb{R}^2} \phi_\delta(\mathbf{x} - \mathbf{y}) \zeta(\mathbf{y}) \, d\mathbf{y} \quad (5.82)$$

where ϕ_δ is a radial kernel with δ referring to the width of the filter. Applying this operation to the barotropic vorticity equation (5.4a), we can write

$$\overline{D_t \zeta} \equiv \partial_t \overline{\zeta} + [\overline{\psi}, \overline{\zeta}] = S + \overline{F} + D^* \overline{\zeta}, \quad (5.83)$$

where D^* is some coarsened dissipation operator and S denotes the eddy source term

$$S = [\overline{\psi}, \overline{\zeta}] - \overline{[\psi, \zeta]} + \overline{D\zeta} - D^* \overline{\zeta}. \quad (5.84)$$

In Mana and Zanna (2014), the authors seek to build a model for S in terms of the divergence of Rivlin–Ericksen stresses which originated in the description of non-Newtonian fluids with infinitesimal memory (see, e.g., Truesdell and Rajagopal 1999). These tensors satisfy material frame invariance and observer objectivity, properties required of a physical material law. For the barotropic vorticity equation, an exact implementation of an inviscid second-grade Rivlin–Ericksen fluid would correspond to

$$S = \alpha \overline{D_t \Delta \zeta} \quad (5.85)$$

which leads to the vorticity formulation of the Euler- α model further discussed in Section 5.5.2 below. Their study, however, finds that a better correlation is obtained by using

$$S = \alpha \Delta \overline{D_t \zeta} \quad (5.86)$$

which differs from (5.85) by nonlinear commutators, but preserves the property of frame invariance. They also find that the coefficient α on the right-hand side is negative, which precludes a straightforward interpretation as advection by a smoothed velocity field.

Grooms and Zanna (2017) provide *a posteriori* justification of (5.86) along the following lines. They argue that, after high-pass filtering, S is the dominant term on the right-hand side of (5.83).¹⁶ In particular, the Laplacian of S dominates the Laplacians of the other two terms so that

$$\Delta D_t \overline{\zeta} \approx \Delta S. \quad (5.87)$$

They proceed to show that S is highly correlated with ΔS , so that (5.87) implies (5.86).

Let us explore such correlations from a more general perspective. We define a family of abstract approximate Laplacians which includes the usual 5-point discrete Laplacian in two dimensions. Suppose that $\{\mu_\varepsilon\}_{\varepsilon>0}$ is a family of finite positive Borel measures on \mathbb{R}^n with $\text{supp } \mu_\varepsilon \subset B(x, \varepsilon)$, the ball centered at x with radius ε , satisfying

¹⁶This assumption was tested with harmonic dissipation. It would be questionable with higher-order dissipation so that, in general, the contribution from dissipation would need to be carried explicitly.

$$\mu_\varepsilon(\{x\}) = 0, \tag{5.88a}$$

$$\lim_{\varepsilon \rightarrow 0} \int_{B(x,\varepsilon)} y_i \, d\mu_\varepsilon(y) = 0, \tag{5.88b}$$

and

$$\lim_{\varepsilon \rightarrow 0} \int_{B(x,\varepsilon)} y_i y_j \, d\mu_\varepsilon(y) = 2 \delta_{ij} \tag{5.88c}$$

for all $i, j = 1, \dots, n$. In particular, various normalized symmetric measures, including point measures and surface Lebesgue measures in lower dimensions, satisfy these conditions. Then,

$$\Delta S(x) = \lim_{\varepsilon \rightarrow 0} |\mu_\varepsilon| (Av_\varepsilon(S) - S(x)), \tag{5.89}$$

where $|\mu_\varepsilon| = \mu_\varepsilon(\mathbb{R}^n)$ and

$$Av_\varepsilon(S) = \frac{1}{|\mu_\varepsilon|} \int_{B(x,\varepsilon)} S(y) \, d\mu_\varepsilon(y). \tag{5.90}$$

Assume now that S is a homogeneous isotropic δ -correlated Gaussian random field with variance σ^2 . Setting

$$\mathbf{w} = \begin{pmatrix} S(x) \\ Av_\varepsilon(S) - S(x) \end{pmatrix} \tag{5.91}$$

and fixing $\varepsilon > 0$ at a small finite value, we find that the covariance matrix of \mathbf{w} is given by

$$\Sigma = \text{Cov}[\mathbf{w}, \mathbf{w}] = \sigma^2 \begin{pmatrix} 1 & -1 \\ -1 & (b+1) \end{pmatrix}, \tag{5.92}$$

with $b = |\mu_\varepsilon^2|/|\mu_\varepsilon|^2$, where $|\mu_\varepsilon^2| = \sum_{y \in B(x,\varepsilon)} \mu_\varepsilon^2(\{y\})$.¹⁷

The eigenvalue ratio corresponding to the subdominant principal component of Σ is given by

$$r \equiv \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{b - \sqrt{b^2 + 4} + 2}{2b + 4}. \tag{5.93}$$

It quantifies the fraction of variance not explained by a linear relationship between the components of \mathbf{w} . When μ_ε does not have point measure components, $b = 0$ and consequently the eigenvalue ratio $r = 0$, indicating perfect correlation between the components of \mathbf{w} . The largest value of b in the class of point measures with equal weights corresponds to the 4-point Laplacian where μ_ε consists of three-unit Dirac

¹⁷The sum in this definition is countable and convergent as $|\mu_\varepsilon| = \mu_\varepsilon(B(x, \varepsilon)) < \infty$. Thus, $|\mu_\varepsilon^2| = 0$ if and only if the measure μ_ε has no nonzero point measure components, which is the case for measures that are absolutely continuous with respect to the Lebesgue measure in the ball or on the sphere.

masses at angles $0, 2\pi/3$, and $4\pi/3$. In this case, $b = 1/3$ and $r = (7 - \sqrt{37})/14 \approx 0.066$. For the usual 5-point stencil as considered in Grooms and Zanna (2017), $b = 1/4$ and $r = 9 - \sqrt{65}/18 \approx 0.052$.

Thus, even for relatively concentrated measures, a major fraction of the variance is explained by a linear relationship between S and the approximate Laplacian of S as defined via the right-hand expression in (5.89) for finite ε . The constant of proportionality is the ratio of the components of the principal eigenvector of Σ , i.e.,

$$S(x) \approx \frac{1}{2} (b - \sqrt{4 + b^2}) (\text{Av}_\varepsilon(S) - S(x)). \quad (5.94)$$

In the concrete case of the 5-point Laplacian, we find

$$S \approx -\frac{1}{8} (1 - \sqrt{65}) \frac{\varepsilon^2}{4} \Delta S = -(c\varepsilon)^2 \Delta S \approx -(c\varepsilon)^2 \Delta D_i \bar{\zeta} \quad (5.95)$$

where $c \approx 0.469782$, which is close to the empirical value found by Mana and Zanna (2014).¹⁸

It is clear from the argument above that the eigenvalue ratio improves when the measure becomes less localized. On the other hand, the assumption of S being a δ -correlated random field must break down on small scales; we expect the decorrelation length to be at or slightly larger than the grid scale. Thus, it should be possible to replace the Laplacian in the argument above with a discrete operator Λ whose stencil nodes are at least a decorrelation length apart and which is effectively acting as a high-pass filter. The form of Λ can then be optimized for its eigenvalue ratio. In this context, we remark that the approximation made in (5.87) does not seem necessary to proceed, as $\Lambda \bar{F}$ and $\Lambda D^* \bar{\zeta}$ are readily computable. We believe that this question is worth further investigation.

A different line of reasoning might be based on a random field model for S with finite spatial correlations. Assuming a given spectrum for S , the characterization of the two-point correlation function via the Wiener–Khinchin theorem (see, e.g., Yaglom 1987) could still allow us to compute the covariance matrix Σ explicitly and subsequently optimize the filter Λ . Finally, a detailed analysis of the structure of S is warranted. Grooms et al. (2015a) provide an argument that the spectrum of S grows like k^5 , which can likely only be true on a limited range of scales as a perfectly δ -correlated random field should have a flat spectrum. Thus, in particular the details of spatial correlation near the grid scale require attention.

¹⁸The slight discrepancy in the value of c as compared with Grooms and Zanna (2017) is due to the different normalizations of the vector \mathbf{w} . Ideally, the components of \mathbf{w} should have unit variance, but both our choices are close enough to being normalized for the point being made.

5.5.2 α -Models

The so-called α -models initially came up in the study of nonlinear waves, not in turbulence. What is now known as the Camassa–Holm equation was first discovered by Fuchssteiner and Fokas (1981) who sought completely integrable generalizations of the Korteweg–de Vries (KdV) equation with a bi-Hamiltonian structure. It was independently re-derived by Camassa and Holm (1993)—for a more detailed exposition, see Camassa et al. (1994)—as a next order correction to the KdV equation in small amplitude expansion of unidirectional surface waves in irrotational shallow water. Camassa and Holm’s work attracted a lot of attention as, in addition to integrability and bi-Hamiltonian structure, they found a family of peaked soliton solutions. Solutions of the Camassa–Holm equation can be seen as geodesics on the diffeomorphism group with respect to a right-invariant H^1 -metric (Kouranbaeva 1999). The striking parallel to Arnold’s (1966) view of ideal three-dimensional hydrodynamics as geodesic flow on the volume-preserving diffeomorphism group endowed with an L^2 -metric was pointed out by Holm et al. (1998) who, replacing the L^2 with an H^1 -metric, obtained a hydrodynamic analog to the Camassa–Holm equations which is now known as the Euler- α equations or the Lagrangian-averaged Euler equations. In velocity–momentum variables, they read

$$\partial_t \mathbf{v} - \mathbf{u} \times (\nabla \times \mathbf{v}) + \nabla p = 0, \quad (5.96a)$$

$$\mathbf{v} = (1 - \alpha^2 \Delta) \mathbf{u}, \quad (5.96b)$$

$$\nabla \cdot \mathbf{u} = 0. \quad (5.96c)$$

The Euler- α equations arise from the “kinetic energy” Lagrangian

$$L_\alpha = \frac{1}{2} \int |\mathbf{u}|^2 + \alpha^2 |\nabla \mathbf{u}|^2 dx, \quad (5.97)$$

which is a constant of the motion.

The connection to turbulence was made soon after its discovery, based on a number of observations. The momentum \mathbf{v} is transported by a velocity field \mathbf{u} which is smoother than the momentum; see (5.96b). This was seen as analogous to Reynolds averaging, even though the two operations are not equivalent; further, the non-viscous terms take the form of a Rivlin–Erikson tensor, so that, in their inviscid form, they coincide with the equations of motion for a non-Newtonian fluid of second grade (Foias et al. 2001). Analytically, the Euler- α equations possess properties which are notably lacking in ideal and Newtonian fluids: In two dimensions, the Euler- α model has unique global point vortex solutions (Oliver and Shkoller 2001), and in three dimensions, the viscous α -equations have global classical solutions (Marsden and Shkoller 2001; Foias et al. 2002). We note that it is not *a priori* clear how to add viscosity to (5.96): The references quoted so far argue that momentum should be diffused; see Chen et al. (1999a) for a discussion of this issue. The classical equations of a viscous second-grade fluid, in contrast, diffuse velocity—a mathematically

weaker form of dissipation so that, correspondingly, the global existence of solution is only known for small initial data (Cioranescu and Girault 1997), much like the situation for the Navier–Stokes equations in three dimensions.

Several authors have given derivations of the Euler- α equations as the equations of motion for some notion of a Lagrangian mean flow. Holm (1999, 2002) recognized a close connection between Lagrangian averaging and the generalized Lagrangian mean (GLM) of Andrews and McIntyre (1978). To provide closure, Holm assumes that first-order fluctuations in a small amplitude expansion are parallel-transported by the mean flow—an assumption he refers to as a *Taylor hypothesis* in analogy with G.I. Taylor’s observation that turbulent fluctuations are correlated in the downstream direction of a flow (Taylor 1938). Marsden and Shkoller (2003), in contrast, assume that first-order fluctuations are transported as a vector field and that parallel transport of second-order fluctuations is, on average, orthogonal to the velocity field. Recently, Gilbert and Vanneste (2018) have pointed out that a geometric view of the Lagrangian mean fixes the higher-order closure conditions. In this framework, the Euler- α equations emerge from Lagrangian averaging under the minimal set of assumptions that (i) the averaged map is the minimizer of geodesic distance, (ii) first-order fluctuations are statistically isotropic, and (iii) first-order fluctuations are transported by the mean flow as a vector field (Oliver 2017; Badin et al. 2018).

The numerical evidence supporting the use of α -models is mixed. Early numerical studies for homogeneous turbulence were encouraging (Chen et al. 1999b; Mohseni et al. 2003). The underlying idea has also been ported to rotating geophysical fluid flow (Holm and Nadiga 2003) and used in various test cases (Hecht et al. 2008a; Aizinger et al. 2015). Careful comparative studies for two-dimensional quasigeostrophic turbulence, however, show that the α -model perturbs the dynamics of two-dimensional turbulence. In particular, it suffers from accumulation of enstrophy at small scales (Lunasin et al. 2007; Graham and Ringler 2013) and has inferior correlation with an empirically observed subgrid stress tensor (Mana and Zanna 2014), where the computationally observed behavior is close to (5.95), a relationship that is similar, but not identical to the α -model closure. In addition, as the inversion of the Helmholtz filter in (5.96b) is non-local, it is not appealing for use in a full ocean model. We note, however, that the idea of filtering in a geometrically intrinsic setting is more general than what is usually pursued and may have some merit even in the setting of the nearly geostrophic turbulence in mesoscale ocean dynamics.

In the final part of the section, we shall sketch a possible nonstandard interpretation of the α -model dynamics as a model for two-dimensional turbulence. For simplicity, we return to the barotropic vorticity equation of Section 5.3.1 with $\beta = 0$ and initially ignore forcing and dissipation. In this setting, the model coincides with the well-studied two-dimensional Euler- α equation whose vorticity dynamics reads

$$\partial_t \xi + [\psi, \xi] = 0, \quad (5.98a)$$

$$\xi = L_\alpha \Delta \psi, \quad (5.98b)$$

where (5.96b) corresponds to the choice $L_\alpha = 1 - \alpha \Delta$, but L_α could also be a more general operator defined via a Fourier symbol $\ell_\alpha(k)$. The α -energy at wavenumber \mathbf{k} is $\mathcal{E}_k = -\frac{1}{2} \psi_k^* \xi_k$, and the α -enstrophy is given by $\mathcal{Z}_k = k^2 \ell_\alpha(k) \mathcal{E}_k$; system (5.98) conserves total α -energy and α -enstrophy.

Now suppose that α -wavenumber k corresponds to a different physical wavenumber $\kappa(k)$ and that there is a corresponding physical energy

$$E(\kappa) = \mathcal{E}(k)/h(k). \quad (5.99)$$

A straightforward computation shows that total physical energy and enstrophy are conserved if and only if

$$\kappa^2(k) = \ell_\alpha(k) k^2 \quad (5.100a)$$

and

$$h(k) = \frac{1}{2} k \ell'_\alpha(k) + \ell_\alpha(k). \quad (5.100b)$$

Looking at the detailed triad interactions of the α -model, we find transfer rate relations similar to (5.15) where the rate of nonlinear energy transfer is with respect to α -wavenumbers, whereas the prefactors on the right-hand side are satisfied with respect to physical wavenumbers. Thus, in general, it is not even approximately true that the α -triad picture corresponds to the physical triad picture under the wavenumber mapping implied by energy and enstrophy conservation. However, there is one class of triads for which this is approximately the case: when one leg of the triad is in the low wavenumbers and two legs are in the high-wavenumber range; to be definite, we take $p \ll k < q$ and set $\delta = p/k \ll 1$. We might call such interactions *catalytic triads* as (5.15) shows that there is an $O(1)$ energy exchange between modes \mathbf{k} and \mathbf{q} while mode \mathbf{p} exchanges energy only at a rate $O(\delta)$. In other words, mode \mathbf{p} takes the role of a catalyst, mediating the transfer of energy in the high-wavenumber regime while not participating in it to leading order. Provided the turbulent regime is dominated by catalytic triads (which is not the classical KLB picture, but it is likely that these triads are key players in the inverse cascade), then under mild assumptions on ℓ_α , an α -model can be interpreted as representing the physical interactions under the mappings (5.100) up to relative errors in rates and mapped wave numbers of $O(\delta)$. The details of this computation involve only elementary estimates and shall be omitted here.

Thus, to interpret the α -dynamics consistently via the remapping of wavenumbers, the same map must be applied when adding forcing and dissipation terms. Dissipation in the α -model momentum equation, in particular, should take the form $D(\kappa(k))$. This corresponds to momentum rather than velocity diffusion and thus coincides with the dissipation operator typically used in connection with α -models as reviewed earlier in this section.

Finally, to consistently interpret the energy spectrum, it must be mapped back to physical wavenumbers. In the comparison of Graham and Ringler (2013), for

example, no such map is applied. This constitutes an interesting open question as, to our knowledge, such analysis has never been done. A related open problem is to formulate the α -model subgrid closure mapped to physical wavenumbers.

5.6 Concluding Remarks

In this chapter, we have reviewed the foundations of geostrophic turbulence and its implications for ocean models in the eddy-permitting regime. In the past decade, a number of authors have looked at the problem of effective parameterizations for subgrid eddy activity and for the resulting backscatter of energy into the resolved grid. Most of the detailed testing so far has been done in the context of quasigeostrophic layer models, with increased attention to full primitive equation setups in recent years.

Our selection of parameterizations for close discussion is necessarily incomplete, highlighting recent developments in favor of older ideas, putting an emphasis on mathematical structure toward systematic, or even rigorous, analysis, and with a view toward applicability for a new generation of global circulation models featuring irregular grids with spatially varying grid resolution which rules out approaches that require explicit Fourier transforms or other constructs tied to a regular grid.

To a large extent, the ideas expressed here are exploratory. None of the parameterizations described here is widely used in operational models so that a major development cycle of introducing more energy-consistent parameterizations lies ahead. It is also not clear which of these approaches will be the most fruitful in the long run or whether some new or possibly old ideas will prevail.

Such old ideas could include the anticipated vorticity method of Sadourny and Basdevant (1985) which seeks to introduce a force $-D\mathbf{k} \times \mathbf{u}$ such that, for example, when D is chosen as an upwind estimate of the layer potential vorticity, the scheme conserves energy exactly while dissipating enstrophy. In practice, this approach is insufficient as it does not remove the component of small-scale numerical noise in \mathbf{u} that does not project on curl as required for numerical stability. Graham and Ringler (2013) report that first-order anticipated vorticity results in either an excess of energy at all scales or dissipation of enstrophy across a too large portion of the spectrum; they suggest that applying a high-order spatial operator within the anticipated PV formalism may solve this issue, but at the expense of easy implementability in current GCMs. Yet, the underlying idea is interesting as the mathematically most direct way to reconcile energy conservation with enstrophy dissipation.

The classical development of Smagorinsky closures has been central in the modeling and simulation of turbulent flow regimes; see, e.g., the review by Meneveau and Katz (2000). However, it is not directly applicable to typical ocean regimes where, due to the scales at which forcing, instabilities, and dissipation act, one is often not in a self-similar scaling regime which is a prerequisite of LES and Smagorinsky-type closures. Dynamical Smagorinsky closures (when the subgrid viscosity is computed by applying an additional coarsening filter and fitting the difference between this and

the original filter to the simulated stresses) could be of interest, although even with these techniques the lack of self-similarity may be an issue.

Stochastic modeling of subgrid interactions and backscatter has been developed, based on the direct interaction approximation of Kraichnan (1959), by Frederiksen and co-workers (Frederiksen and Davies 1997; O’Kane and Frederiksen 2008; Kitsios et al. 2013, 2014, 2016). While their work involves a detailed analysis of unresolved eddy–eddy interactions, it also heavily relies on spectral language, so that it is not clear how applicable this approach is in the context of complex geometries and possibly non-uniform grids and what the trade-offs in terms of skill vs. computational expense are. We also remark that there is similarity between the expression for the subgrid drain dissipation matrix in Kitsios et al. (2013) and the estimation of a dynamic Smagorinsky coefficient in the spirit of Germano et al. (1991).

For systems with an explicit fast–slow scale separation, it may be possible to model the fast timescale component with a stochastic process and use *stochastic mode reduction* to reduce the system to a stochastic equation on the slow timescale. Such methods are reviewed in Section 5 of Franzke et al. (2019). However, it is completely open whether this approach is applicable to subscale modeling in geostrophic turbulence where there is no clear scale separation and whether these techniques scale up to full ocean models.

References

- Aizinger, V., Korn, P., Giorgetta, M., Reich, S.: Large-scale turbulence modelling via α -regularisation for atmospheric simulations. *J. Turbul.* **16**(4), 367–391 (2015)
- Aluie, H.: Scale decomposition in compressible turbulence. *Phys. D* **247**(1), 54–65 (2013)
- Aluie, H., Kurien, S.: Joint downscale fluxes of energy and potential enstrophy in rotating stratified Boussinesq flows. *Europhys. Lett.* **96**(4), 44006 (2011)
- Andrews, D.G., McIntyre, M.E.: An exact theory of nonlinear waves on a Lagrangian-mean flow. *J. Fluid Mech.* **89**(4), 609–646 (1978)
- Anstey, J.A., Zanna, L.: A deformation-based parametrization of ocean mesoscale eddy Reynolds stresses. *Ocean Model.* **112**, 99–111 (2017)
- Arnold, V.: Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l’hydrodynamique des fluides parfaits. *Ann. Inst. Fourier* **16**(1), 319–361 (1966)
- Augier, P., Lindborg, E.: A new formulation of the spectral energy budget of the atmosphere, with application to two high-resolution general circulation models. *J. Atmos. Sci.* **70**, 2293–2308 (2013)
- Badin, G.: On the role of non-uniform stratification and short-wave instabilities in three-layer quasi-geostrophic turbulence. *Phys. Fluids* **26**(9), 096603 (2014)
- Badin, G., Oliver, M., Vasilyevych, S.: Geometric Lagrangian averaged Euler–Boussinesq and primitive equations. *J. Phys. A: Math. Theor.* **51**(45), 455501 (2018)
- Batchelor, G.K.: Computation of energy spectrum in homogeneous two-dimensional turbulence. *Phys. Fluids Suppl.* **12**, II:233–239 (1969)
- Berloff, P.: Dynamically consistent parameterization of mesoscale eddies. Part I: simple model. *Ocean Model.* **87**, 1–19 (2015)
- Berloff, P.: Dynamically consistent parameterization of mesoscale eddies. Part II: eddy fluxes and diffusivity from transient impulses. *Fluids* **1**(3), 22 (2016)
- Callies, J., Ferrari, R.: Interpreting energy and tracer spectra of upper-ocean turbulence in the submesoscale range (1–200 km). *J. Phys. Oceanogr.* **43**, 2456–2474 (2013)

- Camassa, R., Holm, D.D.: An integrable shallow water equation with peaked solitons. *Phys. Rev. Lett.* **71**(11), 1661–1664 (1993)
- Camassa, R., Holm, D.D., Hyman, J.M.: A new integrable shallow water equation. *Adv. Appl. Mech.* **31**(C), 1–33 (1994)
- Capet, X., McWilliams, J.C., Molemaker, M.J., Shchepetkin, A.F.: Mesoscale to submesoscale transition in the California current system. Part III: energy balance and flux. *J. Phys. Oceanogr.* **38**, 2256–2269 (2008)
- Cessi, P.: An energy-constrained parameterization of eddy buoyancy flux. *J. Phys. Oceanogr.* **38**, 1807–1819 (2008)
- Chaouat, B.: Simulations of turbulent rotating flows using a subfilter scale stress model derived from the partially integrated transport modeling method. *Phys. Fluids* **24**(4), 045108 (2012)
- Charney, J.G.: Geostrophic turbulence. *J. Atmos. Sci.* **28**(6), 1087–1095 (1971)
- Chen, S., Foias, C., Holm, D.D., Olson, E., Titi, E.S., Wynne, S.: A connection between the Camassa-Holm equations and turbulent flows in channels and pipes. *Phys. Fluids* **11**(8), 2343–2353 (1999a)
- Chen, S., Holm, D.D., Margolin, L.G., Zhang, R.: Direct numerical simulations of the Navier-Stokes alpha model. *Phys. D* **133**(1–4), 66–83 (1999b)
- Cioranescu, D., Girault, V.: Weak and classical solutions of a family of second grade fluids. *Int. J. Non-Linear Mech.* **32**(2), 317–335 (1997)
- Cooper, F.C.: Optimisation of an idealised primitive equation ocean model using stochastic parameterization. *Ocean Model.* **113**, 187–200 (2017)
- Cooper, F.C., Zanna, L.: Optimisation of an idealised ocean model, stochastic parameterisation of sub-grid eddies. *Ocean Model.* **88**, 38–53 (2015)
- Daly, B., Harlow, F.: Transport equations in turbulence. *Phys. Fluids* **13**, 2634–2649 (1970)
- Danilov, S.: Non-universal features of forced 2D turbulence in the energy and enstrophy ranges. *Discret. Contin. Dyn. Syst. Ser. B* **5**(1), 67–78 (2005)
- Danilov, S.: Ocean modeling on unstructured meshes. *Ocean Model.* **69**, 195–210 (2013)
- Danilov, S., Gurarie, D.: Forced two-dimensional turbulence in spectral and physical space. *Phys. Rev. E* **63**(6), 061208 (2001)
- Danilov, S., Gurarie, D.: Scaling, spectra and zonal jets in beta-plane turbulence. *Phys. Fluids* **16**, 2592–2603 (2004)
- Danilov, S., Wang, Q.: Resolving eddies by local mesh refinement. *Ocean Model.* **93**, 75–83 (2015)
- Deardorff, J.W.: The use of subgrid transport equations in a three-dimensional model of atmospheric turbulence. *J. Fluids Eng.* **95**(3), 429–438 (1973)
- Duan, J., Nadiga, B.T.: Stochastic parameterization for large eddy simulation of geophysical flows. *Proc. Am. Math. Soc.* **135**(4), 1187–1196 (2007)
- Eden, C.: Parameterising meso-scale eddy momentum fluxes based on potential vorticity mixing and a gauge term. *Ocean Model.* **32**(1–2), 58–71 (2010)
- Eden, C.: Closing the energy cycle in an ocean model. *Ocean Model.* **101**, 30–42 (2016)
- Fjørtoft, R.: On the changes in the spectral distribution of kinetic energy for twodimensional, nondivergent flow. *Tellus* **5**(3), 225–230 (1953)
- Foias, C., Holm, D.D., Titi, E.S.: The Navier-Stokes-alpha model of fluid turbulence. *Phys. D* **152/153**, 505–519. *Advances in nonlinear mathematics and science* (2001)
- Foias, C., Holm, D.D., Titi, E.S.: The three dimensional viscous Camassa-Holm equations, and their relation to the Navier-Stokes equations and turbulence theory. *J. Dyn. Differ. Equ.* **14**(1), 1–35 (2002)
- Fox-Kemper, B., Menemenlis, D.: Can large eddy simulation techniques improve mesoscale rich ocean models? In: Hecht, M.W., Hasumi, H. (eds.) *Ocean Modeling in an Eddy Regime*, pp. 319–337. American Geophysical Union (2008)
- Franzke, C., O’Kane, T., Berner, J., Williams, P., Lucarini, V.: Stochastic climate theory and modelling. *WIREs Clim. Chang.* **6**, 63–78 (2015)
- Franzke, C.L.E., Oliver, M., Rademacher, J.D.M., Badin, G.: Multi-scale methods for geophysical flows. This volume, Chapter 1 (2019)

- Frederiksen, J.S., Davies, A.G.: Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.* **54**, 2475–2492 (1997)
- Fuchssteiner, B., Fokas, A.S.: Symplectic structures, their Bäcklund transformations and hereditary symmetries. *Phys. D* **4**(1), 47–66 (1981)
- Gassmann, A.: A global hexagonal C-grid non-hydrostatic dynamical core (ICON-IAP) designed for energetic consistency. *Q. J. Roy. Meteorol. Soc.* **139**, 152–175 (2013)
- Gent, P.R.: The Gent-McWilliams parameterization: 20/20 hindsight. *Ocean Model.* **39**(1–2), 2–9 (2011)
- Gent, P.R., McWilliams, J.C.: Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.* **20**, 150–155 (1990)
- Germano, M., Piomelli, U., Moin, P., Cabot, W.H.: A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids A* **3**(7), 1760–1765 (1991)
- Gilbert, A.D., Vanneste, J.: Geometric generalised Lagrangian-mean theories. *J. Fluid Mech.* **839**, 95–134 (2018)
- Gkioulekas, E., Tung, K.K.: A new proof on net upscale energy cascade in two-dimensional and quasi-geostrophic turbulence. *J. Fluid Mech.* **576**, 173–189 (2007)
- Graham, J.P., Ringler, T.: A framework for the evaluation of turbulence closures used in mesoscale ocean large-eddy simulations. *Ocean Model.* **65**, 25–39 (2013)
- Grooms, I., Lee, Y., Majda, A.J.: Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbulence. *Multiscale Model. Sim.* **13**(3), 1001–1021 (2015a)
- Grooms, I., Majda, A.J.: Efficient stochastic superparameterization for geophysical turbulence. *Proc. Nat. Acad. Sci.* **110**, 4464–4469 (2013)
- Grooms, I., Majda, A.J.: Stochastic superparameterization in quasigeostrophic turbulence. *J. Comput. Phys.* **271** (2014)
- Grooms, I., Majda, A.J., Smith, K.S.: Stochastic superparameterization in a quasigeostrophic model of the Antarctic Circumpolar Current. *Ocean Model.* **85**, 1–15 (2015b)
- Grooms, I., Zanna, L.: A note on ‘Toward a stochastic parameterization of ocean mesoscale eddies’. *Ocean Model.* **113**, 30–33 (2017)
- Hallberg, R.: Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Model.* **72**, 92–103 (2013)
- Hecht, M.W., Holm, D.D., Petersen, M.R., Wingate, B.A.: The LANS- α and Leray turbulence parameterizations in primitive equation ocean modeling. *J. Phys. A: Math. Theor.* **41**(34), 344009 (2008a)
- Hecht, M.W., Hunke, E., Maltrud, M., Petersen, M.R., Wingate, B.A.: Lateral mixing in the eddy-regime and a new broad-ranging formulation. In: Hecht, M.W., Hasumi, H., (eds.) *Ocean Modeling in an Eddy Regime*, pp. 339–352. American Geophysical Union (2008b)
- Held, I.M., Larichev, V.D.: A scaling theory for horizontally homogeneous, baroclinically unstable flow on a beta-plane. *J. Atmos. Sci.* **53**, 946–952 (1996)
- Holm, D.D.: Fluctuation effects on 3D Lagrangian mean and Eulerian mean fluid motion. *Phys. D* **133**(1), 215–269 (1999)
- Holm, D.D.: Averaged Lagrangians and the mean effects of fluctuations in ideal fluid dynamics. *Phys. D* **170**(3), 253–286 (2002)
- Holm, D.D., Marsden, J.E., Ratiu, T.S.: The Euler-Poincaré equations and semidirect products with applications to continuum theories. *Adv. Math.* **137**(1), 1–81 (1998)
- Holm, D.D., Nadiga, B.T.: Modeling mesoscale turbulence in the barotropic double-gyre circulation. *J. Phys. Oceanogr.* **33**(11), 2355–2365 (2003)
- Ilicak, M., Adcroft, A.J., Griffies, S.M., Hallberg, R.W.: Spurious diapycnal mixing and the role of momentum closure. *Ocean Model.* **45**, 37–58 (2012)
- Jansen, M.F., Held, I.M.: Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Model.* **80**, 36–48 (2014)
- Jansen, M.F., Held, I.M., Adcroft, A.J., Hallberg, R.: Energy budget-based backscatter in an eddy permitting primitive equation model. *Ocean Model.* **94**, 15–26 (2015)

- Kitsios, V., Frederiksen, J., Zidikheri, M.: Scaling laws for parameterisations of subgrid eddy-eddy interactions in simulations of oceanic circulations. *Ocean Model.* **68**, 88–105 (2013)
- Kitsios, V., Frederiksen, J.S., and Zidikheri, M.J.: Scaling laws for parametrizations of subgrid interactions in simulations of oceanic circulations. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **372**(2018) (2014)
- Kitsios, V., Frederiksen, J.S., Zidikheri, M.J.: Theoretical comparison of subgrid turbulence in atmospheric and oceanic quasi-geostrophic models. *Nonlinear Proc. Geophys.* **23**, 95–105 (2016)
- Klein, P., Hua, B.L., Lapeyre, G., Capet, X., Le Gentil, S., Sasaki, H.: Upper-ocean turbulence from high-resolution 3D simulations. *J. Phys. Oceanogr.* **38**, 1748–1763 (2008)
- Klingbeil, K., Burchard, H., Danilov, S., Goetz, C., Iske, A.: Reducing spurious diapycnal mixing in ocean models. This volume, Chapter 8 (2019)
- Korn, P.: Formulation of an unstructured grid model for global ocean dynamics. *J. Comput. Phys.* **339**, 525–552 (2017)
- Kouranbaeva, S.: The Camassa-Holm equation as a geodesic flow on the diffeomorphism group. *J. Math. Phys.* **40**(2), 857–868 (1999)
- Kraichnan, R.H.: The structure of isotropic turbulence at very high Reynolds numbers. *J. Fluid Mech.* **5**(4), 497–543 (1959)
- Kraichnan, R.H.: Inertial ranges in two-dimensional turbulence. *Phys. Fluids* **10**, 1417–1423 (1967)
- Lapeyre, G.: What vertical mode does the altimeter reflect? On the decomposition in baroclinic modes and a surface-trapped mode. *J. Phys. Oceanogr.* **39**, 2857–2874 (2009)
- Large, W.G., McWilliams, J.C., Doney, S.C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.* **32**(4), 363–403 (1994)
- Leith, C.E.: Diffusion approximation for two-dimensional turbulence. *Phys. Fluids* **11**, 671–673 (1968)
- Leith, C.E.: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.* **28**, 145–161 (1971)
- Lunasin, E., Kurien, S., Taylor, M.A., Titi, E.S.: A study of the Navier-Stokes- α model for two-dimensional turbulence. *J. Turbul.* **8**, N30 (2007)
- Maltrud, M.E., Vallis, G.K.: Energy and enstrophy transfer in numerical simulations of two-dimensional turbulence. *Phys. Fluids A* **5**(7), 1760–1775 (1993)
- Mana, P.P., Zanna, L.: Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Model.* **79**, 1–20 (2014)
- Marsden, J.E., Shkoller, S.: Global well-posedness for the Lagrangian averaged Navier-Stokes (LANS- α) equations on bounded domains. *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.* **359**(1784), 1449–1468 (2001)
- Marsden, J.E., Shkoller, S.: The anisotropic Lagrangian averaged Euler and Navier-Stokes equations. *Arch. Ration. Mech. Anal.* **166**(1), 27–46 (2003)
- Marshall, D.P., Adcroft, A.J.: Parameterization of ocean eddies: potential vorticity mixing, energetics and Arnold’s first stability theorem. *Ocean Model.* **32**(3–4), 188–204 (2010)
- Marshall, D.P., Maddison, J.R., Berloff, P.S.: A framework for parameterizing eddy potential vorticity fluxes. *J. Phys. Oceanogr.* **42**(4), 539–557 (2012)
- Meneveau, C., Katz, J.: Scale-invariance and turbulence models for large-eddy simulation. *Ann. Rev. Fluid Mech.* **32**(1), 1–32 (2000)
- Moffatt, H.: Note on the triad interactions of homogeneous turbulence. *J. Fluid Mech.* **741**, R3 (2014)
- Mohammadi-Aragh, M., Klingbeil, K., Brüggemann, N., Eden, C., Burchard, H.: The impact of advection schemes on restratification due to lateral shear and baroclinic instabilities. *Ocean Model.* **94**, 112–127 (2015)
- Mohseni, K., Kosović, B., Shkoller, S., Marsden, J.E.: Numerical simulations of the Lagrangian averaged Navier-Stokes equations for homogeneous isotropic turbulence. *Phys. Fluids* **15**(2), 524–544 (2003)
- O’Kane, T.J., Frederiksen, J.S.: Statistical dynamical subgrid-scale parameterizations for geophysical flows. *Phys. Scripta* **2008**(T132), 014033 (2008)

- Oliver, M.: Lagrangian averaging with geodesic mean. *Proc. R. Soc. A* **473**(2207), 20170558, 9 (2017)
- Oliver, M., Shkoller, S.: The vortex blob method as a second-grade non-Newtonian fluid. *Commun. Partial Differ. Equ.* **26**(1–2), 295–314 (2001)
- Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., Weisheimer, A.: Stochastic parametrization and model uncertainty. Technical report, ECMWF (2009)
- Rhines, P.B.: Waves and turbulence on a beta-plane. *J. Fluid Mech.* **69**, 417–443 (1975)
- Rhines, P.B.: The dynamics of unsteady currents. In: Goldberg, E.A., McCane, I.N., O'Brien, J.J., Steele, J.H. (eds.) *The Sea*, vol 6, pp. 189–318. J. Wiley (1977)
- Ringler, T., Gent, P.: An eddy closure for potential vorticity. *Ocean Model.* **39**(1–2), 125–134 (2011)
- Sadourny, R., Basdevant, C.: Parameterization of subgrid scale barotropic and baroclinic eddies in quasi-geostrophic models: anticipated potential vorticity method. *J. Atmos. Sci.* **42**, 1353–1363 (1985)
- Salmon, R.: Baroclinic instability and geostrophic turbulence. *Geophys. Astrophys. Fluid Dyn.* **15**(1), 167–211 (1980)
- Salmon, R.: *Lectures on Geophysical Fluid Dynamics*. Oxford University Press (1998)
- San, O.: A dynamic eddy-viscosity closure model for large eddy simulations of two-dimensional decaying turbulence. *Int. J. Comput. Fluid D.* **28**, 363–382 (2014)
- Schiestel, R., Dejoan, A.: Towards a new partially integrated transport model for coarse grid and unsteady turbulent flow simulations. *Theor. Comput. Fluid Dyn.* **18**(6), 443–468 (2005)
- Schmidt, H., Schumann, U.: Coherent structure of the convective boundary layer derived from large-eddy simulations. *J. Fluid Mech.* **200**, 511–562 (1989)
- Schumann, U.: Subgrid scale model for finite difference simulations of turbulent flows in plane channels and annuli. *J. Comput. Phys.* **18**(4), 376–404 (1975)
- Schumann, U.: Subgrid length-scales for large-eddy simulation of stratified turbulence. *Theor. Comput. Fluid Dyn.* **2**(5–6), 279–290 (1991)
- Scott, R.B., Arbic, B.K.: Spectral energy fluxes in geostrophic turbulence: implications for ocean energetics. *J. Phys. Oceanogr.* **37**, 673–688 (2007)
- Smith, K.S., Boccaletti, G., Henning, C.C., Marinov, I.N., Tam, C.Y., Held, I.M., Vallis, G.K.: Turbulent diffusion in the geostrophic inverse cascade. *J. Fluid Mech.* **469**, 13–48 (2002)
- Stone, P.H.: A simplified radiative-dynamical model for the static stability of rotating atmospheres. *J. Atmos. Sci.* **29**, 405–418 (1972)
- Taylor, G.I.: The spectrum of turbulence. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **164**(919), 476–490 (1938)
- Truesdell, C., Rajagopal, K.: *An Introduction to the Mechanics of Fluids*. Birkhäuser (1999)
- Umlauf, L., Burchard, H.: A generic length-scale equation for geophysical turbulence models. *J. Marine Res.* **61**(2), 235–265 (2003)
- Vallis, G.K.: *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*. Cambridge University Press (2006)
- von Storch, J.-S., Badin, G., Oliver, M.: The interior energy pathway: inertial gravity wave emission by oceanic flows. This volume Chapter 2 (2019)
- Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., Jung, T., Schröter, J.: The finite element sea ice–ocean model (FESOM) v.1.4: formulation of an ocean general circulation model. *Geosci. Model Dev.* **7**(2), 663–693 (2014)
- Wunsch, C.: The vertical partition of oceanic horizontal kinetic energy. *J. Phys. Oceanogr.* **27**, 1770–1794 (1997)
- Xun, Q.-Q., Wang, B.-C.: A dynamic forcing scheme incorporating backscatter for hybrid simulation. *Phys. Fluids* **26**(7), 075104 (2014)
- Yaglom, A.M.: *Correlation Theory of Stationary and Related Random Functions. Basic Results*, vol. 1. Springer (1987)
- Zanna, L., Mana, P.P., Anstey, J., David, T., Bolton, T.: Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. *Ocean Model.* **111**, 66–80 (2017)

Chapter 6

Diagnosing and Parameterizing the Effects of Oceanic Eddies



Alexa Griesel, Julia Dräger-Dietel and Kerstin Jochumsen

Abstract Oceanic eddies, fluctuations on scales on the order of one km to hundreds of km, derive their energy primarily from baroclinic instability processes. Currently, climate models do not incorporate the space and time variability of the effects of eddies and sub-mesoscale processes in an energy-consistent way. Eddy diffusivities are specified without connection to the energy budget and, more fundamentally, it is unclear to what extent, where and on what scales the downgradient eddy diffusion model is appropriate at all. Rotational components of the eddy fluxes associated with the advective terms in the eddy variance equation are generally large, so that production and dissipation of eddy energy do not balance locally. We will review here the current understanding of the spatial and temporal variability of eddy diffusivities and eddy–mean flow interactions that have been inferred in both observations and eddying ocean models. A focus will be on Lagrangian particle statistics as an ideal tool to describe the effects of eddies on a time mean transport and to assess the limits and validities of the eddy diffusion model. Eddy diffusivity diagnostics and the current state of eddy parameterizations in ocean models will be discussed as well as prospects for energy-consistent parameterizations.

6.1 Introduction

Oceanic eddies, fluctuations on scales on the order of one km to hundreds of km, and with typical timescales on the order of a month, derive their energy primarily from baroclinic instability processes, a key component in the oceanic energy cycle

A. Griesel (✉) · J. Dräger-Dietel
Center for Earth System Research and Sustainability (CEN), Universität Hamburg,
Hamburg, Germany
e-mail: alexa.griesel@uni-hamburg.de

J. Dräger-Dietel
e-mail: julia.draeger-dietel@uni-hamburg.de

K. Jochumsen
Federal Maritime and Hydrographic Agency, Hamburg, Germany
e-mail: kerstin.jochumsen@bsh.de

leading to the release of available potential energy. Most current ocean components of climate models are still rather coarse (about 1° horizontal resolution), do not resolve mesoscale eddies, and need to parameterize their effects on the mean variables. Here, we discuss the aspects and challenges of the parameterization of the effects of eddies as diffusion down mean tracer gradients (“Austauschansatz”). The focus is on the stratified oceanic interior, while mixed-layer eddies will not be specifically covered and parameterization in models that partly resolve eddies and their effect in the mean momentum equation is covered in Chapter 5 (Danilov et al. 2019).

The effects of mesoscale eddies in the ocean are usually parameterized using three different components. The first component, sometimes referred to as a skew diffusion, can be interpreted as advection of properties by an additional eddy-driven velocity, which can be related to mixing of isopycnal layer thickness (Gent and McWilliams 1990; Griffies 1998). The first component can be parameterized in terms of horizontal diffusion of buoyancy and leads to the dissipation of available potential energy. The second component, referred to as isopycnal diffusion, can be interpreted as turbulent mixing of properties along isopycnal surfaces. The third component, diapycnal diffusion, considers a possible eddy-induced diapycnal mixing of density and other properties.

In this chapter, we discuss the challenges that arise when diagnosing and parameterizing the eddy diffusivities that arise from the diffusive parameterizations, where the focus is on the isopycnal and skew diffusion that dominate in the oceanic interior.

To derive the different components of the parameterizations, we start with considering the conservation equation for a tracer T , for example, temperature, salinity, or buoyancy (in the Boussinesq approximation), where the overline is a time average, with averaging properties of a Reynolds average¹

$$\partial_t \overline{T} + \overline{\mathbf{u}} \cdot \nabla \overline{T} + \nabla \cdot \overline{\mathbf{u}'T'} = \overline{Q}, \quad (6.1)$$

where \overline{Q} are sources and sinks of tracer. Using the transformed Eulerian mean (TEM) framework (e.g., Andrews et al. 1987), the eddy tracer flux is decomposed into components across and along isosurfaces of the mean tracer, i.e., a diffusive and skew diffusive flux, $\overline{\mathbf{u}'T'} = -\mathbf{K} \cdot \nabla \overline{T} + \mathbf{B} \times \nabla \overline{T}$, which yields²

$$\partial_t \overline{T} + (\overline{\mathbf{u}} + \nabla \times \mathbf{B}) \cdot \nabla \overline{T} = \nabla \cdot \mathbf{K} \cdot \nabla \overline{T} + \overline{Q} \quad (6.2)$$

where \mathbf{B} is the vector streamfunction that defines the eddy-driven advection velocity $\nabla \times \mathbf{B}$, which depends on the tracer T , and \mathbf{K} is a 3×3 diffusivity tensor quantifying irreversible mixing of the tracer T . Note that in the eddy flux decomposition, an arbitrary rotational flux $\nabla \times \boldsymbol{\theta}$ that drops out in the divergence of the eddy flux can be added to the diffusive and skew fluxes :

¹We refer here to transient eddies, i.e., $T' = T - \overline{T}$ and $\mathbf{u}' = \mathbf{u} - \overline{\mathbf{u}}$, denote the fluctuations in time due to eddies. For statistically stationary flows, time averaging is equivalent to ensemble averaging.

²The divergence of the skew part of the eddy flux $\nabla \cdot \mathbf{F}_{\text{skew}} = \nabla \cdot (\mathbf{B} \times \nabla \overline{T}) = (\nabla \times \mathbf{B}) \cdot \nabla \overline{T} \equiv \mathbf{u}^* \cdot \nabla \overline{T}$ and hence can be expressed as advection by an eddy-driven velocity $\mathbf{u}^* = \nabla \times \mathbf{B}$.

$$\overline{\mathbf{u}'T'} = -\mathbf{K} \cdot \nabla \bar{T} + \mathbf{B} \times \nabla \bar{T} + \nabla \times \boldsymbol{\theta} \equiv \mathbf{F}_{\text{diff}}(\bar{T}) + \mathbf{F}_{\text{skew}}(\bar{T}) + \mathbf{F}_{\text{rot}}(\bar{T}). \quad (6.3)$$

We will discuss this gauge freedom and how it can be exploited in section 6.2.3.

6.1.1 Isopycnal and Diapycnal Diffusion

The form of the diffusion tensor \mathbf{K} depends on the magnitude of mixing in the different spatial directions. A simple form that accounts for the much larger horizontal scales of the eddies as compared to the vertical ones would be to choose $K_{11} = K_{22} = K_h$ and $K_{33} = K_v$, with $K_h \gg K_v$ and $K_{mn} = 0$ for $n \neq m$. More accurately though, the principal directions of mixing are expected to be in the directions along and across the local neutral density surfaces. They are defined as surfaces perpendicular to their normal vector $\mathbf{e} = \gamma \rho \nabla S - \alpha \rho \nabla \Theta$, where S is salinity, Θ is (conservative) temperature, α, γ are thermal expansion and saline contraction coefficients, respectively. A diffusion tensor that distinguishes between isopycnal and diapycnal components is, for example, of the form $\mathbf{K} = \mathbf{K}_d + \mathbf{K}_i$ with $\mathbf{K}_d = \kappa_d \frac{\mathbf{e}\mathbf{e}}{e^2}$ and $\mathbf{K}_i = \kappa_i \left(\mathbf{I} - \frac{\mathbf{e}\mathbf{e}}{e^2} \right)$. \mathbf{K} in this case is symmetric (i.e. $K_{mn} = K_{nm}$), and \mathbf{K}_i assumes mixing is isotropic in the along isopycnal direction. In small-isoneutral slope approximation,³ the isopycnal diffusion tensor has the form

$$\mathbf{K}_i = \kappa_i \begin{pmatrix} 1 & 0 & s_1 \\ 0 & 1 & s_2 \\ s_1 & s_2 & s_1^2 + s_2^2 \end{pmatrix} \quad (6.4)$$

where s_1, s_2 are the isoneutral slopes in zonal and meridional direction of the horizontal slope vector $\mathbf{s} = -\mathbf{e}_h/e_3$ (Olbers et al. 2012). In general, isopycnal mixing can also be anisotropic, for instance, it is expected to be reduced in the cross-stream direction, then yielding a more complicated formulation of \mathbf{K}_i .

6.1.2 Skew Diffusion

The tensor \mathbf{K} describes the isopycnal and diapycnal diffusion components of the eddy parameterization. To understand the skew diffusive component, we now consider the TEM framework applied to buoyancy $b = -g/\rho_0\rho$, where ρ is density and ρ_0 is a background density as used in the Boussinesq approximation. Then the isopycnal component $\mathbf{K}_i \cdot \nabla \bar{b}$ vanishes and only the diagonal component κ_d of \mathbf{K} remains. Analogous to equation (6.2), decomposing the eddy buoyancy flux in components along and across mean buoyancy contours (see, e.g., Olbers et al. 2012) (at this point

³In the oceanic interior, isoneutral slopes are small, whereas in the well-mixed boundary layer of the ocean, this would not be a good approximation.

omitting any possible rotational parts) yields

$$\partial_t \bar{b} + (\bar{\mathbf{u}} + \nabla \times \mathbf{B}_b) \cdot \nabla \bar{b} = \bar{\mathbf{Q}}_b + \nabla \cdot \kappa_d \nabla \bar{b}, \quad (6.5)$$

with the eddy advection velocity $\nabla \times \mathbf{B}_b$. In the oceanic interior, $\bar{\mathbf{Q}}_b$ and the diffusivity κ_d are small.⁴ Then the residual velocity $(\bar{\mathbf{u}} + \nabla \times \mathbf{B}_b)$ is along mean isopycnals in steady state. Neglecting κ_d , and using the gauge condition $\mathbf{B}_b \cdot \nabla \bar{b} = 0$,⁵ the vector streamfunction can be expressed with the eddy buoyancy fluxes, and in small-slope approximation

$$\mathbf{B}_b = -|\nabla \bar{b}|^{-2} \begin{pmatrix} \overline{v'b'\partial_z \bar{b}} - \overline{w'b'\partial_y \bar{b}} \\ \overline{w'b'\partial_x \bar{b}} - \overline{u'b'\partial_z \bar{b}} \\ \overline{u'b'\partial_y \bar{b}} - \overline{v'b'\partial_x \bar{b}} \end{pmatrix} \approx -|\nabla \bar{b}|^{-2} \begin{pmatrix} \overline{v'b'\partial_z \bar{b}} \\ -\overline{u'b'\partial_z \bar{b}} \\ \overline{u'b'\partial_y \bar{b}} - \overline{v'b'\partial_x \bar{b}} \end{pmatrix}. \quad (6.6)$$

The parameterization for the horizontal eddy buoyancy flux as a diffusion down the mean horizontal buoyancy gradient $\overline{u'_h b'} = -\kappa_b \nabla_h \bar{b}$ leads to the parameterization by Gent and McWilliams (1990), GM in the following, with the eddy advection velocity $u'_h = -\partial_z(\kappa_b \mathbf{s})$ and $w' = \nabla_h \cdot (\kappa_b \mathbf{s})$. This form of the eddy-induced advection velocity \mathbf{u}^* was chosen because it mimics the effects of baroclinic instability and leads to a global sink of available potential energy. The horizontal component of the GM skew flux of ρ , $\mathbf{F}(\rho)_{skew} = \mathbf{B}_b \times \nabla \bar{\rho}$, is downgradient, while the vertical component is always directed up the vertical density gradient. For the total skew flux $\mathbf{F}_{skew}(\rho) \cdot \nabla \bar{\rho} = 0$, meaning there is no net diapycnal GM skew flux of ρ (hence the adiabatic nature of GM). The upgradient tendency of the vertical skew flux results in a tendency for the GM closure to reduce potential energy locally, i.e., to rotate the density surfaces leading to a flattening of the isopycnals.

We note that with the GM parameterization, it is assumed that all tracers are advected by the same eddy-induced velocity $\nabla \times \mathbf{B}_b$. The advective and skew fluxes however depend on the tracer. As was shown, e.g., by Griffies (1998), for an arbitrary tracer T , the divergence of the GM advective flux $F_{adv} = \mathbf{u}^* T$, which is equivalent to the divergence of the skew flux $F_{skew} = \mathbf{B}_b \times \nabla T$, can be attributed to an antisymmetric skew diffusive tensor \mathbf{A} with $\mathbf{u}_* = -\nabla \cdot \mathbf{A}$. Hence, the small-slope isoneutral diffusion tensor \mathbf{K}_i can be combined with the antisymmetric GM stirring tensor \mathbf{A} to yield

$$\mathbf{J} = \mathbf{A} + \mathbf{K}_i = \begin{pmatrix} \kappa_i & 0 & (\kappa_i - \kappa_b)S_1 \\ 0 & \kappa_i & (\kappa_i - \kappa_b)S_2 \\ (\kappa_i + \kappa_b)S_1 & (\kappa_i + \kappa_b)S_2 & \kappa_i S^2 \end{pmatrix} \quad (6.7)$$

⁴By exploiting the gauge freedom of rotational eddy flux addition, the diapycnal diffusivity κ_d can be defined such that $\kappa_d = 0$ follows from zero diabatic forcing $\bar{\mathbf{Q}}_b$ as in Eden et al. (2007a).

⁵The component of \mathbf{B} in equation 6.3 parallel to $\nabla \bar{T}$ plays no role and without loss of generality we can use the gauge condition $\mathbf{B}_b \cdot \nabla \bar{b} = 0$. The solution for \mathbf{B} can be found taking $\mathbf{F}_b \times \nabla \bar{b} = -\nabla \bar{b} \times (\mathbf{B}_b \times \nabla \bar{b}) = -\mathbf{B}_b (\nabla \bar{b} \cdot \nabla \bar{b}) + \nabla \bar{b} (\mathbf{B}_b \cdot \nabla \bar{b})$ and the eddy streamfunction becomes $\mathbf{B}_b = -(\mathbf{F}_b \times \nabla \bar{b}) |\nabla \bar{b}|^{-2}$ (where we have neglected κ_d and a possible rotational flux in the eddy flux decomposition).

which is currently implemented in many ocean models to represent the isopycnal effects of eddies. It requires the specification of the isopycnal and skew diffusivities, κ_i and κ_b .

6.1.3 *Diagnosing and Parameterizing the Diffusivities*

The production of eddy kinetic energy in the ocean through baroclinic instability is spatially and temporally highly variable, and this variation needs to be captured by the diffusivities. Hence, significant horizontal and vertical variation in eddy diffusivities have been indicated by both observations and eddying model studies (e.g., Eden 2006; Marshall et al. 2006; Griesel et al. 2010; Naveira Garabato et al. 2011; Vollmer and Eden 2013; Griesel et al. 2014; Klocker and Abernathey 2014), although different methods to infer diffusivities can result in conflicting magnitudes and spatial distributions. It is also largely unclear how and where the eddy kinetic energy is dissipated. In most current eddy parameterizations, the dissipation of the available potential energy by the GM parameterization is simply lost and not considered further. One process of eddy energy dissipation is for example considered in Chapter 2 (von Storch et al. 2018).

Closures for both, the skew and isopycnal eddy diffusivities have been based on mixing length arguments, motivated by Green (1970), where the diffusivity, κ , depends on a length scale, L_e , and a typical velocity scale, v_e , $\kappa = L_e v_e$. Here, we also discuss approaches based on quasigeostrophic (QG) theory (e.g., Killworth 1997; Eden 2011) which offers a physically more elaborate alternative to the classical mixing length framework, from which potential vorticity (PV) and buoyancy diffusivities based on eddy fluxes resulting from the fastest growing unstable waves can be inferred. Quasigeostrophic linear stability analysis (LSA) results in PV diffusivities, κ_{PV} , that are reduced at the surface in strong jets such as the Antarctic Circumpolar Current (ACC) and enhanced at the steering level where the velocity of the background flow equals the phase speed of the unstable waves (e.g., Killworth 1997; Vollmer and Eden 2013). Steering levels are hence areas where enhanced dissipation of energy is expected, and classical mixing length theory can be extended to include this effect (Ferrari and Nikurashin 2010; Klocker and Abernathey 2014; Griesel et al. 2015).

On a more fundamental note, it still remains unclear whether, where, or on what scales the eddy diffusion model is appropriate at all. If eddy diffusivities are diagnosed from observations or models, correlations between the eddy fluxes and mean gradients are typically low and the distributions are unphysical and noisy, which can be attributed to the presence of strong rotational parts (equation 6.3) that play no role in the mean tracer equation but bias the definition of diffusivities. When diffusivities are diagnosed from Lagrangian particle dispersion, Griesel et al. (2010, 2014) showed that floats travel large distances before their mean square displacement can

be characterized as diffusion, in analogy to the molecular diffusion of Brownian particles, if at all. The eddy flux depends on the history of mean gradients and not just on the gradients at the point under consideration (Davis 1987). This non-local effect, and how it depends on the scale of interest, has never been systematically quantified on a global scale. The presence of large rotational parts of the eddy fluxes are connected to the non-locality of the eddy energy cycles, where the production and dissipation of eddy energy do not occur at the same location (e.g., Wilson and Williams 2006).

In the next section, we will give an overview over some Eulerian and Lagrangian methods to diagnose eddy diffusivities from both observations and models. Section 6.3 will summarize eddy diffusivity distributions found in the ocean with the different models. Section 6.4 comments on the applicability of the eddy diffusion model itself with a focus on Lagrangian dispersion statistics, and Section 6.5 introduces concepts for the parameterization of the eddy diffusivities taking into account energy consistency.

6.2 Eddy Diffusivity Diagnostics

Diffusivities can be quantified using both Eulerian and Lagrangian methods. The Lagrangian approach is based on the analysis of the spreading of a tracer or particles as they follow the flow (Taylor 1921), whereas Eulerian diffusivities can be quantified from Eulerian eddy tracer fluxes (e.g., Eden 2006; Fox-Kemper et al. 2013). Estimates of eddy diffusivities in the ocean have been made using all of these methods, with a focus on surface distributions (Ferreira et al. 2005; Marshall et al. 2006; Eden 2006; Sallée et al. 2008; Smith and Marshall 2009; Shuckburgh et al. 2009; Abernathey et al. 2010; Ferrari and Nikurashin 2010; Griesel et al. 2010; Naveira Garabato et al. 2011).

6.2.1 Lagrangian Particle Dispersion

The Lagrangian perspective can be used to describe the effects of eddy variability on a time mean transport (Davis 1991) and to evaluate linear QG theory as discussed in section 6.2.2 (Griesel et al. 2015). Einstein's theory of Brownian motion for molecules predicts the diffusive regime, where the mean square displacement of the particles grows linearly with time and diffusivity is constant. A ballistic regime where the dispersion grows quadratically with time is observed for shorter time lags. In turbulence, single-particle dispersion has been shown by Taylor (1921), under certain conditions, to be similar to Einstein's theory of Brownian motion. Following Taylor (1921), the Lagrangian single-particle diffusivity is half the time rate of change of the dispersion, which is the mean square particle displacement from a starting point. It can also be written as the integral of the Lagrangian velocity autocovariance function (Taylor 1921; Davis 1987; Griesel et al. 2010, 2014):

$$\kappa_L(\mathbf{x}, \tau) = \int_{-\tau}^0 d\tilde{\tau} \langle u'(t_0|\mathbf{x}, t_0) u'(t_0 + \tilde{\tau}|\mathbf{x}, t_0) \rangle_L, \quad (6.8)$$

where $u'(t_0 + \tau|\mathbf{x}, t_0)$ denotes the velocity of a particle at time $t_0 + \tau$ passing through x at time t_0 with the background mean flow at the particle's location subtracted. The brackets $\langle \rangle_L$ denote the Lagrangian average over many trajectories. Here, the theory of Taylor was extended to apply to oceanic flows with a background mean flow (Davis 1987), and hence the single-particle diffusivity as defined in equation 6.8 is a mixed Eulerian–Lagrangian quantity. If κ is converging to a constant for some time $t > T$, where T is a typical eddy timescale, then we can conclude that the behavior can be described as diffusion for times $t > T$. In the oceanic interior, if particles are advected along isopycnal surfaces, and if the dispersion is isotropic, the Lagrangian diffusivity is equivalent to the isopycnal diffusivity κ_i in the tensor \mathbf{K} from equation 6.1. As discussed below, a linearized form of the Lagrangian diffusivity is equivalent to the PV diffusivity from linear stability analysis.

The approach taken by Davis (1987), strictly speaking, is valid for an ensemble of particles passing through a fixed position at different moments in a shear flow. In practice, particles do not all pass through exactly the same point, and instead an ensemble of particles taken from a finite area (or bin) is considered. Generally, the Lagrangian diffusivity in equation (6.8) is a symmetric tensor for homogeneous statistics and negligible shear dispersion. As shown by Oh et al. (2000), if shear dispersion cannot be neglected, it is the along-stream diffusivity that is affected by shear dispersion. If we remove a spatially uniform average from each float velocity, we are left with a residual due to the mean shear that may dominate the diffusivity estimate (Bauer et al. 1998; Koszalka et al. 2011). Griesel et al. (2010, 2014) computed the Lagrangian diffusivity tensor in the Southern Ocean of an eddying ocean model in a coordinate system projected along and across the Eulerian mean velocity and found that the off-diagonal components were indistinguishable from zero, while the diffusivity in the along-stream direction was about 6 times larger than in the cross-stream direction. This means that in the tensor \mathbf{K}_i (equation 6.4) anisotropic diffusion should be implemented.

In the real ocean, it is unclear whether and where the diffusive regime exists. We will discuss this question in section 6.4. If the integral converges to a constant, κ_L^∞ , then the probability density function (PDF) of the particle displacements should be Gaussian (LaCasce 2008). An alternative method to assess convergence properties of equation (6.8) is hence to compare the observed PDFs with Gaussians using different diffusivities (LaCasce et al. 2014). In the presence of coherent vortices, like eddies, the velocity autocovariance does not just decay to zero but oscillates with time lag (Figure 6.1). The linearized Lagrangian velocity autocovariance, assuming advection of particles in the presence of small amplitude Rossby waves with wavenumber \mathbf{k} , can be written as (Ferrari and Nikurashin 2010; Klocker et al. 2012a)

$$R(\tau) = E_{ck} e^{-\gamma\tau} \cos(|\mathbf{k}|(\mathbf{n} \cdot \mathbf{U} - c_r)\tau), \quad (6.9)$$

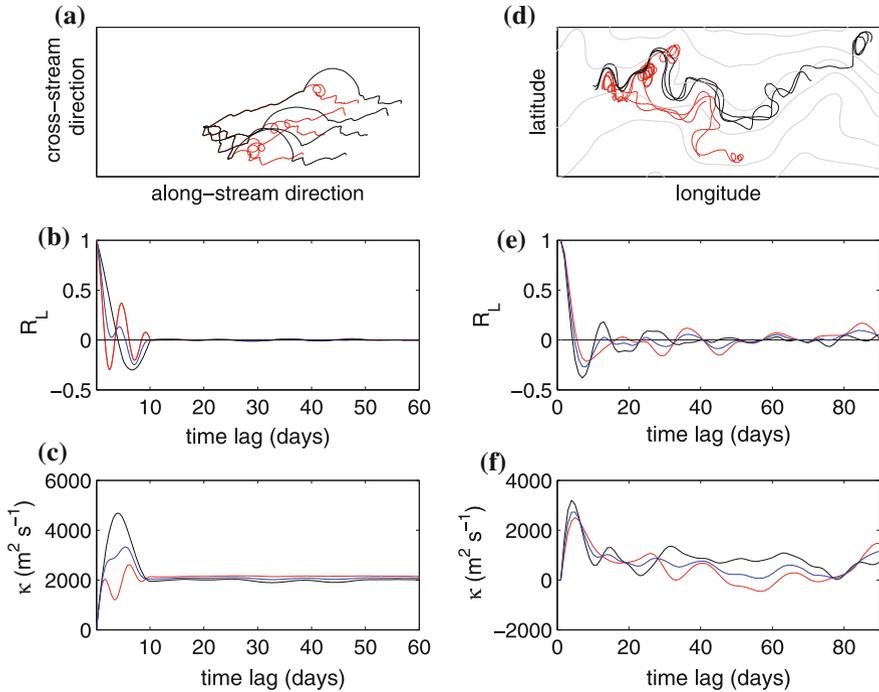


Fig. 6.1 Circling and meandering trajectories produce no net dispersion but oscillations in the velocity autocovariance, leading to the mixing barrier effect in the diffusivity. Left panels: idealized scenario: a random velocity field plus a constant mean flow in x -direction (“along-stream”) superimposed on a velocity component producing particle trajectories that meander and circle around eddies. **a** Resulting trajectories that meander (black) and circle around eddies (red). **b** Velocity autocovariance in meridional (“cross-stream”) direction, **c** corresponding cross-stream diffusivities as a function of time lag. Right panels: more realistic scenario from an eddying ocean model. **d** Some circling and meandering numerical trajectories at 300 m depth. **e** Corresponding velocity autocovariance **f** diffusivities as a function of time lag (from Griesel et al. 2014)

where E_{ek} is eddy kinetic energy, \mathbf{U} is the background mean flow, \mathbf{n} is the direction of wave/eddy propagation, c_r is the phase velocity of the waves/eddies, and γ is the Lagrangian decay scale or the linear damping term of the stochastic forcing as in Klocker et al. (2012a). This velocity autocovariance oscillates in the presence of large differences between mean flow and wave speeds (Figure 6.1). The integral, in the limit $\tau \rightarrow -\infty$, becomes

$$\kappa_{LAG} = \frac{E_{ek} \gamma}{\gamma^2 + |\mathbf{k}|^2 (\mathbf{n} \cdot \mathbf{U} - c_r)^2}. \quad (6.10)$$

This is equivalent to the PV diffusivity from linear stability analysis as will be shown in section 6.2.2, if γ , k , c_r are the growth rate, wavenumber, and phase velocity of the most unstable wave, respectively.

The Lagrangian single-particle diffusivity can be written as the product of the Lagrangian integral timescale, which is the integral of the normalized Lagrangian velocity autocorrelation and eddy kinetic energy. Near the surface, in the presence of strong jets, when the phase speed of the eddies and mean flow differs, the floats are carried through meanders and circle around eddies, leading to oscillations in the velocity autocovariance and large negative lobes (Figure 6.1) that reduce the Lagrangian integral timescale, consistent with the idea of a reduced mixing length (Klocker et al. 2012a).

Single-particle Lagrangian statistics are an ideal tool to investigate the relative importance of eddy kinetic energy and mixing barrier/steering level effects on the diffusivity, as well as quantifying the non-localness of the mixing and applicability of the diffusive model.

6.2.2 Quasigeostrophic Linear Stability Analysis

We now discuss the eddy diffusivities in the framework of quasigeostrophic linear theory, where one can define a buoyancy diffusivity, equivalent to the skew diffusivity, and a potential vorticity (PV) diffusivity that is equivalent to the linearized Lagrangian diffusivity. Starting point for the linear stability analysis is the linearized quasigeostrophic potential vorticity equation around a basic state $\Psi = \bar{\Psi} + \Psi'$, with

$$\partial_t q' + \mathbf{U}_h \cdot \nabla q' + \mathbf{u}'_h \cdot \nabla Q = A_h \nabla^2 q' \quad (6.11)$$

$$q' = \nabla^2 \Psi' + \partial_z \left(\frac{f^2}{N^2} \partial_z \Psi' \right) \quad (6.12)$$

$$= \nabla^2 \Psi' + \Gamma \Psi' \quad (6.13)$$

with the operator $\Gamma = \partial_z \left(\frac{f^2}{N^2} \partial_z \right)$, N and U_h are background stratification and horizontal velocity, respectively, q' and \mathbf{u}'_h are the perturbation PV and velocity, respectively, Ψ is the quasigeostrophic streamfunction. The lateral viscosity A_h is related to subgrid-scale friction and is introduced to filter fast growing small-scale modes that are often related to dynamically less important surface instabilities. Assuming horizontally homogeneous conditions, the mean PV gradient becomes

$$\nabla Q = \beta \mathbf{e}_y - \partial_z \left(\frac{f^2}{N^2} \partial_z \mathbf{U}_h \right) \quad (6.14)$$

where \mathbf{e}_y is the unit vector in the meridional direction. Inserting solutions of the form $\psi' = \Psi_0 \Phi(z) e^{i(k_x x + k_y y - \omega t)}$ with vertical structure function Φ and constant amplitude Ψ_0 leads to a vertical eigenvalue equation

$$\Gamma \Phi = \left(\frac{\overline{\mathbf{n}} \cdot \nabla Q}{c - \overline{\mathbf{n}} \cdot \mathbf{U}_h + i A_h k} + \mathbf{k}^2 \right) \Phi. \quad (6.15)$$

in the interior, with no normal flow boundary conditions, where $c = \omega/|\mathbf{k}|$ is the complex phase velocity, $\mathbf{n} = \mathbf{k}/|\mathbf{k}|$ is the direction of phase propagation. Vectors with the subscript $\overline{}$ are rotated anticlockwise in the horizontal, e.g., $\overline{\nabla} = (-\partial_y, \partial_x)$.

The eigenvalue problem (6.15) can be solved numerically following Smith (2007) and Vollmer and Eden (2013). Eigenfunctions Φ and eigenvalues ω might be complex. For a positive imaginary part of ω , the amplitude grows exponentially in time.

6.2.2.1 Eddy PV Flux and Diffusivity

With $\mathbf{u}'_h = \overline{\nabla \psi'} = \overline{\mathbf{k} \psi'}$ and $q' = \psi_0 (-\mathbf{k}^2 \Phi + \Gamma \Phi) e^{i(k_x + K_y - \omega t)}$, the eddy PV flux, averaged over one wave cycle

$$\overline{\mathbf{u}'_h q'} = -\frac{\psi_0^2}{2} \left(\frac{\overline{\mathbf{n}} \cdot \nabla Q}{c - \overline{\mathbf{n}} \cdot \mathbf{U}_h} \right) \Phi^2 c_i \overline{\mathbf{k}} \quad (6.16)$$

is in the direction perpendicular to the wave propagation and defines a PV diffusivity

$$\kappa_{PV} = \frac{\overline{\mathbf{u}'_h q'}}{|\nabla Q|^2} = \frac{\psi_0^2}{2} |\mathbf{k}| \frac{c_i |\Phi|^2}{|\mathbf{n} \cdot \mathbf{U}_h - c|^2} \quad (6.17)$$

where $c_i = \omega_i/|\mathbf{k}|$ is the imaginary part of the phase velocity. Equation (6.17) implies again the assumption that there is a local relationship between the eddy fluxes and the mean gradient, and that the unstable waves locally produce the mixing. The streamfunction amplitude can be parameterized using

$$\psi_0 = K_w c_i L_{max}, \quad (6.18)$$

with scale of largest growth L_{max} , and constant scaling factor K_w (Killworth 1997; Eden 2011; Vollmer and Eden 2013). With $E_{ek} = \frac{1}{2} \psi_0^2 |\mathbf{k}|^2 \Re(\Phi \Phi^*)$, the PV diffusivity can be written as

$$\kappa_{PV} = \frac{\omega_i E_{ek}(z)}{\omega_i^2 + |\mathbf{k}|(\mathbf{n} \cdot \mathbf{U}_h(\mathbf{z}) - c_r)^2} \quad (6.19)$$

where c_r is the real part of the (Doppler shifted) phase speed of the most unstable wave and $\omega_i = c_i |\mathbf{k}|$ is the growth rate. Note that the diffusivity has no singularity since for unstable waves $\omega_i \neq 0$ (and for stable waves $\kappa_{PV} = 0$).

6.2.2.2 Eddy Buoyancy Flux and Diffusivity

The eddy buoyancy flux related to the wave can be written as

$$\overline{\mathbf{u}'b'} = \overline{\mathbf{k} \times \nabla \Psi' f \partial_z \Psi'} = -\frac{\Psi_0^2}{2} \overline{\mathbf{k} f \Re \left(i \Phi \frac{d\Phi^*}{dz} \right)} = \kappa_b \nabla \bar{b}, \quad (6.20)$$

where Φ^* is the complex conjugate eigenfunction. The eddy buoyancy flux projected across mean buoyancy contours defines a diffusivity κ_b , that in the quasigeostrophic framework is equivalent to the GM skew diffusivity. Both PV and buoyancy diffusivity can be estimated by solving the linear stability problem using local profiles of vertical shear and stratification (Smith and Marshall 2009; Vollmer and Eden 2013; Griesel et al. 2015).

The eddy PV (equation 6.13) contains contributions from both the relative and stretching PV, and in the quasigeostrophic framework, the eddy PV and buoyancy fluxes are related in the Taylor relation

$$\overline{\mathbf{u}'q'} = \nabla \cdot \overline{\mathbf{u}'\mathbf{u}'} + f_0 \partial_z \left(\frac{\overline{\mathbf{u}'b'}}{N^2} \right) - \frac{1}{2} \frac{\nabla \bar{b}^2}{N^2} \quad (6.21)$$

The eddy buoyancy as well as the eddy momentum flux divergence contribute to the eddy PV flux. Hence, equation (6.21) suggests that the convergence of the eddy momentum flux that appears in the mean momentum budget can be parameterized in terms of eddy PV and buoyancy fluxes (see, e.g., Marshall et al. 2012 and Chapter 5; Danilov et al. 2019). If the buoyancy flux term dominates in equation (6.21), inserting the definition of the diffusivities, the buoyancy diffusivity is related to the vertical integral of the PV diffusivity

$$\kappa_{PV} \nabla Q = f_0 \partial_z \left(\frac{\kappa_b \nabla \bar{b}}{N^2} \right). \quad (6.22)$$

The advantage of using quasigeostrophic linear stability analysis is that it leads to diffusivities that are a direct consequence of the properties of the mean flow to generate eddies and does not rely on mixing length hypotheses. A disadvantage is that it assumes horizontally homogeneous conditions, relies on the linearized equations, and implies that the energy of the unstable waves is locally “dissipated.” Eddy energy that was generated elsewhere and advected into the region of interest is not taken into account.

A large fraction of the eddy field appears as coherent structures with large amplitudes and non-linear processes likely play a large role (Chelton et al. 2011). The effects of the non-local, non-linear nature of eddy–mean flow interaction on energy cascades, eddy properties, and eddy fluxes remain unclear (Chen et al. 2014a). Nevertheless Griesel et al. (2015) have shown how Lagrangian statistics for the fully

non-linear case can be used to assess the QG diffusivities, and that the Lagrangian decay scale is on average four times larger than the growth rate of the most unstable linear waves and depth dependent.

6.2.3 Diffusivities from Eulerian Eddy Fluxes

We now discuss how Eulerian eddy tracer fluxes from non-linear eddying ocean models can be used to define physically meaningful eddy diffusivities. From a Lagrangian perspective, “mixing” means decorrelation of particle trajectories and is generally a non-local process since particle trajectories need to be long enough to have reached the largest scales of the eddies to average over the oscillations in the velocity autocorrelation. Similarly, Eulerian eddy heat fluxes, averaged, must have a net component down the mean gradient so that mean available potential energy may be released. However, locally the association between eddy flux and mean gradient is not so strong because large rotational parts in the Eulerian eddy fluxes with both up- and downgradient components locally bias estimates of eddy diffusivities. To still estimate physically meaningful diffusivities, the idea is to isolate the local irreversible downgradient mixing processes from other reversible transport processes, taking into account the eddy variance budget (Osborn and Cox 1972; Marshall and Shutts 1981; Nakamura 2001; Medvedev and Greatbatch 2004; Eden et al. 2007a). This approach can be applied to both GM and PV diffusivity, using the eddy buoyancy variance (related to eddy potential energy) and eddy enstrophy (eddy PV variance) budgets. Generally, a large part of the eddy fluxes are not directed downgradient, as shown in Figure 6.2 for eddy buoyancy flux and mean buoyancy gradients in an eddying ocean model. Even if the part of the eddy flux that is across mean contours is considered, it is not necessarily downgradient.

In order to estimate physically meaningful diffusivities, analogous to the three-dimensional decomposition in equation (6.3) the horizontal eddy buoyancy flux can be decomposed into a component that is across the horizontal buoyancy gradient with buoyancy diffusivity κ_b as used in the GM parameterization, a component along isolines of mean buoyancy \bar{b} with parameter ν , and a rotational component with rotational potential θ that does not play a role in the divergence of the eddy fluxes but influences the definition of the buoyancy diffusivity

$$\overline{\mathbf{u}'_h b'} = -\kappa_b \nabla_h \bar{b} + \nu \overline{\nabla \bar{b}} + \overline{\nabla \theta}. \quad (6.23)$$

The buoyancy diffusivity can then be diagnosed from the Eulerian quantities in the models

$$\kappa_b = - \left(\overline{\mathbf{u}'_h b'} - \overline{\nabla \theta} \right) \cdot \nabla_h \bar{b} / |\nabla_h \bar{b}|^2, \quad (6.24)$$

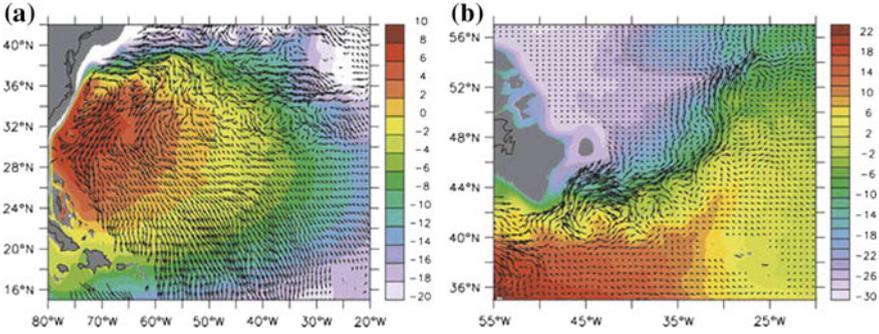


Fig. 6.2 **a** Horizontal eddy buoyancy flux (vectors) and mean buoyancy (color, in m min^{-2}) in an eddy-resolving model of the North Atlantic ($1/12^\circ$ resolution) at 300 m depth. Shown are 5-year averages for two different regions (from Eden et al. (2007b)). While some part (south of about 26°N) of the eddy flux is downgradient in the subtropical gyre (**a**), it is along mean buoyancy contours particularly in the North Atlantic current (**b**)

where rotational parts of the eddy fluxes can be defined by taking into account the equation for eddy variance $\Phi_2 = \frac{\overline{b'^2}}{2}$

$$\partial_t \Phi_2 + \nabla_h \cdot \overline{\mathbf{u}_h \Phi_2} + \overline{\mathbf{u}'_h b'} \cdot \nabla_h \bar{b} = -N^2 \overline{w' b'}. \quad (6.25)$$

which is equivalent to the conservation equation for available potential energy in QG. The second term on the left-hand side is the advection of eddy potential energy by both the mean flow and eddies, the third term is the eddy flux across \bar{b} contours and related to baroclinic instability, and the term on the right-hand side is the exchange between eddy potential and eddy kinetic energy and hence can be considered as “dissipation” of eddy potential energy. We have neglected any diabatic forcing. N^2 is the buoyancy frequency of the background stratification. Using the flux decomposition, equation (6.23) in steady state leads to

$$\nabla_h \cdot \overline{\mathbf{u}_h \Phi_2} + \overline{\nabla \theta \cdot \nabla_h \bar{b}} = \kappa_b |\nabla_h \bar{b}|^2 - N^2 \overline{w' b'}. \quad (6.26)$$

Similar to approaches of Marshall and Shutts (1981), Greatbatch (2011), Medvedev and Greatbatch (2004), Eden et al. (2007a), one can associate the advection of eddy variance (first term of the LHS of equation (6.26)) with the projection of the rotational flux along the buoyancy gradient (second term of the LHS of equation 6.26). This assumption allows then to associate the diffusivity κ_b with dissipation of eddy variance (RHS of equation (6.26)) and is hence related to the release of mean available potential energy as intended in the GM parameterization. However, also the eddy variance flux can have components along and across the mean buoyancy contours, and only the part that is along \bar{b} contours should be associated with the rotational potential, whereas the flux of eddy variance across \bar{b} contours should contribute to the horizontal diffusion of \bar{b} expressed with κ_b . Hence, we decompose also the eddy

variance flux,

$$\overline{\mathbf{u}_h \Phi_2} = -\kappa_2 \nabla \bar{b} + \nu_2 \overline{\nabla \bar{b}}, \quad (6.27)$$

inserted into the eddy variance equation leads to

$$\nabla_h \cdot \kappa_2 \nabla_h \bar{b} + \overline{\nabla} (\theta - \nu_2) \cdot \nabla_h \bar{b} = \kappa_b |\nabla_h \bar{b}|^2 - N^2 \overline{w' b'}. \quad (6.28)$$

Choosing the rotational potential $\theta = \nu_2 = \overline{\mathbf{u}_h \Phi_2} \cdot \overline{\nabla \bar{b}} / |\nabla_h \bar{b}|^2$, the diffusivity

$$\kappa_b = |\nabla_h \bar{b}|^{-2} (N^2 \overline{w' b'} + \nabla_h \cdot \kappa_2 \nabla \bar{b}) \quad (6.29)$$

is also associated with the eddy variance flux across \bar{b} contours, where $\kappa_2 = \overline{\mathbf{u}_h \Phi_2} \cdot \nabla \bar{b} / |\nabla_h \bar{b}|^2$. However, also the eddy variance flux in κ_2 contains rotational parts and we can now consider higher order moments and add a rotational flux $\overline{\theta_2}$ in the eddy variance flux decomposition equation (6.27) which can be determined considering the equation for $\Phi_3 = \overline{b'^3}/3$.

Even higher order moments can be taken into account analogous to Eden et al. (2007a), and the diffusivity, in steady state, can be expressed as

$$\kappa_b = \frac{N^2}{|\nabla_h \bar{b}|^2} \overline{w' b'} - \frac{N^2}{|\nabla_h \bar{b}|^2} \nabla_h \cdot \left(\frac{\nabla_h \bar{b}}{|\nabla_h \bar{b}|^2} \overline{w \Phi_2} \right) + O(\overline{w b'^n}; n > 2). \quad (6.30)$$

This means that in principle, eddy diffusivities associated with the production of eddy variance and higher order moments in steady state can be devised. The question remains whether in practice such a procedure will indeed converge to a local diffusivity.

6.3 Eddy Diffusivity Estimates in the Global Ocean

We discuss now results from different eddy PV, isopycnal and buoyancy (GM) diffusivity estimates, in both observations and eddying ocean models, and their consistency with predictions from linear stability analysis. In the linear QG framework, PV and Lagrangian diffusivity are equivalent, and PV diffusivity is expected to display a maximum at the steering level. Abernathy et al. (2013) showed principal consistency and agreement of PV, isopycnal diffusivity for different tracers and Lagrangian eddy diffusivity, using various diagnostics in a circumpolar channel without topography. They also found a steering level signature as predicted from linear QG theory consistent with the findings in the zonal channel of Treguer (1999). They also concluded that the buoyancy diffusivity is not identical to the isopycnal diffusivity. Riha and Eden (2011) compared Lagrangian and Eulerian PV diffusivities from a flux–gradient relationship in a zonal flat-bottom channel and found that the Eulerian estimate shows

an increase with depth and clear minima of meridional diffusivities within the zonal jets, indicating mixing barriers. In that study, the Lagrangian estimates agreed with the Eulerian method on the vertical variation.

In the zonally averaged flat-bottom framework, however, the downgradient relationship is expected to be more robust since, e.g., rotational fluxes are reduced. Zonal asymmetries, introduced in the form of topographic barriers, lead to different eddy statistics upstream and downstream of the barriers. For example, smaller-scale baroclinic eddies can be organized by larger-scale topography generated stationary waves to form localized storm tracks (locally enhanced eddy kinetic energy) (e.g., Abernathey and Cessi 2014; Bischoff and Thompson 2014; Chapman et al. 2015). The strong along-stream variation of eddy properties requires considering local time mean–eddy statistics and taking into account strong rotational parts (e.g., Wilson and Williams 2004). As a consequence, diffusivity estimates become challenging in more realistic configurations with realistic topography.

For example, Lagrangian methods need to consider long enough trajectories or tracer contours to estimate diffusivities, and those diffusivities correspond to an average over trajectories and tracer contours rather than providing local 2D diffusivity maps, a direct consequence of the non-localness of the eddy mixing (section 6.4). In spite of the challenges, and since observations of Eulerian eddy fluxes are too sparse and biased by rotational parts, a large part of observational estimates of eddy diffusivities rely on Lagrangian dispersion statistics in various regions (e.g., Davis 1991; LaCasce 2008; Koszalka et al. 2011; Sallée et al. 2008). Some studies did not use long enough time lags for the diffusivity to converge and thereby overestimate or underestimate diffusivity, leading to large correlation with eddy kinetic energy (Griesel et al. 2010). Some include the effect of shear dispersion which is no eddy effect. Estimates of surface eddy diffusivity were obtained by Marshall et al. (2006), Ferrari and Nikurashin (2010), Klocker and Abernathey (2014), Abernathey and Marshall (2013) advecting a tracer with velocities as computed from satellite altimetry. Effective diffusivity directly measures the enhancement of mixing due to the stretching of tracer contours by eddies, but is also a non-local estimate. Generally, surface eddy diffusivities are in the order $100\text{--}10,000\text{ m}^2\text{ s}^{-1}$ and agree in that they are reduced in the cross-stream direction in areas with strong mean flows.

While estimates of surface diffusivities are more widely available, inferences of depth dependence rely mostly on numerical models and theories. Vollmer and Eden (2013) analyze both GM and PV diffusivities from linear stability analysis using observed stratification and geostrophic velocity shear, showing enhanced PV diffusion at the steering level. Griesel et al. (2015) tested the predictions from linear theory in an eddying model in the Southern Ocean and found that the Lagrangian integral timescale was enhanced at the steering level at many places in the ACC (upper right panel of Figure 6.3), but not everywhere. However, the spatial variation of the Lagrangian diffusivity also crucially depends on the vertical structure of eddy kinetic energy (EKE) and was diagnosed in Griesel et al. (2015) to be without clear steering level signals (lower right panel of Figure 6.3).

Estimates of isopycnal diffusivity and depth dependence from observations are limited. Naveira Garabato et al. (2011) and Cole et al. (2015) used a mixing length

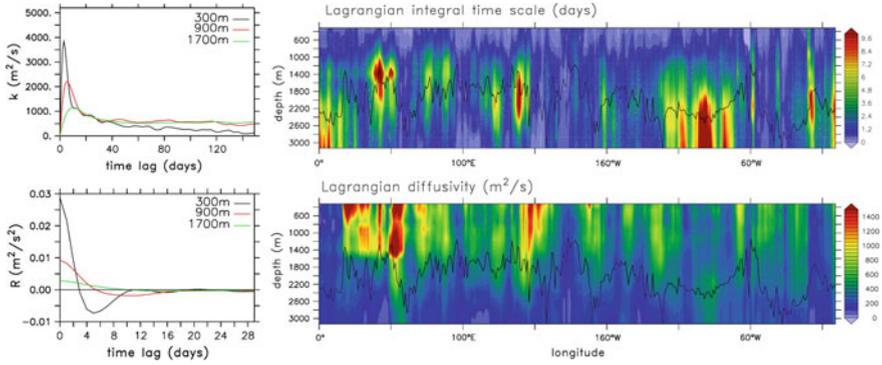


Fig. 6.3 Left: Lagrangian diffusivities (upper panel) and velocity autocovariance (lower panel) as a function of time lag for a region southwest of Australia. Right: Lagrangian integral timescale (upper panel) and Lagrangian diffusivity (lower panel) diagnosed from numerical particle dispersion in an eddying ocean model in the Southern Ocean as a function of depth and longitude along the main jets of the Antarctic Circumpolar Current. The black line is the steering level depth where enhanced mixing is expected from linear theory (from Griesel et al. 2015)

framework that relates observed tracer anomalies to gradients of the mean tracer to estimate isopycnal diffusivity. Cole et al. (2015) used salinity anomalies on density surfaces from ARGO float profiles, defining a mixing length (as described by $(S'^2)^{1/2}/|\nabla S|$, where S' , ∇S are salinity anomaly and salinity gradient on isopycnal surfaces, respectively) and diffusivity computed as mixing length times u'^2 . The study by Cole et al. (2015) supports the picture that while mixing lengths have both surface and subsurface maxima, diffusivity has less pronounced subsurface maxima due to the decay with depth of the velocity field. The significance, extent and magnitude, of a subsurface maximum and whether it is consistent with the depth at which mean flow and phase speeds of unstable linear Rossby waves are the same, remains under debate (Abernathey et al. 2010; Griesel et al. 2010, 2014; Klocker et al. 2012a, b; Tulloch et al. 2014; Chen et al. 2014b, 2015; LaCasce et al. 2014). Roach et al. (2016) recently used trajectories of ARGO floats together with an eddying ocean model of the Southern Ocean, SOSE and found cross-stream diffusivities at 1000 m ranged between 300 and 2500 $\text{m}^2 \text{s}^{-1}$, with peaks corresponding to topographic features. Eddy diffusivities in the regions of topographic steering were greater than what would be theoretically expected and the ACC experienced localized enhanced cross-stream mixing in these regions. Estimates of the GM diffusivity have been obtained considering eddy buoyancy fluxes in eddying ocean models. Nakamura and Chao (2000) and Tanaka et al. (2007) estimated the GM diffusivity from the divergence of the eddy fluxes directly, yielding extremely noisy distributions. Eden (2006); Eden et al. (2007b) estimate GM diffusivity from a high-resolution model of the Southern Ocean and North Atlantic, respectively, subtracting rotational parts using approaches similar to equation (6.24) and find that the subtraction of rotational parts improves the estimates.

The full diffusivity tensor (equation 6.7) has been estimated by Fox-Kemper et al. (2013) and Liu et al. (2012) from global eddying ocean models. Fox-Kemper et al. (2013) used multiple tracers, assuming they share the same diffusion tensor, to estimate its components with a least squares method (Bratseth 1998), without subtracting rotational parts. Their results suggested isopycnal and GM diffusivities to be similar, something that is not predicted by LSA (equation 6.22) for example. Liu et al. (2012) estimated both isopycnal and GM diffusivities with an adjoint technique, including the parameter ν from equation (6.23), finding GM diffusivities ranging from -800 to 2500 m, i.e., including negative ones. They also found some evidence of enhanced mixing at the steering level. Overall, the studies suggest significant spatial variations in eddy diffusivities with open questions regarding their depth dependence, whether GM and isopycnal diffusivities should be similar, and, more fundamentally over what scales and where a downgradient diffusive model is appropriate.

6.4 Limits of the Eddy Diffusion Model and Anomalous Diffusion

As noted before, the presence of rotational parts (or in other words, the reversible components in the eddy variance equation, like advection of eddy variance by the eddies and the mean flow) makes it difficult to locally parameterize the effects of eddies as a diffusive process, since production of EKE and “dissipation” does not occur in the same place. The subtraction of rotational parts including higher order moments as in Eden et al. (2007a) presents in theory the possibility to estimate a local eddy diffusivity that is related to dissipation of variance. But those rotational parts of the eddy fluxes may overwhelm the total eddy tracer flux making a local estimate challenging if possible at all (e.g., Griesel et al. 2009). Looking at the eddy variance budget of a passive tracer advected by geostrophic velocities derived from satellite altimetry, Abernathey and Marshall (2013) concluded that on large-scale mixing is “approximately local.” More explicitly, they argued, that variance is generally dissipated within 500 km, of where it is produced. The areas where the advection terms and thereby the rotational parts are significant are mostly in boundary currents, the ACC, or near the equator.

The lack of local relationship between eddy fluxes and mean gradients is also reflected in the fact that longtime lags are needed for the Lagrangian diffusivity to reach the diffusive regime (see also Figure 6.3). Specifically this might lead to experimental situations where the diffusive regime is not reached at all during the time of observation. For example, as discussed in section 6.2.1 in the presence of vortices, or eddies, the velocity autocovariance oscillates. A major part of the eddy flux rotates around contours of eddy variance (upper panel of Figure 6.4) and is both up and down the mean tracer gradient, producing rotational fluxes that do not influence mean tracer. The eddy heat fluxes may have a net downgradient component only when averaged (lower panel of Figure 6.4). Meanders and coherent eddies dominate the Lagrangian dispersion at small time lags, leading to no net dispersion (the mixing barrier effect as discussed in section 6.2.1, and cf Figure 6.1) and still well-correlated trajectories.

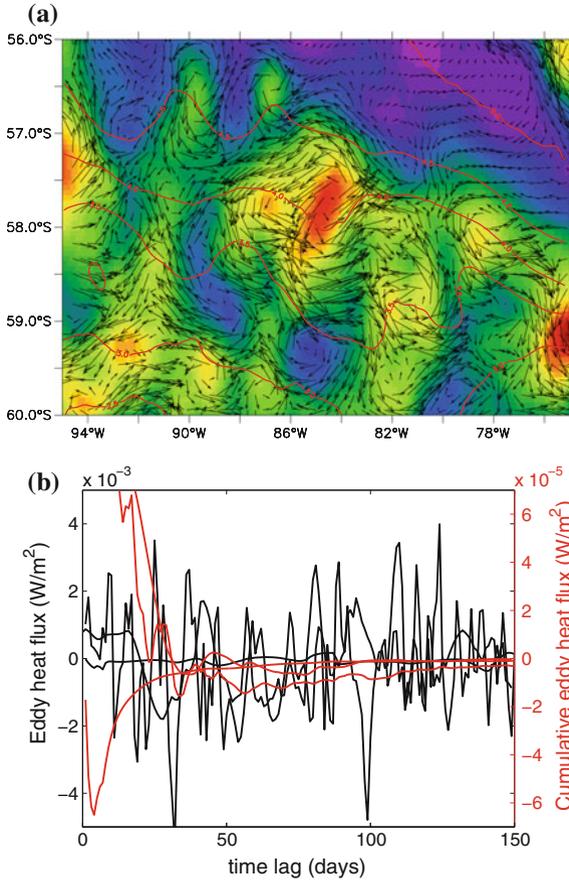


Fig. 6.4 Upper panel: horizontal eddy heat flux (vectors in W m^{-2}) rotating around eddy temperature variance $\overline{\theta'^2}$ (color in $^{\circ}\text{C}^2$), leading to eddy fluxes up and down the mean temperature $\overline{\theta}$ (red contours in $^{\circ}\text{C}$) gradient for a region in the Southern Ocean west of Drake Passage. Lower panel: cross-stream eddy heat flux along three Lagrangian trajectories (black, left axis; W m^{-2}) and associated cumulative mean cross-stream eddy heat flux averaged along each trajectory (red, right axis; W m^{-2}). The averaged heat flux converges for time lags greater than 50 days to a net downgradient (negative) one that is much smaller than the local eddy heat flux (black lines) (from Griesel et al. 2010)

Only when the particles become uncorrelated the diffusive regime can be reached, where the dispersion grows linearly with time as is characteristic of Brownian motion. For smaller scales, due to the underlying complex dynamics described above, strong deviations from normal diffusion occur. The transport can be ballistic, where the dispersion, that is the mean square displacement, $\langle r^2 \rangle$, grows quadratically in time. But even anomalous dispersion

$$\langle r^2(t) \rangle \sim t^\beta \tag{6.31}$$

with power laws different from linear ($\beta = 1$) or quadratic ($\beta = 2$) can take place. Superdiffusion where $\beta > 1$ has been observed in numerical and laboratory experiments, but also, e.g., for sea ice in the Fram Strait region (Gabrielski et al. 2015), where $\beta \approx 5/4$ has been measured. Anomalous diffusion, its hallmark being the non-linear growth of the mean square displacement with time (equation 6.31), is found in a wide diversity of complex systems, like liquid polymers, proteins, biopolymers, glasses or even in economics. In general, the temporal evolution of, and within, such systems deviates from Fick's diffusive law. With the development of more sophisticated experimental methods leading to higher resolutions in laboratory experiments or with the emerging higher resolutions in ocean and climate models, these deviations have become more prominent. From a mathematical point of view, anomalous diffusion rests on the generalized central limit theorem for such situations where underlying distributions are so broad that not all of their moments exist, which is the hallmark of so-called Lévy distributions. A Lévy flight, for instance, does not converge toward a Gaussian stochastic process, but instead is "attracted" toward a Lévy stable process with infinite variance (Mandelbrot 1982; Bouchaud and Georges 1990). The requirement of finite variance of a physical process can be met by introducing spatial-temporal coupling (Montroll and Weiss 1965; Shlesinger et al. 1993). While for superdiffusion ($\beta > 1$), long-range correlations dictate the temporal evolution of the system, broad distributions of residence times, e.g., due to geometrical constraints, can lead to subdiffusion ($\beta < 1$). As also discussed in section 6.2.1 subdiffusive dispersion in the ocean occurs, for example, due to longtime trapping of particles by coherent structures or eddies and planetary waves (e.g., Berloff and McWilliams 2002). Another diffusional anomaly, which has been explored much less, displays even logarithmically slow dependence on time of the so-called Sinai type $\langle r^2(t) \rangle \sim \ln^\beta t$ and can also be described within the same framework (Draeger and Klafter 2000).

In practice, estimating the Lagrangian diffusivities from ocean field experiments relies on grouping different trajectories together in geographical bins. As a consequence of the above-described features the bin sizes, then need to be large enough so that the trajectory segments are sufficiently long for convergence. Due to the complex dynamics on smaller and intermediate timescales, Lagrangian convergence properties and scales have not yet been quantified in a consistent way on a global scale. In some previous studies, the Lagrangian diffusivity has been evaluated as the maximum diffusivity in a given time lag interval (Krauss and Böning 1987; Lumpkin et al. 2001; Zhurbas and Oh 2003), which is not the right measure of diffusivity in the presence of meandering mean flows and coherent eddies (Veneziani et al. 2005; Griesel et al. 2010) and overestimates or underestimates the diffusivity. Fox-Kemper et al. (2013) and Zhurbas et al. (2014) discuss estimates of Lagrangian surface eddy diffusivities that were computed from available surface drifter trajectories. They obtained their values from comparatively short time lags (10, 15, or 20 days). Sallée et al. (2008) used surface drifters and fit an analytical expression of the autocorrelation, consisting of an exponential decay and a cosine oscillation with a period of the

first zero crossing. While the analytical expression is similar to the one derived by Klocker et al. (2012b) from dynamical arguments (equation (6.9)), it underestimates the mixing barrier effect by emphasizing the early lag times and hence consistently overestimates diffusivity (Griesel et al. 2015). Instead, Sallée et al. (2011) used time lags of up to 300 days in a cross-stream coordinate system and detected two regimes: a transition regime with periods less than approximately 50 days where dispersion grows quadratically (ballistic transport) and a second regime where the dispersion converges toward linear growth, starting around day 70. Griesel et al. (2010, 2014, 2015) have shown that in the Southern Ocean of an eddying model time lags of at least 30 days were required for the Lagrangian diffusivity to converge (Figure 6.1). LaCasce et al. (2014) reported that diffusivities from observed trajectories in the region west of Drake Passage failed to converge before the particles had left the region at depth levels above 1000 m. At the deeper levels, convergence took a whole year.

While single-particle dispersion yields information about the mean flow and diffusivity, multiple particle statistics measure the cross-correlation between velocities of different particles and thereby enables to quantify straining, which reflects the spread about the center of a cloud of tracers with time. In particular, Lagrangian pair dispersion statistics can shed light on the physics at different scales and can give insight into energy cascades that are present at the scale of the separation (e.g., LaCasce 2008). The study of pair (“relative”) dispersion goes back to Richardson’s famous work on atmospheric diffusion in 1926 (Richardson 1926) and has been derived from surface drifters in the ocean in recent years as more Lagrangian data become available (e.g., LaCasce 2008; Koszalka et al. 2009; LaCasce and Ohlmann 2003; Poje et al. 2014). Relative dispersion, computed from particle pairs, is quantified by the mean square separation d of the pairs

$$\langle d^2(t) \rangle = \langle |\mathbf{x}_1(t) - \mathbf{x}_2(t)|^2 \rangle \quad (6.32)$$

where the averages are taken over all pairs of particles 1 and 2 with $\mathbf{x}_{1,2}(t)$ being their location at time t . Similarly to single-particle diffusivity (equation 6.8), one can define the relative diffusivity

$$\kappa_r(t) = \frac{1}{2} \frac{d}{dt} \langle d(t)^2 \rangle = \langle d_0 v \rangle + \int_{t_0}^t \langle v(t)v(t') \rangle dt' \quad (6.33)$$

where v is the pairs’ separation velocity, and d_0 is their initial separation. For small enough initial separations the contribution of the first term can be neglected and the time dependence of the relative diffusivity is governed by the second term. When the particle separation is large enough, relative dispersion has the same asymptotic diffusive behavior as absolute (single) particle dispersion but the relative diffusivity approaches twice the absolute diffusivity from equation (6.8). At earlier times, due to velocity correlations induced by the Eulerian flow, relative dispersion shows a different behavior (see below). Under conditions of stationary Lagrangian statistics and homogeneous incompressible flow, relative dispersion can be related to the Eulerian

energy spectrum. If the spectrum has a power-law dependence $E(k) = k^{-\alpha}$, the dispersion scales as $\langle d^2(t) \rangle \sim t^{4/(3-\alpha)}$ for shallow spectra $1 < \alpha < 3$. For the turbulent inertial range, for which $\alpha = 5/3$, the relative dispersion increases cubically in time

$$\langle d^2(t) \rangle = C\varepsilon t^3, \quad (6.34)$$

where ε is the energy transfer rate and C is the Obukhov–Richardson constant. While equation (6.34) originally was derived by Richardson from empirical findings, later Obukhov and Batchelor derived the t^3 -dispersion law by applying scaling arguments of Kolmogorov. Richardson dispersion can be interpreted by a scale-dependent diffusivity $\kappa \propto d^{4/3}$ (see, e.g., Bennett 1984; LaCasce 2008). This anomalous dispersion which also referred to as local dispersion is controlled by eddies that are comparable in size to the pair’s separation.

For steeper Eulerian spectra, $\alpha \geq 3$ particle pair separations grow exponentially $\langle d^2(t) \rangle \sim \exp(t/\tau_s)$, where the unfolding time τ_s is related to the strain rate. In this non-local regime, the dispersion is determined by eddies which are larger than the pair’s distance. The diffusivity then scales linearly with the dispersion. In 2D turbulence in the presence of baroclinic instability, one would expect to observe exponential growth of relative dispersion until the deformation radius is reached, and then dispersion would increase cubically in time, until the scale of the largest eddies is reached from which on the dispersion would increase linearly in time with constant diffusivity. In the asymptotic (diffusive) regime, the pair’s separation is larger than the largest eddy in the system. Roughly speaking in the diffusive regime, the two particles are no more affected by the same eddy or more precisely this regime is characterized by uncorrelated pair velocities. Similarly to Brownian motion, where the particles are kicked independently by different molecules, pair dispersion in the ocean asymptotically grows linearly with time, i.e., it can be described by a constant diffusivity.

Observations of relative dispersion in the ocean from floats or drifters that were not deployed in pairs as it is done in most field experiments are limited since in these cases the analysis has to rely on so-called chance pairs, that is on pairs which randomly come close to each other. When drifters or floats are released in pairs (or triplets) Richardson-type dispersion is often found within certain separation scales, which depends on the region (Okubo 1971; LaCasce and Bower 2000; LaCasce and Ohlmann 2003; Ollitraut et al. 2005; LaCasce 2008; Koszalka et al. 2009; Draeger-Dietel et al. 2018). The diffusive regime in contrast is observed for much larger scales beyond the local deformation radius. Koszalka et al. (2009), for example, found evidence for the diffusive regime for scales larger than 100 km in the Nordic Seas from relative dispersion. For the Benguela upwelling region in the South Atlantic Draeger-Dietel et al. (2018) found the diffusive regime to take place at scales larger than about 200 km. Ohlmann and Niiler (2005) observed relative dispersion with initial pair separations smaller than 1 km in the Gulf of Mexico, where the deformation radius is roughly 45 km. The dispersion never reached the diffusive regime and a power-law dependence persisted up to the largest observed scales (a few hundred kilometers).

While the Richardson regime is expected to occur under a turbulent energy cascade, it also can occur in the presence of a mean shear. Since, as noted above, in the first case the velocities of the pairs are correlated, this leads to an increasing mean square separation velocity. In the latter case however due to the absence of correlations, a constant mean square separation velocity should be detected. By means of analyzing the velocity correlation Koszalka et al. (2009) discounted shear dispersion to be responsible for the Richardson dispersion observed at intermediate scales in their experiment in the Nordic Seas and concluded that their results are consistent with an inverse turbulent energy cascade in the region.

To conclude, since particles cover large distances before the corresponding dispersion becomes diffusive, if at all, parameterizations of the effects of eddies that are based on the concept of a downgradient diffusion with a scale-independent diffusivity need to be aware of the scales and regions for which they are applicable. Especially with higher resolution of climate models, interpretation of the diffusivities becomes more challenging. Therefore, approaches which take into account the intermediate anomalous diffusive regimes and go beyond the concept of diffusive system possibly in terms of recent stochastic methods can be discussed; see Chapter 1 (Franzke et al. 2019).

6.5 Eddy Diffusivity Parameterization

To account for the spatial variability of the eddy diffusivities, closures of the diffusivities κ_b , κ_{PV} , κ_i have been based on mixing length arguments, first proposed by Green (1970) and Stone (1972). Mixing length is similar to the concept of free mean path in thermodynamics, where a parcel retains its property for a characteristic length scale until it mixes with its surroundings. Assuming the mixing length is related to the size of the eddies, parameterizations of the form

$$\kappa_b = c_1 E_{ek} T_e = c_2 E_{ek}^{1/2} L_e = c_3 L_e^2 \sigma. \quad (6.35)$$

have been employed, where E_{ek} is eddy kinetic energy, L_e , T_e are typical eddy length and timescales, respectively. The c_i are tuning parameters. Several different variants of equation (6.35) have been used. Visbeck et al. (1997) chose the inverse of the vertically averaged Eady growth rate σ as a timescale and the maximum of the local Rossby radius, the grid spacing and the width of the baroclinic zone of the region of interest as the eddy length scale L_e . By adjusting the parameter c_3 , using the Rossby radius as length scale yields either too small values of κ_b in high latitudes or too large ones in low latitudes so Eden and Greatbatch (2008) concluded to use instead of the minimum of Rossby radius and Rhines scale, which is consistent with the theory of Theiss (2004). To include depth dependence Danabasoglu and Marshall (2007) used a parameterization that depended on the local stability frequency, N ,

$\kappa_b = \kappa_0 N^2 / N_{ref}^2$. Marshall et al. (2012) derived a geometric interpretation of eddy–mean flow interactions and proposed a GM coefficient similar to equation (6.35), however using the total (kinetic plus potential) eddy energy. The parameterization of Visbeck et al. (1997) can be considered as a simple version of the parameterization proposed by Eden and Greatbatch (2008), that will be discussed below, if $(\sigma L_e)^{1/2}$ is an approximation of E_{ek} , and the local, instead of vertically averaged Eady growth rate is used.

However, as suggested by, e.g., Ferrari and Nikurashin (2010) and Klocker and Abernathey (2014), and as discussed in section 6.2, mixing length scales like eddy size in the absence of mean currents, but is reduced when eddies propagate with speeds smaller than the mean flow speed, leading to mixing length that is reduced at the surface increasing with depth. This mixing suppression is supported by observations. The starting point is based on quasigeostrophic LSA (equation 6.19). This expression is strictly valid for the PV diffusivity, but the GM diffusivity is related to the PV diffusivity under certain assumptions (equation 6.22). It is not entirely clear how the variables that appear in equation (6.19) and also in (6.10) are related to the non-linear eddies in the ocean. The kinetic energy of the growing linear waves could be replaced by the EKE of the turbulent eddies and parameterized as in the framework of Eden and Greatbatch (2008) (next section). The linear growth rate ω_i may not be applicable to the non-linear eddies in the ocean that grow and decay. The Lagrangian parameter γ from equation (6.10) is the Lagrangian decay scale of the velocity autocovariance and is related to the Lagrangian decorrelation time. The Lagrangian decay scale was diagnosed to be about 4 times the growth rate of linear waves by Griesel et al. (2015) in the Southern Ocean, and also depth dependent exhibiting a maximum close to the steering level. In the stochastic theory of Ferrari and Nikurashin (2010), γ is the decay rate of a random superposition of Rossby waves. The inverse damping rate can be considered as a lifetime of the eddies, or the time it takes for large-scale straining motions to distort the eddy (Salmon 1998; Bates et al 2014). Under the assumption that the wavenumber $k = |\mathbf{k}|$ of the unstable waves is related to the size of the eddies $L_e = 2\pi/k$ (which might be larger than k in an inverse cascade), and that the phase speed of the eddies is related to the phase speed of the linear Rossby waves, one can estimate diffusivities based on equations (6.19), (6.10) (Ferrari and Nikurashin 2010; Klocker and Abernathey 2014; Bates et al 2014; Chen et al. 2014b; Griesel et al. 2015). We discuss in the next section how to parameterize EKE from equation (6.19).

6.5.1 EKE Equation

To connect to the energy budget, in the approach of Eden and Greatbatch (2008) the terms in the equation for eddy kinetic energy, E_{ek} ,

$$D_t E_{ek} = \overline{S} + \overline{w'b'} - \nabla \cdot \mathbf{M} - \varepsilon \quad (6.36)$$

are parameterized in terms of the large-scale variables. $\bar{S} = -\overline{\mathbf{u}'\mathbf{u}'} \cdot \nabla_h \bar{\mathbf{u}} - \overline{\mathbf{u}'\mathbf{v}'} \cdot \nabla_h \bar{\mathbf{v}}$ describes the production of eddy kinetic energy related to eddy momentum fluxes and exchange with mean kinetic energy, the energy production term $\overline{w'b'}$ due to baroclinic instability, the flux divergence term $\nabla \cdot \mathbf{M}$ containing the advection of eddy kinetic energy by the eddies and correlations between pressure and velocity fluctuations, and ε is the dissipation of eddy kinetic energy.

The baroclinic production term (denoting the conversion of eddy potential to eddy kinetic energy) is parameterized with κ_b as

$$\overline{w'b'} = \kappa_b |\nabla_h \bar{b}|^2 / N^2. \quad (6.37)$$

From a global eddying ocean model von Storch et al. (2012) estimate a global eddy energy production of 0.83 TW through baroclinic instability. The global conversion from mean to eddy kinetic energy through barotropic instability, (\bar{S}), was estimated to be 0.1 TW. This term can also locally be negative if the eddies transfer energy to the mean flow, i.e., upgradient. Nevertheless, a simple downgradient momentum flux closure is usually applied in ocean models that dissipates mean kinetic energy; see Chapter 5 (Danilov et al. 2019). An alternative to parameterize the effect of eddy momentum fluxes is to consider the downgradient diffusion of PV, as mentioned in section 6.2.2. Spatial variations of κ_{PV} can be linked to the eddy momentum flux convergence and generation of zonal jets, that act as mixing barriers (e.g., Eden 2010).

Several mechanisms have been proposed for the dissipation of EKE (ε). Flows in quasigeostrophic balance feature an inverse kinetic energy cascade from smaller to larger scales (see, e.g., Vallis 2006), while flows on much smaller scales transfer energy in the opposite direction, i.e., from larger to smaller scales until dissipation occurs at the molecular scales (Kolmogorov 1941). Processes like, e.g., lee wave generation (Nikurashin and Ferrari 2011), loss of balance (Bell (1975) and Chapter 2 (von Storch et al. 2018)) or ageostrophic instabilities can cause a damping of the EKE. Dissipation by bottom friction has been estimated to be between 0.12 and 0.65 TW (Arbic et al. 2009)—however, this estimate strongly depends on the drag coefficient c_D . How much of the eddy energy is dissipated at the smallest scale as the result of a forward energy cascade seems to depend on the presence of ageostrophic components of the flow (e.g., Molemaker et al. 2010; Brüggemann and Eden 2014). For example, simulations of Capet et al. (2008a, b, c) of an idealized Eastern boundary current system show sub-mesoscale features and filaments that are out of geostrophic balance which are accompanied by a forward energy flux at spatial scales smaller than the first baroclinic Rossby radius. On the other hand, Molemaker et al. (2005) find that for Richardson numbers (i.e., the ratio of the vertical density stratification to the vertical shear of the horizontal velocity) $Ri = O(1)$ as found in the mixed layer, the dominating baroclinic instabilities are still mainly geostrophically balanced. Instabilities that arise for $Ri \gg 1$ are mainly in quasigeostrophic balance (e.g., Stone 1966). The strength of the direct route to dissipation can hence be related to the Richardson number—a low Richardson number flow likely contains ageostrophic components and will foster a downscale energy cascade at the expense

of the inverse cascade. The study by Brüggemann and Eden (2014) has suggested to use a dissipation with a coefficient r that is dependent on the local Richardson number Ri , $\varepsilon = r(Ri)E_{ek}$. The relative contribution of the dissipation mechanisms to the total is unclear.

The flux term $\nabla \cdot \mathbf{M}$ is important to connect regions where E_{ek} is produced with the regions where it is dissipated. It can be interpreted as wave radiation, e.g., radiation of Rossby waves for quasigeostrophic dynamics, or eddy advection. As a first step it can be parameterized in the horizontal by simple isotropic horizontal diffusion of E_{ek} with diffusivity K . It is also important to account for the vertical fluxes, since eddy energy is predominantly produced close to the surface but may be dissipated at the bottom.

6.6 Conclusions

From observational estimates and analyzes of eddying ocean models, is there a consensus on spatial variation of eddy mixing coefficients? It has to be clear what kind of eddy diffusivity is considered. One distinguishes between advective and diffusive effects of the eddies, which can be represented as symmetric and antisymmetric (skew diffusive) components of a mixing tensor, respectively. In general, different tracers with different mean tracer gradient orientations may be associated with different diffusivities. Eddy diffusion is also non-isotropic (e.g., in the presence of strong mean flows), and one has to distinguish along- and cross-stream diffusion. The diffusivities can be diagnosed by quantifying that part of the downgradient eddy flux that is associated with irreversible mixing, e.g., through defining appropriate rotational fluxes.

Also, different eddy advection velocities can be defined, depending on tracer and averaging method. The commonly used GM skew diffusivity parameterizes the eddy advective effects as horizontal diffusion of buoyancy. PV and GM diffusivities generally are not the same, consistent with QG theory and observations (Abernathey et al. 2013; Vollmer and Eden 2013), whereas similar magnitudes and spatial distributions for passive tracers and PV have been inferred. The Lagrangian diffusivity is equivalent to PV diffusivity in linear theory if the growth rate of the unstable waves is a multiple of the Lagrangian decay scale. In the following, we distinguish between diffusivity, which can be diffusion of PV, passive and active tracers, and skew diffusivity, e.g., the diffusivity employed in the GM parameterization.

1. Surface distributions of eddy diffusivities are suppressed in regions where the speed of the mean flow and eddies differ, such as the ACC jets, western boundary currents, and hence are better represented by theories based on quasigeostrophic dynamics that include steering level effects, than using mixing length based on Rossby radius or eddy size only. Linear QG theory can predict eddy diffusivities as long as the flow is in a QG regime, and non-linear effects are negligible.

2. Apart from steering level effects, there is strong control of topography over mixing. EKE is often enhanced in the wake of topographic features, where mixing barrier effect breaks down. It is hence as important to correctly predict the enhanced EKE and associated mixing hot spots, as it is to include steering level effects.
3. The depth dependence of eddy diffusivities is less well observed than surface distributions. Linear QG predicts eddy PV diffusivities to be enhanced at the steering level, but since EKE decays rapidly with depth, it is unclear how significant a mid-depth maximum is in the ocean. Also, the linear growth rate ω_i may not be applicable to the non-linear eddies in the ocean that grow and decay. Since EKE predicted by LSA does not take into account eddy propagation, methods to parameterize the diffusivity based on linear QG theory may be improved by including non-linear effects through larger and depth-dependent decay rates, correct eddy propagation speeds and with the predictive EKE model of Eden and Greatbatch (2008), including EKE transport and dissipation mechanisms.
4. Skew diffusivities have been estimated largely in eddying ocean models or from linear QG theory using observed geostrophic mean flow. In QG, the skew diffusivity is related to the vertical integral of the PV diffusivity and can, under certain assumptions, be predicted once the PV diffusivity has been parameterized. The skew diffusivity is central to the theory of Southern Ocean residual overturning circulation and setting the transport of the ACC. So far, ocean models which permit eddies have predicted less sensitivity of overturning circulations and ACC transport to increases in Southern Ocean wind stress (the so-called eddy compensation and saturation mechanisms) than models in which eddies are parameterized, even though the extent of the compensation and saturation depends on the timescale, wind stress magnitude and location, and the surface buoyancy forcing (see Munday et al. 2013; Farneti et al. 2015, for an overview). Flow-dependent parameterizations (e.g., Viebahn and Eden 2010; Hofmann and Morales 2011; Gent and Danabasoglu 2011; Mak et al. 2017) have decreased this sensitivity. On the other hand, Jochum and Eden (2015) show the meridional overturning circulation is insensitive to Southern Ocean winds despite using a coarse resolution model with parameterized eddies as long as there is enough diapycnal mixing in the Indo-Pacific. Understanding and parameterizing the response of EKE and subsequent dissipation to changes in wind energy input are crucial. It should also be noted that numerical diffusivities, associated with the discretization of advection, can play a large role affecting ocean dynamics (Chapter 8, Klingbeil et al. 2019) and need to be considered on top of the physical ones.
5. It is important to quantify the limits and validities of the eddy diffusion model. Lagrangian diffusivities converge to a constant often only after several eddy length scales if at all. This “non-locality” makes estimating the “true diffusivity” challenging in regions with spatially variable mean flows and eddy characteristics, if possible at all. Local diffusivities can in principle be estimated by the method of Eden et al. (2007a), by defining that part of the eddy flux that is associated with the irreversible mixing, taking into account higher order moments (if the estimate converges). The question is how to interpret the Lagrangian diffusivity if

convergence takes place only after the floats have traveled out of the region of interest. Pair dispersion statistics can help identify which dispersion regimes exist at what scale. The question then is when and how can scale-dependent diffusivities be implemented. An idea was recently presented by Cushman-Roisin (2008) and Kämpf and Cox (2016) based on linear increase of the width of the dispersion patches (instead of with the square root of time as in Fickian diffusion), which is closer to the Richardson regime observed for scales smaller than the dominant oceanic eddies. The diffusivity then specifically depends on the scale of separation of the particles and cannot be considered a constant. Those approaches are particularly interesting in models where eddies are partially resolved. It is crucial to quantify on a global basis what dispersion regimes can be found where and for what scales, how long it takes to reach a diffusive limit, if at all, and how the diffusivity then depends on the regimes found at earlier lag times.

References

- Abernathey, R., Cessi, P.: Topographic enhancement of eddy efficiency in baroclinic equilibration. *J. Phys. Oceanogr.* **44**, 2107–2126 (2014)
- Abernathey, R., Marshall, J.: Global surface eddy diffusivities derived from satellite altimetry. *J. Geophys. Res.* **118**, 901–916 (2013)
- Abernathey, R., Marshall, J., Mazloff, M., Shuckburgh, E.: Enhanced isopycnal mixing at steering levels in the Southern Ocean. *J. Phys. Oceanogr.* **40**, 170–184 (2010)
- Abernathey, R., Ferreira, D., Klocker, A.: Diagnostics of eddy mixing in a circumpolar channel. *Ocean Model.* **72**, 1–16 (2013)
- Andrews, D.G., Holton, J.R., Leovy, C.B.: *Middle Atmosphere Dynamics*. Academic Press (1987)
- Arbic, B.K., et al.: Estimates of bottom flows and bottom boundary layer dissipation of the oceanic general circulation from global high-resolution models. *J. Geophys. Res.* **114**, C02024 (2009)
- Bates, M., Tulloch, R., Marshall, J., Ferrari, R.: Rationalizing the spatial distribution of mesoscale eddy diffusivity in terms of mixing length theory. *J. Phys. Oceanogr.* **44**, 1523–1540 (2014)
- Bauer, S., Swenson, M.S., Griffa, A., Mariano, A.J., Owens, K.: Eddy-mean flow decomposition and eddy-diffusivity estimates in the tropical Pacific Ocean. 1. Methodology. *J. Geophys. Res.* **103**, 30855–30871 (1998)
- Bell, T.H.: Topographically generated internal waves in the open ocean. *J. Geophys. Res.* **80**, 320–327 (1975)
- Bennett, A.F.: Relative dispersion: local and nonlocal dynamics. *J. Atmos. Sci.* **41**, 1881–1886 (1984)
- Berloff, P.S., McWilliams, J.C.: Material transport in oceanic gyres. Part I: phenomenology. *J. Phys. Oceanogr.* **32**, 764–796 (2002)
- Bischoff, T., Thompson, A.: Configuration of a Southern Ocean storm track. *J. Phys. Oceanogr.* **44** (2014)
- Bouchaud, J.P., Georges, A.: Anomalous diffusion in disordered media: statistical mechanisms, models and physical applications. *Phys. Rep.* **195**, 127–129 (1990)
- Bratseth, A.M.: On the estimation of transport characteristics of atmospheric data sets. *Tellus* **50A**, 451–467 (1998)
- Brüggemann, N., Eden, C.: Evaluating different parameterizations for mixed layer eddy fluxes induced by baroclinic instability. *J. Phys. Oceanogr.* **44**, 2525–2546 (2014)

- Capet, X., McWilliams, J.C., Molemaker, M.J., Shchepetkin, A.F.: Mesoscale to submesoscale transition in the California current system. Part i: flow structure, eddy flux, and observational tests. *J. Phys. Oceanogr.* **38**, 29–43 (2008a)
- Capet, X., McWilliams, J.C., Molemaker, M.J., Shchepetkin, A.F.: Mesoscale to submesoscale transition in the California current system. Part ii: frontal processes. *J. Phys. Oceanogr.* **38**, 44–64 (2008b)
- Capet, X., McWilliams, J.C., Molemaker, M.J., Shchepetkin, A.F.: Mesoscale to submesoscale transition in the California current system. Part iii: energy balance and flux. *J. Phys. Oceanogr.* **38**, 2256–2269 (2008c)
- Chapman, C., Hogg, A., Kiss, A., Rintoul, S.: The dynamics of Southern Ocean storm tracks. *J. Phys. Oceanogr.* **45**, 884–903 (2015)
- Chelton, D.B., Schlax, M.G., Samelson, R.M.: Global observations of nonlinear mesoscale eddies. *Prog. Oceanogr.* **91**, 167–216 (2011)
- Chen, R., Flierl, G., Wunsch, C.: A description of local and nonlocal eddy-mean flow interaction from an eddy state estimate. *J. Phys. Oceanogr.* **44**, 2336–2352 (2014a)
- Chen, R., McClean, J.L., Gille, S.T., Griesel, A.: Isopycnal eddy diffusivities and critical layers in the Kuroshio extension from an eddy ocean circulation model. *J. Phys. Oceanogr.* **44**, 2191–2211 (2014b)
- Chen, R., Gille, S., McClean, J., Flierl, G., Griesel, A.: A multi-wavenumber theory for eddy diffusivities and its application to the southeast Pacific (DIMES) region. *J. Phys. Oceanogr.* (2015). <https://doi.org/10.1175/JPO-D-14-0229.1>
- Cole, S.T., Wortham, C., Kunze, E., Owens, W.B.: Eddy stirring and horizontal diffusivity from argo float observations: geographic and depth variability. *Geophys. Res. Lett.* **42**, 3989–3997 (2015). <https://doi.org/10.1002/2015GL063827>
- Cushman-Roisin, B.: Beyond eddy diffusivity: an alternative model for turbulent dispersion. *Env. Fluid. Mech.* **8**, 543–549 (2008)
- Danilov, S., Juricke, S., Kutsenko, A., Oliver, M.: Toward consistent subgrid momentum closures in ocean models (This volume, Chapter 5) (2019)
- Danabasoglu, G., Marshall, J.: Effects of vertical variations of thickness diffusivity in an ocean general circulation model. *Ocean Model.* **18**, 122–141 (2007)
- Davis, R.: Modelling eddy transport of passive tracers. *J. Mar. Res.* **45**, 635–666 (1987)
- Davis, R.E.: Observing the general circulation with floats. Deep sea research Part A. *Oceanogr. Res. Pap.* **38**, S531–S571 (1991). [https://doi.org/10.1016/S0198-0149\(12\)80023-9](https://doi.org/10.1016/S0198-0149(12)80023-9)
- Draeger, J., Klafter, J.: Strong anomaly in diffusion generated by an iterated map. *Phys. Rev. Lett.* **84**, 5998–6001 (2000)
- Draeger-Dietel, J., Jochumsen, K., Griesel, A., Badin, G.: Relative dispersion of surface drifters in the Benguela upwelling region. *J. Phys. Oceanogr.* **48** (10), 2325–2341 (2018). <https://doi.org/10.1175/JPO-D-18-0027.1>
- Eden, C.: Parameterizing meso-scale eddy momentum fluxes based on potential vorticity mixing and a gauge term. *Ocean Model.* **32**, 58–71 (2010)
- Eden, C.: Thickness diffusivity in the Southern Ocean. *Geophys. Res. Lett.* **33** (2006). <https://doi.org/10.1029/2006GL026157>.
- Eden, C.: A closure for meso-scale eddy fluxes based on linear instability theory. *Ocean Model.* **39**, 362–369 (2011)
- Eden, C., Greatbatch, R.J.: Towards a mesoscale eddy closures. *Ocean Model.* **20**, 223–239 (2008)
- Eden, C., Greatbatch, R.J., Olbers, D.: Interpreting eddy fluxes. *J. Phys. Oceanogr.* **37**, 1282–1296 (2007a)
- Eden, C., Greatbatch, R.J., Willebrand, J.: A diagnosis of thickness fluxes in an eddy-resolving model. *J. Phys. Oceanogr.* **37**, 727–742 (2007b)
- Farneti, R., et al.: An assessment of Antarctic Circumpolar Current and Southern Ocean meridional overturning circulation during 1958–2007 in a suite of interannual CORE-II simulations. *Ocean Model.* **93**, 84–120 (2015)

- Ferrari, R., Nikurashin, M.: Suppression of eddy mixing across jets in the Southern Ocean. *J. Phys. Oceanogr.* **40**, 1501–1519 (2010)
- Ferreira, D., Marshall, J., Heimbach, P.: Estimating eddy stresses by fitting dynamics to observations using a residual-mean ocean circulation model and its adjoint. *J. Phys. Oceanogr.* **35**, 1891–1910 (2005)
- Fox-Kemper, B., Lumpkin, R., Bryan, F.: Lateral transport in the ocean interior. In: *Ocean Circulation and Climate: A 21st century perspective*, vol. 103, pp. 185–209 (2013)
- Franzke, C.L.E., Oliver, M., Rademacher, J.D.M., Badin, G.: Multi-scale methods for geophysical flows (2019) (This volume, Chapter 1)
- Gabrielski, A., Badin, G., Kaleschke, L.: Anomalous dispersion of sea ice in the Fram Strait region. *J. Geophys. Res.* **120**, 1809–1824 (2015)
- Gent, P., Danabasoglu, G.: Response to increasing Southern Hemisphere winds in CCSM4. *J. Clim.* **24**, 4992–4998 (2011). <https://doi.org/10.1175/JCLI-D-10-05011.1>
- Gent, P.R., McWilliams, J.C.: Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.* **20**, 150–155 (1990)
- Greatbatch, R.J.: A framework for mesoscale eddy parameterization based on density-weighted averaging at fixed height. *J. Phys. Oceanogr.* **31**, 2797–2806 (2011)
- Green, J.S.: Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Quart. J. R. Meteorol. Soc.* **96**, 157–185 (1970)
- Griesel, A., Gille, S.T., Sprintall, J., McClean, J.L., Maltrud, M.E.: Assessing eddy heat flux and its parameterization: a wavenumber perspective from a $1/10^\circ$ ocean simulation. *Ocean Model.* **29**, 248–260 (2009). <https://doi.org/10.1016/j.ocemod.2009.05.004>
- Griesel, A., Gille, S.T., Sprintall, J., McClean, J.L., LaCasce, J.H., Maltrud, M.E.: Isopycnal diffusivities in the Antarctic Circumpolar Current inferred from Lagrangian floats in an eddying model. *J. Geophys. Res.* **115** (2010). <https://doi.org/10.1029/2009JC005821>
- Griesel, A., McClean, J.L., Gille, S.T., Sprintall, J., Eden, C.: Eulerian and Lagrangian isopycnal eddy diffusivities in the Southern Ocean of an eddying model. *J. Phys. Oceanogr.* **44**, 644–661 (2014)
- Griesel, A., Eden, C., Koopmann, N., Yulaeva, E.: Comparing isopycnal eddy diffusivities in the Southern Ocean with predictions from linear theory. *Ocean Model.* **94**, 33–45 (2015). <https://doi.org/10.1016/j.ocemod.2015.08.001>
- Griffes, S.M.: The Gent-McWilliams skew flux. *J. Phys. Oceanogr.* **28**, 831–841 (1998)
- Hofmann, M., Morales, M.: The response of Southern Ocean eddies to increased midlatitude westerlies: a non-eddy resolving model study. *Geophys. Res. Lett.* **38** (2011). <https://doi.org/10.1029/2010GL045972>
- Jochum, M., Eden, C.: The connection between Southern Ocean winds, the Atlantic Meridional Overturning circulation, and Indo-Pacific upwelling. *J. Clim.* **28**, 9250–9257 (2015)
- Kämpf, J., Cox, D.: Towards improved numerical schemes of turbulent lateral dispersion. *Ocean Modell.* **106**, 1–11 (2016)
- Killworth, P.D.: On the parameterization of eddy transfer. Part I. Theory. *J. Mar. Res.* **55**, 1171–1197 (1997)
- Klingbeil, K., Burchard, H., Danilov, S., Goetz, C., Iske, A.: Reducing spurious diapycnal mixing in ocean models (2019) (This volume, Chapter 8)
- Klocker, A., Abernathy, R.: Global patterns of mesoscale eddy properties and diffusivities. *J. Phys. Oceanogr.* **44**, 1030–1046 (2014)
- Klocker, A., Ferrari, R., LaCasce, J.H.: Estimating suppression of eddy mixing by mean flows. *J. Phys. Oceanogr.* **9**, 1566–1576 (2012a)
- Klocker, A., Ferrari, R., LaCasce, J.H., Merrifield, S.T.: Reconciling float-based and tracer-based estimates of eddy diffusivities. *J. Mar. Res.* **70**, 569–602 (2012b)
- Kolmogorov, A.N.: Dissipation of energy in the locally-isotropic turbulence. *Proc. Math. Phys. Sci.* **434**, 15–17 (1941)
- Koszalka, I.M., LaCasce, J.H., Orvik, K.A.: Relative dispersion in the Nordic Seas. *J. Mar. Res.* **67** (2009)

- Krauss, W., Böning, C.: Lagrangian properties of eddy fields in the northern North Atlantic as deduced from satellite-tracked buoys. *J. Mar. Res.* **45**, 259–291 (1987)
- LaCasce, J.H.: Lagrangian statistics from oceanic and atmospheric observations. In: Weiss, J.B., Provenzale, A. (eds.) *Transport and Mixing in Geophysical Flows*. Springer, Berlin (2008)
- LaCasce, J.H., Bower, A.: Relative dispersion in the subsurface North Atlantic. *J. Mar. Res.* **58**, 863–894 (2000)
- LaCasce, J.H., Ferrari, R., Marshall, J., Tulloch, R., Balwada, D., Speer, K.: Float-derived isopycnal diffusivities in the DIMES experiment. *J. Phys. Oceanogr.* **44**, 764–780 (2014)
- Koszalka, I., LaCasce, J.H., Andersson, M., Orvik, K.A., Mauritzen, C.: Surface circulation in the Nordic Seas from clustered drifters. *Deep Sea Res.* **58**, 468–485 (2011)
- LaCasce, J.H., Ohlmann, C.: Relative dispersion at the surface of the Gulf of Mexico. *J. Mar. Res.* **61**, 285–312 (2003)
- Liu, C., Köhl, A., Stammer, D.: Adjoint-based estimation of eddy-induced tracer mixing parameters in the global ocean. *J. Phys. Oceanogr.* **42**, 1186–1206 (2012)
- Lumpkin, R., Treguer, A.-M., Speer, K.: Lagrangian eddy scales in the Northern Atlantic Ocean. *J. Phys. Oceanogr.* **32**, 2425–2440 (2001)
- Mak, J., Marshall, D.P., Maddison, J.R., Bachmann, S.D.: Emergent eddy saturation from an energy constrained eddy parameterisation. *Ocean Model.* **112**, 125–138 (2017)
- Mandelbrot, B.: *The Fractal Geometry of Nature*. ISBN 0-7167-1186-9. W.H. Freeman & Co (1982)
- Marshall, J., Shutts, G.: A note on rotational and divergent eddy fluxes. *J. Phys. Oceanogr.* **11**, 1677–1680 (1981)
- Marshall, J., Shuckburgh, E., Jones, H., Hill, C.: Estimates and implications of surface eddy diffusivity in the Southern Ocean derived from tracer transport. *J. Phys. Oceanogr.* **36**, 1806–1821 (2006)
- Marshall, D.P., Maddison, J.R., Berloff, P.S.: A framework for parameterizing eddy potential vorticity fluxes. *J. Phys. Oceanogr.* **42**, 539–557 (2012)
- Medvedev, A.S., Greatbatch, R.J.: On advection and diffusion in the mesosphere and lower thermosphere: the role of rotational fluxes. *J. Geophys. Res.* **109**, D07104 (2004). <https://doi.org/10.1029/2003JD003931>
- Molemaker, M.J., McWilliams, J.C., Yavneh, I.: Baroclinic instability and loss of balance. *J. Phys. Oceanogr.* **35**, 1505–1517 (2005)
- Molemaker, M.J., McWilliams, J.C., Capet, X.: Balanced and unbalanced routes to dissipation in an equilibrated Eady flow. *J. Fluid Mech.* **654**, 35–63 (2010)
- Montroll, E.W., Weiss, G.H.: Random walks on lattices. *J. Math. Phys.* **6** (1965). <https://doi.org/10.1063/1.1704269>
- Munday, D.R., Johnson, H.L., Marshall, D.P.: Eddy saturation of equilibrated circumpolar currents. *J. Phys. Oceanogr.* **43**, 507–532 (2013). <https://doi.org/10.1175/JPO-D-12-095.1>
- Nakamura, M., Chao, Y.: On the eddy isopycnal thickness diffusivity of the Gent-McWilliams subgrid mixing parameterization. *J. Clim.* **13**, 502–510 (2000)
- Nakamura, N.: A new look at eddy diffusivity as a mixing diagnostic. *J. Atmos. Sci.* **58**, 3685–3702 (2001)
- Naveira Garabato, A.C., Ferrari, R., Polzin, K.L.: Eddy stirring in the Southern Ocean. *J. Geophys. Res.* **116**, C09019 (2011). <https://doi.org/10.1029/2010JC006818>
- Nikurashin, M., Ferrari, R.: Global energy conversion rate from geostrophic flows into internal lee waves in the deep ocean. *Geophys. Res. Lett.* **38**, L08610 (2011)
- Oh, S.I., Zhurbas, V., Park, W.S.: Estimating horizontal diffusivity in the East Sea (Sea of Japan) and the northwest Pacific from satellite-tracked drifter data. *J. Geophys. Res.* **105**, 6483–6492 (2000)
- Ohlmann, J.C., Niiler, P.: A two-dimensional response to a tropical storm on the Gulf of Mexico shelf. *Prog. Oceanogr.* **29** (2005)
- Okubo, A.: Oceanic diffusion diagrams. *Deep-Sea Res.* **18** (1971)
- Olbers, D., Willebrand, J., Eden, C.: *Ocean Dynamics*. Springer, Berlin (2012)

- Ollittraut, M., Gabillet, C., de Verdiere, A.C.: Open ocean regimes of relative dispersion. *J. Fluid Mech.* **533** (2005)
- Osborn, T.R., Cox, C.S.: Oceanic fine structure. *Geol. Astron. Fluid Dyn.* **3** (1972)
- Poje, A.C., et al.: The nature of surface dispersion near the Deepwater Horizon oil spill. *Proc. Nat. Acad. Sci. USA* **111**, 12693–12698 (2014)
- Richardson, L.F.: Atmospheric diffusion shown on a distance-neighbor graph. *Proc. R. Soc. A. Bd* **110**, 709–737 (1926)
- Riha, S., Eden, C.: Lagrangian and Eulerian lateral diffusivities in zonal jets. *Ocean Modell.* **39**, 114–124 (2011)
- Roach, C.J., Balwada, D., Speer, K.: Horizontal mixing in the Southern Ocean from Argo float trajectories. *J. Geophys. Res.* 5570–5586 (2016). <https://doi.org/10.1002/2015JC011440>
- Sallée, J.B., Speer, K., Morrow, R., Lumpkin, R.: An estimate of Lagrangian eddy statistics and diffusion in the mixed layer of the Southern Ocean. *J. Mar. Res.* **66**, 441–463 (2008)
- Sallée, J.B., Speer, K., Rintoul, S.R.: Mean-flow and topographic control on surface eddy-mixing in the Southern Ocean. *J. Mar. Res.* **69**, 753–777 (2011)
- Salmon, R.: *Lectures on Geophysical Fluid Dynamics*. Oxford University Press, Oxford (1998)
- Shlesinger, M.F., Zaslavsky, G.M., Klafter, J.: Strange kinetics. *Nature* (1993)
- Shuckburgh, E., Jones, H., Marshall, J., Hill, C.: Understanding the regional variability of eddy diffusivity in the Pacific sector of the Southern Ocean. *J. Phys. Oceanogr.* **39**, 2011–2023 (2009)
- Smith, K.: The geography of linear baroclinic instability in the Earth's Ocean. *J. Mar. Res.* **65**, 655–683 (2007)
- Smith, K., Marshall, J.: Evidence for deep eddy mixing in the Southern Ocean. *J. Phys. Oceanogr.* **39**, 50–69 (2009)
- Stone, P.H.: On non-geostrophic baroclinic instability. *J. Atmos. Sci.* **27**, 390–400 (1966)
- Stone, P.H.: A simplified radiative-dynamical model for the static stability of rotating atmospheres. *J. Atmos. Sci.* **29**, 405–418 (1972)
- Tanaka, Y., Hasumi, H., Endoh, M.: The distribution of the thickness diffusivity inferred from a high-resolution ocean model. In: Ohfuchi, W., Hamilton, K. (eds.) *High Resolution Numerical Modeling of the Atmosphere and Ocean*. Springer, Berlin (2007)
- Taylor, G.I.: Diffusion by continuous movements. *Proc. R. Soc. A* **64**, 476–490 (1921)
- Theiss, J.: Equatorward energy cascade, critical latitude and the predominance of cyclonic vortices in geostrophic turbulence. *J. Phys. Oceanogr.* **34**, 1663–1678 (2004). <https://doi.org/10.1002/2015JC011440>
- Treguier, A.M.: Evaluating eddy mixing coefficients from eddy-resolving ocean models: a case study. *J. Mar. Res.* **57**, 89–108 (1999)
- Tulloch, R., Ferrari, R., Jahn, O., Klocker, A., LaCasce, J.H., Ledwell, J.R., Marshall, J., Messias, M.-J., Speer, K., Watson, A.: Direct estimate of lateral eddy diffusivity upstream of Drake Passage. *J. Phys. Oceanogr.* **44**, 2593–2616 (2014)
- Vallis, G.K.: *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press, Cambridge, UK (2006)
- Veneziani, M., Griffa, A., Reynolds, A.M., Garraffo, Z.D., Chassignet, E.P.: Parameterizations of lagrangian spin statistics and particle dispersion in presence of coherent vortices. *J. Mar. Res.* **63**, 1057–1083 (2005)
- Viebahn, J., Eden, C.: Towards the impact of eddies on the response of the southern ocean to climate change. *Ocean Model.* **34** (2010). <https://doi.org/10.1016/j.ocemod.2010.05.005>
- Visbeck, M., Marshall, J., Haine, T.: Specification of eddy transfer coefficients in coarse-resolution ocean circulation models. *J. Phys. Oceanogr.* **27**, 381–402 (1997)
- Vollmer, L., Eden, C.: A global map of meso-scale eddy diffusivities based on linear stability analysis. *Ocean Modell.* **72**, 198–209 (2013)
- von Storch, J.-S., Eden, C., Fast, I., Haak, H., Hernandez-Deckers, D., Maier-Reimer, E., Marotzke, J., Stammer, D.: An estimate of the Lorenz energy cycle for the World Ocean based on the 1/10° STORM/NCEP simulation. *J. Phys. Oceanogr.* **42**, 2185–2205 (2012)

- von Storch, J.-S., Badin, G., Oliver, M.: The interior energy pathway: inertial gravity wave emission by oceanic flows, p. 2 (2019) (This volume, Chapter 2)
- Wilson, C., Williams, R.G.: Why are eddy fluxes of potential vorticity difficult to parameterize? *J. Phys. Oceanogr.* **34**, 142–155 (2004)
- Wilson, C., Williams, R.G.: When are eddy tracer fluxes directed downgradient? *J. Phys. Oceanogr.* **36**, 189–201 (2006)
- Zhurbas, V., Lyzhkov, D., Kuzmina, N.: Drifter-derived estimates of lateral eddy diffusivity in the World Ocean with emphasis on the Indian Ocean and problems of parameterization. *Deep Sea Res. I* **83**, 1–11 (2014)
- Zhurbas, V., Oh, S.I.: Drifter-derived maps of lateral diffusivity in the Pacific and Atlantic Oceans in relation to surface circulation patterns. *J. Geophys. Res.* **109**, C05015 (2003). <https://doi.org/10.1029/2003JC002241>

Chapter 7

Entropy Production in Turbulence Parameterizations



Almut Gassmann and Richard Blender

Abstract The atmosphere is a forced-dissipative system. It has to export more entropy than it imports in a steady state. Therefore, the entropy inflow and outflow have to be distinguished from the internal entropy production, which has to be positive in the mean on long timescales. This principle does not only hold for the whole atmosphere, but also for subsystems like individual grid boxes in a numerical model. However, the constraint of positive internal entropy production was not taken into account when developing contemporary subgrid-scale parameterization schemes for atmospheric models. Some of these schemes suit automatically into this framework; some do not. This article discusses the current understanding and scientific discussion of this topic and illustrates possible future development paths.

7.1 The Numerically Modeled Atmosphere as a Forced-Dissipative System

The atmosphere is a forced-dissipative system. Its dynamics is forced either directly or indirectly by solar radiation. A part of this thermal forcing enters the atmosphere by the lower boundary through latent and sensible heat fluxes. Another part is the absorption of long-wave radiation inside the atmosphere itself. A pure radiative convective equilibrium in every atmospheric column is already in itself dissipative with regard to the radiation field but also with regard to the dissipation associated with heat, water vapor, and precipitation fluxes. Horizontal temperature differences may not be kept stable with respect to dynamic phenomena and lead to the generation of baroclinic eddies. The supplied energy to the dynamics of the atmosphere is opposed by kinetic energy dissipation which occurs when initial wave energy has passed through the macro-, meso-, and microscale energy cascade. Lorenz (1967) proves

A. Gassmann (✉)
Leibniz-Institut für Atmosphärenphysik e.V., Universität Rostock, Rostock, Germany
e-mail: gassmann@iap-kborn.de

R. Blender
Center for Earth System Research and Sustainability (CEN), Universität Hamburg,
Hamburg, Germany
e-mail: richard.blender@uni-hamburg.de

© Springer Nature Switzerland AG 2019
C. Eden and A. Iske (eds.), *Energy Transfers in Atmosphere and Ocean*,
Mathematics of Planet Earth 1, https://doi.org/10.1007/978-3-030-05704-6_7

that the net heating of the atmosphere equals the frictional heating. The exact pathway of energy dissipation is difficult to estimate as it occurs on molecular and viscous scales in nature. Those scales may not be explicitly simulated by atmospheric models which are used for climate projection or numerical weather prediction. The role of the diverse energy-dissipating processes due to friction, heat fluxes, constituent fluxes, and phase changes is not yet well known. Pauluis and Held (2002a, b) made large efforts to quantify all these processes in their relative contributions.

When atmospheric modelers build their models, they look at the dissipative processes appearing on the unresolved subgrid scale from a process-oriented viewpoint. Turbulence modelers use terms like countergradient fluxes when they refer to thermal fluxes of unresolved, but large, eddies. Cloud process modelers talk about entrainment and detrainment rates with respect to updrafts or downdrafts of convective cloud systems. Mathematicians tend to associate any diffusion in their equations with dissipative processes. All these concepts are very convenient but disguise the fact that the second law of thermodynamics contains a clear and unique definition of energy dissipation. Energy dissipation is simply defined as the temperature times the internal entropy production.

The way to account for energy dissipation is constrained by the second law of thermodynamics. It is often assumed that the second law of thermodynamics does not apply for the coarse-grained equations used in climate modeling or numerical weather prediction. This is however not true, because reversible phenomena which exchange energy between different energy compartments in a two-way fashion can only be imitated on the resolved scales, but not on unresolved scales. Only irreversible processes may occur on unresolved scales. The exclusive way to modulate the strength of dissipation is thus to tune the parameterization of diffusion coefficients of entropy-consistent algebraic flux formulations. It is known that models are often overdissipative. They tend to exhibit too steep kinetic energy spectra compared to nature indicating that too much energy is dissipated at too large scales. Therefore, one needs some fluctuations of turbulence in space and time which are able to obstruct the energy cascade or even lead to energy backscatter during short times. Such events occur with some finite probability according to the Fluctuation Theorem (FT) and can be described by negative diffusion coefficients. To explore this probabilistic nature is a future task for numerical modeling and faces several challenges. Numerical stability has to be guaranteed for occasionally negative diffusion coefficients. Numerical methods for stochastic differential equations have to be employed. And even before that, the statistical properties of fluxes have to be estimated in order to generate a suitable probability density function.

Most of the frictional and thermal dissipation occurs in the atmospheric boundary layer. But for the general comprehension of the full circulation of the atmosphere, the dissipation of Rossby waves and gravity waves in the middle atmosphere plays an important role, too. Here, apart from the turbulence parameterization schemes, gravity wave parameterization schemes imitate the dissipative processes. It is a challenge to prove compliance with the second law of thermodynamics for those highly sophisticated parameterization schemes. The same holds for the heat and moisture fluxes associated with deep convection schemes.

7.2 The Entropy Budget Equation in Numerical Models of the Atmosphere

We are interested in thermodynamic entropy and refer to the literature for the problem of the entropy of the radiation field. Yet the entropy budget equation for atmospheric flow is problematic with respect of the perception of turbulent heat fluxes. It has been noted by many authors (Goody 2000; Romps 2008; Raymond 2013; Gassmann and Herzog 2015) that if one is tempted to diffuse potential temperature instead of temperature, the diagnosed entropy production will be negative in case of stable stratification, just in contrast to the demands of the second law of thermodynamics. As discussed, for instance, by Verkley and Gerkema (2004), the diffusion of temperature would drive the atmosphere to an isothermal profile in equilibrium, which is not observed. Already Priestley and Swinbank (1947) argued in view of the stability of the dry-adiabatic lapse rate that eddy diffusion is from regions of high to regions of low potential temperature. They noted already that at first sight this is a contravention of the well-established second law of thermodynamics, but argued that the work done by wind shear essentially provides the energy to diffuse temperature against its own gradient. Similar arguments were recently repeated by Akmaev (2008). The work done by wind shear is the mechanical turbulent friction, which clearly leads to frictional heating. The momentum diffusion tensor and the heat flux are, however, unrelated thermodynamic fluxes, because they are of distinct tensor order (de Groot and Mazur 1984). This thermodynamic argumentation questions the common conception that shearing stresses provide the energy for thermal diffusion. In order to resolve this dilemma, Gassmann (2016) suggested to distinguish heat fluxes and buoyancy fluxes explicitly. This allows to diffuse potential temperature for any stratification, but needs a further subscale buoyancy term in the vertical velocity equation acting like a Rayleigh damping in case of stable stratification. The kinetic energy of the vertical motion lost by this buoyancy term provides the energy for pushing isentropes down.

A further problem in the numerical modeling of dissipation is the perception of air constituents like water vapor as passive tracers. Moist air is slightly lighter than dry air, and therefore, water vapor is slightly more likely mixed upward than a passive constituent. Such subtleties are not yet found in the practice of modeling.

It is also often speculated that numerical problems of the dynamical core create or destroy entropy arbitrarily (Egger 1999; Woolings and Thuburn 2006; Raymond 2013). In modern dynamical cores, this is no longer the case, because they often reside on Hamiltonian dynamics and imitate energy conversion correctly. Numerical cores that mimic the product rule of differentiation $\nabla \cdot \mathbf{F} = \psi \nabla \cdot (\mathbf{F}/\psi) + \mathbf{F}/\psi \cdot \nabla \psi$ for any scalar ψ and any flux \mathbf{F} will correctly imitate energy conversions and will also provide the correct discrimination of entropy flux divergences and internal entropy production terms.

Let us now become more specific and give a mathematical framework to the dynamics of moist air. Gassmann and Herzog (2015) gave a detailed derivation of the entropy budget equation. Their framework did not yet contain the mentioned

Table 7.1 Thermodynamic and dynamic quantities

Symbol	Explanation
$\bar{\psi}, \psi''$	Barycentric mean and deviation $\psi = \hat{\psi} + \psi''$, $\hat{\psi} = \overline{\rho\psi}/\bar{\rho}$
$\bar{\psi}, \psi'$	Reynolds mean and deviation $\psi = \bar{\psi} + \psi'$
$q_i = \rho_i/\rho, \sum_i q_i = 1$	Specific content of an air constituent $i = d, v, l, f$; dry air, water vapor, liquid, and frozen water
$\rho_i, \rho = \sum_i \rho_i$	Density of i , total density
$u = \sum_i u_i q_i$	Internal energy
$h = \sum_i h_i q_i$	Enthalpy
$s = \sum_i s_i q_i$	Entropy
$\mu = \sum_i \mu_i q_i, \mu_i = h_i - T s_i$	Chemical potential
$u_i = c_{vi}(T - T_0) + u_i^0$	Specific internal energy
$h_i = c_{pi}(T - T_0) + h_i^0$	Specific enthalpy
$c_{vi}, c_v = \sum_i c_{vi} q_i$	i -th/total heat capacity at constant volume
$c_{pi}, c_p = \sum_i c_{pi} q_i$	i -th/total heat capacity at constant pressure
$R_i = c_{pi} - c_{vi}, R = R_d q_d + R_v q_v$	i -th/total specific gas constant
$l_v = h_v - h_l$	Latent heat of vaporization
$l_s = h_v - h_f$	Latent heat of sublimation
T	Temperature
$p_i, p = p_d + p_v$	Partial pressure of i , total pressure
$\theta = T/\Pi, \theta_v = \theta R/R_d$	Potential temperature, virtual potential temperature
$\Pi = (p/p_0)^{R_d/c_{pd}}$	Exner pressure
$\mathbf{v}_i, \mathbf{v} = \sum_i \mathbf{v}_i \rho_i / \rho$	Velocity of i , barycentric velocity
$\mathbf{J}_i^d = \overline{\rho q_i (\mathbf{v}_i - \mathbf{v})} = \overline{\rho q_i \mathbf{v}_i^d}$	Diffusive flux, significant during precipitation, $\sum_i \mathbf{J}_i^d = 0$
$\mathbf{J}_i^t = \overline{\rho \mathbf{v}'' q_i''}$	Turbulent flux of i , $\sum_i \mathbf{J}_i^t = 0$
I_i	Source term of i -th constituent, $\sum_i I = 0$
$\mathbf{J}_s = c_p \overline{\rho \mathbf{v}'' T''}$	Sensible heat flux
$\mathbf{J}_\theta = \overline{\rho \mathbf{w}'' \theta''} \mathbf{k}$	Potential temperature (buoyancy) flux
$\theta_v^d = \mathbf{k} \cdot \mathbf{J}_\theta / (\bar{\rho} \hat{w})$	Diffusive addend to θ_v
$\underline{\boldsymbol{\tau}} = \overline{\rho \mathbf{v}'' \mathbf{v}''}$	Momentum diffusion tensor
$\varepsilon_{fr} = -\underline{\boldsymbol{\tau}} : \nabla \hat{\mathbf{v}} \geq 0$	Frictional heating
$K = \mathbf{v}^2/2$	Specific kinetic energy
$\Phi = gz$	Geopotential
$\boldsymbol{\omega}_a = \nabla \times \mathbf{v} + 2\boldsymbol{\Omega}$	Absolute vorticity vector
$N^2 = g \partial_z \ln \hat{\theta}$	Brunt–Väisälä frequency
$N_{iso}^2 = g^2 / (c_{pd} T)$	Brunt–Väisälä frequency as for an isothermal atmosphere
\mathbf{R}	Radiation flux
g	Gravity acceleration

distinction between buoyancy fluxes and heat fluxes. Here we give a formulation which comprises all processes. Table 7.1 defines the involved variables. A complete set of prognostic equations for atmospheric flow is found to be

$$\bar{\rho} \frac{d\hat{\mathbf{v}}}{dt} = -c_{pd} \bar{\rho} (\theta_v \mathbf{E} + \theta^d \mathbf{k}\mathbf{k}) \cdot \nabla \Pi - \bar{\rho} \nabla \Phi - \bar{\rho} 2\boldsymbol{\Omega} \times \mathbf{v} - \nabla \cdot \underline{\boldsymbol{\tau}} \quad (7.1)$$

$$\bar{\rho} \frac{d\hat{u}}{dt} = -\bar{\rho} \nabla \cdot \mathbf{v} + \varepsilon_{fr} - c_{pd} \bar{\Pi} \nabla \cdot \mathbf{J}_\theta - \nabla \cdot (\mathbf{R} + \mathbf{J}_s + \sum_i \hat{h}_i (\mathbf{J}_i^t + \mathbf{J}_i^d)) \quad (7.2)$$

$$\frac{d\bar{\rho}}{dt} = -\bar{\rho} \nabla \cdot \mathbf{v} \quad (7.3)$$

$$\bar{\rho} \frac{d\hat{q}_i}{dt} = -\nabla \cdot (\mathbf{J}_i^t + \mathbf{J}_i^d) + I_i. \quad (7.4)$$

The internal energy is usually not a prognostic variable. Rather temperature or potential temperature is predicted. In view of the derivation of the entropy budget equation, the internal energy equation is, however, useful. The Gibbs relation is a starting point for the derivation of the entropy budget equation

$$\bar{\rho} \frac{d\hat{u}}{dt} = -\bar{\rho} \nabla \cdot \mathbf{v} + \sum_i \hat{\mu}_i \bar{\rho} \frac{d\hat{q}_i}{dt} + \hat{T} \bar{\rho} \frac{d\hat{s}}{dt}. \quad (7.5)$$

Inserting the internal energy equation (7.2) and treating moisture quantities as in Gassmann and Herzog (2015) yield as the desired entropy budget equation

$$\begin{aligned} \bar{\rho} \frac{d\hat{s}}{dt} = & -\nabla \cdot \left(\frac{\mathbf{J}_s}{\hat{T}} + \frac{c_{pd} \mathbf{J}_\theta}{\hat{\theta}} + \sum_i \hat{s}_i (\mathbf{J}_i^t + \mathbf{J}_i^d) \right) + \frac{\varepsilon_{fr}}{\hat{T}} - \mathbf{J}_s \cdot \frac{\nabla \hat{T}}{\hat{T}^2} - \mathbf{J}_\theta \cdot \frac{c_{pd} \nabla \hat{\theta}}{\hat{\theta}^2} \\ & - \frac{\sum_i I_i \hat{\mu}_i}{\hat{T}} - \frac{\sum_i (\mathbf{J}_i^t + \mathbf{J}_i^d) \cdot \nabla \hat{\mu}_i |_{\hat{T}}}{\hat{T}}. \end{aligned} \quad (7.6)$$

This form discriminates between entropy flux divergences and internal entropy production terms σ related to distinct processes. The dissipative processes are in the order as they appear in the previous equation: friction, heat and buoyancy fluxes, irreversible phase changes, and mixing of air constituents. The associated single entropy productions have to be positive definite for long timescales. This requirement shapes the structure of the still unknown fluxes $\underline{\boldsymbol{\tau}}$, \mathbf{J}_s , \mathbf{J}_θ , I_i , \mathbf{J}_i^t , and \mathbf{J}_i^d . A hypothetical atmosphere has reached an equilibrium state if the single entropy productions vanish and the associated gradients are erased. This does not mean that all fluxes are vanishing instantaneously, but the probabilities of fluxes in the one or the other direction are the same.

7.3 Moisture and Precipitation Fluxes

As was shown in Gassmann and Herzog (2015), the turbulent fluxes of gaseous air constituents \mathbf{J}_i^t and the diffusive fluxes in case of precipitation \mathbf{J}_i^d are constrained by the requirement that the last term of (7.6) has to be positive definite. The gradients of chemical potentials $\hat{\mu}_i$ at constant temperature can be expressed as gradients of partial pressures $\bar{\rho}\hat{q}_{d,v}\nabla\hat{\mu}_{d,v}|_{\hat{T}} = \nabla\bar{p}_{d,v}$.

Then the vertical turbulent flux of water vapor becomes

$$\mathbf{J}_v^t \cdot \mathbf{k} = -\bar{\rho}K^v \left(\frac{\hat{q}_d\partial_z\bar{p}_v - \hat{q}_v\partial_z\bar{p}_d}{\bar{p}} \right). \quad (7.7)$$

This is different from the usual assumption that water vapor is mixed as a passive tracer

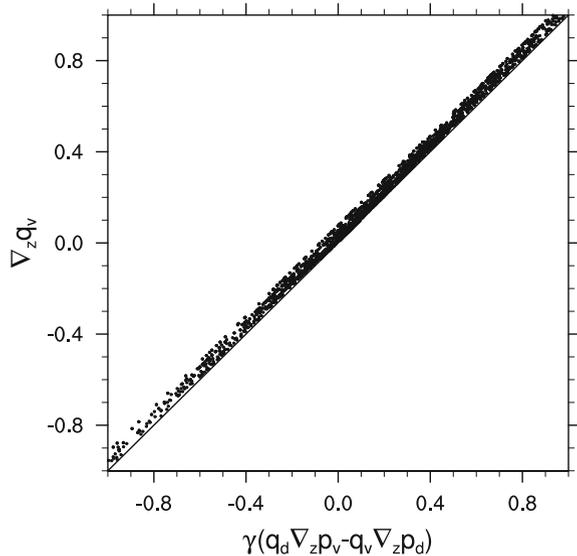
$$\mathbf{J}_v^t \cdot \mathbf{k} = -\bar{\rho}K^{v'}\partial_z\hat{q}_v. \quad (7.8)$$

The vertical gradient of \hat{q}_v can be alternatively expressed as

$$\partial_z\hat{q}_v = \frac{R}{R_v\hat{q}_d + R_d\hat{q}_v} \left(\frac{\hat{q}_d\partial_z\bar{p}_v - \hat{q}_v\partial_z\bar{p}_d}{\bar{p}} \right) + \frac{(R_d - R_v)\hat{q}_v\hat{q}_d}{R_v\hat{q}_d + R_d\hat{q}_v} \partial_z \ln\left(\frac{\bar{p}}{R}\right). \quad (7.9)$$

Figure 7.1 displays a scatter plot of the terms on the left and the immediate right of this equation for arbitrary radiosonde data in the tropics. The second term on the right makes for the deviation of the scatter points from the diagonal. The fact that all points are found to the left of the diagonal signifies that the vertical flux with the entropy-compliant formulation (7.7) is slightly more upward directed than the flux

Fig. 7.1 Scatter plot of the conventional and the entropy-compliant vertical gradients needed for water vapor fluxes. γ is an abbreviation for $R/(R_v\hat{q}_d + R_d\hat{q}_v)$. See the discussion of equation (7.9)



(7.8) under the passive tracer assumption. This is due to the different gas constants of water vapor and dry air. Water molecules have a smaller molecular weight than dry air considered as the mixture of N_2 and O_2 . The passive tracer perspective would presume an air constituent which is indistinguishable from dry air in its thermodynamic properties. This tracer could be diffused for its own but would not contribute to energy dissipation and thermodynamic entropy production. Regarding this example, the difference between the concepts of diffusion and dissipation becomes most obvious.

Pauluis and Held (2002a) fathered the dissipation associated with precipitation on friction occurring in the microscopic shear zones surrounding the hydrometeors. The inspection of the last term of (7.6) reveals, however, that this dissipation is formally due to mixing of air and precipitation. The governing equations consider air parcels moving with a barycentric velocity. Due to the sedimentation of the hydrometeors inside these air parcels, the non-sedimenting air constituents have an upward-directed relative velocity w^d with respect to the barycentric total velocity. This is expressed in the condition $\sum \mathbf{J}_i^d = 0$. As $\nabla \hat{\mu}_{l,f} |_{\hat{\gamma}} = 0$ for sedimenting particles, the energy dissipation is

$$\hat{T} \sigma |_{precip} = -w^d \partial_z \bar{p} \approx w^d \bar{\rho} g \geq 0. \quad (7.10)$$

The concept of the barycentric velocity with its relative velocities for sedimenting and non-sedimenting parts has to be correctly taken into account in numerical modeling as a prerequisite for the correct imitation of the discussed energy dissipation.

7.4 Thermal Fluxes

As discussed in Gassmann and Herzog (2015), the buoyancy flux term has to be discretized in a downgradient manner

$$\mathbf{J}_\theta = \overline{\rho w'' \theta''} \mathbf{k} = -\bar{\rho} K^\theta \partial_z \hat{\theta} \mathbf{k} \quad (7.11)$$

in order to obtain $-\mathbf{J}_\theta \cdot c_{pd} \nabla \hat{\theta} / \hat{\theta}^2 \geq 0$. Hence, when inserting an additional subgrid-scale buoyancy term $-c_{pd} \theta^d \partial_z \bar{\Pi}$ in the vertical velocity equation and diffusing the potential temperature downgradient, energy is conserved and a positive entropy source is guaranteed. Further inspection reveals that the buoyancy term acts like a Rayleigh friction term in the vertical velocity equation if the atmosphere is stably stratified

$$\partial_t \hat{w} |_{turb} = -c_{pd} \theta^d \partial_z \bar{\Pi} \approx g \frac{\theta^d}{\hat{\theta}} = -\frac{N^2 K^\theta}{\hat{w}} = -r_w \hat{w}. \quad (7.12)$$

As the Rayleigh friction term is not allowed to become singular, the associated Rayleigh damping coefficient $r_w = N^2 K^\theta / \hat{w}^2$ has to remain finite also for $\hat{w} \rightarrow 0$. This constrains in turn the diffusion coefficient K^θ which is then not connected with the momentum diffusion coefficient K^m via a fixed Prandtl number. The assumption

that vertical diffusion of horizontal momentum and Rayleigh damping of vertical momentum act on similar timescales for shortest resolvable waves suggests a flow-dependent Prandtl number

$$Pr = \frac{(\Delta z)^2 N^2}{\pi^2 \hat{w}^2}. \tag{7.13}$$

Here, the vertical grid distance parameter $\Delta z/\pi$ may also be associated with a vertical mixing length Λ . This Prandtl number leads to the buoyancy flux

$$\mathbf{J}_\theta \cdot \mathbf{k} = -\frac{\rho K^m \pi^2 \hat{w}^2 \hat{\theta}}{(\Delta z)^2 g}. \tag{7.14}$$

This flux is downward directed and vanishes for vanishing vertical velocity. The latter property is the main difference to the conventional potential temperature flux. When buoyancy fluxes are applied within a breaking gravity wave, only the flanks of the isentropes are pushed down, but not the wave trough, which leads to a flattening of the wave trough. This is schematically displayed in the upper example of Figure 7.2. As displayed in the lower example, a conventional scheme pushes the isentropes down in the wave trough, too. This increases the wave amplitude instead of supporting the wave overturning by just widening the wave trough.

The energy enabling the diffusion counter the temperature gradient stems from subscale buoyancy forces and not from shearing stresses. Kinetic energy proportional to $\bar{\rho} K^\theta N^2 = \bar{\rho} K^m \pi^2 \hat{w}^2 / (\Delta z)^2$ is converted into internal energy. Shearing stresses play only a role for the determination the momentum diffusion coefficient K^m .

In case of unstable stratification, the subscale buoyancy term (7.12) would result in a further amplification of the vertical velocity, hence a generation of kinetic energy out of unavailable internal energy. This is not intended by a dissipative process which should increase the unavailable internal energy. Therefore, the sensible heat flux \mathbf{J}_s instead of the buoyancy flux shall be used. Since the vertical gradients of temperature and potential temperature are parallel in case of unstable stratification, the specific algebraic form of the vertical heat flux is unimportant

$$\mathbf{J}_s \cdot \mathbf{k} = -\bar{\rho} K^\theta \bar{\Pi} \partial_z \hat{\theta} = -\bar{\rho} K^T \partial_z \hat{T}. \tag{7.15}$$

The diffusion coefficients K^T and K^θ are both positive and related by

$$K^T = \frac{-N^2}{N_{iso}^2 - N^2} K^\theta. \tag{7.16}$$

Fig. 7.2 Schematic of the pushing down of isentropes due to potential temperature fluxes. The upper example is for the entropy-compliant flux, and the lower panel for a conventional flux



The classical formulation of the sensible heat flux as downgradient with respect to θ and with fixed Prandtl number remains possible for unstable stratification.

The described combination of the heat flux \mathbf{J}_s for unstable stratification and the buoyancy flux \mathbf{J}_θ for stable stratification delivers an explanation for the stability of the adiabatic lapse rate in the atmosphere. With the new philosophy, the long-standing question of entropy production in case of stable stratification is solved.

It is of interest to investigate the practical implications of the proper discrimination of heat and buoyancy fluxes. Gassmann (2016) simulated breaking gravity waves in the mesosphere and found that the filamental structure of resulting mesospheric inversion layers was more realistically simulated if subscale buoyancy fluxes were used instead of the traditional diffusion of potential temperature at stable stratification as in (7.15). Figure 7.3 displays the specific energy deposition $K^\theta N^2$ as a measure for the downward pushing of the isentropes by subscale turbulence for a gravity wave breaking event in the mesosphere. As can be inferred from the figure, the energy deposition of the entropy-compliant model run vanishes in the wave trough where the vertical velocity is zero. This is not the case for the conventional run. Since the zero vertical wind line bends with height toward the horizontal, the temperature inversion layer at the bottom of the breaking region represented by the tight isentropes is sharper for the entropy-compliant run. This sharpness is also seen in typical Rayleigh lidar measurements by Liu and Meriwether (2004).

Turbulence parameterizations have to be applicable to the whole atmosphere. The successful simulation of breaking gravity waves is only a first application example of entropy-consistent heat and buoyancy fluxes. The simulation of atmospheric boundary layers, jets, and orographic wave breaking in the troposphere span future research fields.

Boundary layer schemes often employ a countergradient term inside the heat flux

$$\mathbf{J}_s \cdot \mathbf{k} = -\bar{\rho} K^\theta \bar{\Gamma} \left(\partial_z \hat{\theta} - \gamma \right). \quad (7.17)$$

This countergradient term imitates continuing upward heat flux even in case of slightly stable stratification. Therefore, it prevents the occurrence of negative entropy production $-\mathbf{J}_s \cdot \nabla \hat{T} / \hat{T}^2 < 0$ to some extent in the upper part of the planetary boundary layer. The countergradient term is said to mimic large eddies of the order of magnitude of the whole boundary layer depth. This heat flux is called the non-local heat flux.

According to the discussed distinction between heat fluxes for unstable stratification and buoyancy fluxes with Rayleigh damping for stable stratification, another perspective on countergradient schemes could be taken into account. Future research should investigate whether the countergradient flux is really needed or whether other means might replace it. One direction of research could be motivated by the observation that boundary layer eddies are very intermittent. Therefore, when simulating the respective heat fluxes in unstable stratification, the subscale fluxes do not have to be equal for equal mean ambient gradients. This indicates that heat fluxes should have some stochastic nature. The different intensities of the heat fluxes give rise to

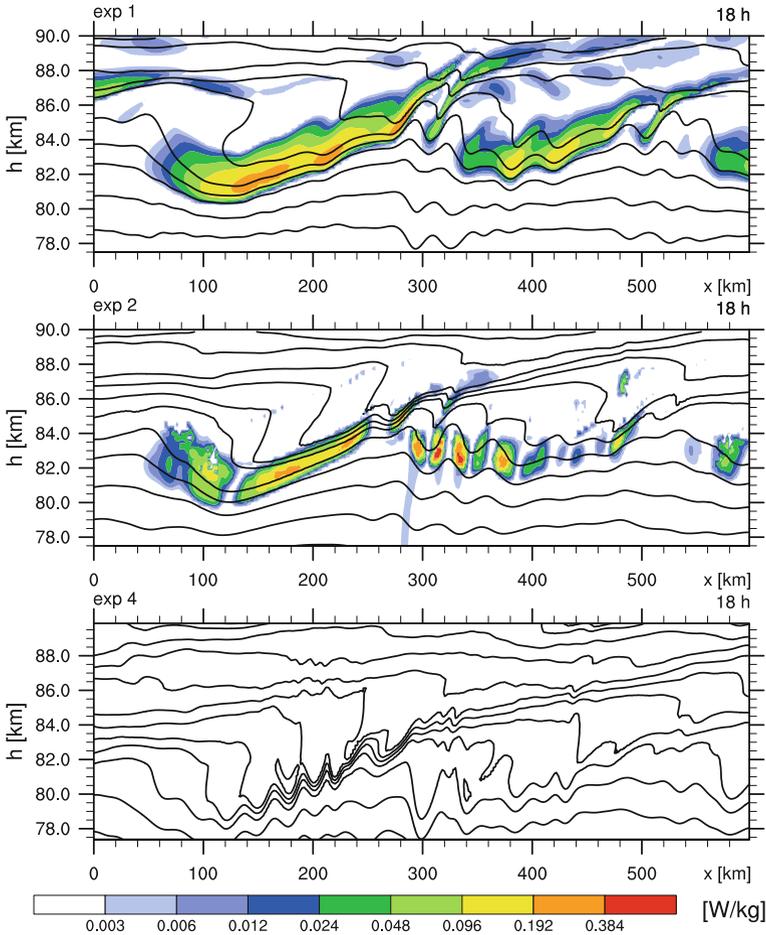


Fig. 7.3 $K^\theta N^2$ (colors) as a measure of the downward pushing of isentropes (isolines) due to subscale turbulence. The results are shown for a experiments with a conventional heat flux $\mathbf{J}_s \cdot \mathbf{k}$ as in (7.15) (upper panel), with an entropy-compliant buoyancy flux $\mathbf{J}_\theta \cdot \mathbf{k}$ (middle panel), and without any subscale heat or buoyancy flux (lower panel) in case of stable stratification

the bumping of subscale eddies onto the top of the boundary layer. The such generated horizontal temperature gradients lead to resolved dynamical motions including nonzero vertical velocities in the stably stratified top of the boundary layer. Such resolved vertical velocities are a prerequisite for the buoyancy fluxes which in turn entrain air into the boundary layer from above. This full cycle may deliver a realistic representation of the structure of the boundary layer.

7.5 Momentum Fluxes

It has been early recognized that atmospheric models need some amount of horizontal and vertical momentum diffusion and that this diffusion is not only needed in the atmospheric boundary layer. When running early general circulation models, models crashed because of nonlinear instabilities. Two pathways were explored to solve this problem. Arakawa (1966) proposed to enforce the conservation of potential enstrophy in addition to potential vorticity and energy by the introduction of the Arakawa Jacobian. Phillips (1959) suggested to introduce a momentum diffusion term just as a numerical means to keep the model stable. In retrospect, the Arakawa Jacobian was a success for the dynamics of vortical motions, but could not be applied when divergent motions were allowed to evolve unconstrained by an explicit balance equation as it had been the case during the time of quasi-geostrophic numerical models. With the emergence of the primitive equations for numerical modeling, enstrophy-conserving schemes lead to the danger of the Hollingsworth instability (Arakawa and Lamb 1981; Hollingsworth et al. 1983; Gassmann 2013). Consequently, the momentum diffusion is the surviving method of choice. However in early applications, the horizontal diffusion coefficient was only crudely chosen or set to a constant. This led to significant overdiffusion for second-order diffusion operators. Therefore, hyperviscosity operators were introduced which only damp the smallest resolved scales. At this time, compliance with the second law of thermodynamics was not a topic of interest. Hyperviscosity schemes were not tested on their conformance with the second law of thermodynamics. As the frictional heating was not considered at all as a source term in the internal energy equation, it was not even noticed that the eventual frictional heating of hyperviscosity schemes would sometimes lead to a frictional cooling, just in contrast to the second law of thermodynamics. Finally, any type of nonlinear momentum diffusion of Smagorinsky type (Smagorinsky 1993) is the remaining form of momentum diffusion and the challenge of research is to determine suitable and flexible formulations of the diffusion coefficients. For instance, the dynamic Smagorinsky parameterization delivers flow-dependent diffusion coefficients which lead to a reasonable development of the kinetic energy spectrum.

Still today, the state-of-the-art conception of momentum diffusion is problematic regarding a structural break occurring for the transition from anisotropic to isotropic turbulence. Intuitively, one would assume that in this limit case the horizontal and vertical diffusion coefficient should equalize and later converge to the molecular viscosity for viscous scales where the Navier–Stokes stress tensor

$$\boldsymbol{\tau}_{NS} = -\mu\rho \left(\nabla\mathbf{v} + \mathbf{v}\nabla - \frac{2}{3}(\nabla \cdot \mathbf{v})\mathbf{I} \right) \quad (7.18)$$

determines the diffusion of momentum. This tensor is symmetric and trace-free and guarantees positive entropy production. The problematic aspects of this definition are twofold. First, it is not clear where the distinction of horizontal and vertical diffusion coefficients can consistently be introduced. Second, this tensor is not invariant to a

rotation of the coordinate system in one plane. If one writes $\underline{\tau}_{NS}$ down in components, it yields

$$\underline{\tau}_{NS} = -\mu\rho \begin{pmatrix} \frac{2}{3}E + \frac{2}{3}I & F & G \\ F & -\frac{2}{3}E + \frac{2}{3}J & H \\ G & H & -\frac{2}{3}I - \frac{2}{3}J \end{pmatrix} \quad (7.19)$$

with the shear deformations

$$F = \partial_x v + \partial_y u \quad G = \partial_z u + \partial_x w \quad H = \partial_y w + \partial_z v, \quad (7.20)$$

and the strain deformations

$$E = \partial_x u - \partial_y v \quad I = \partial_x u - \partial_z w \quad J = \partial_y v - \partial_z w. \quad (7.21)$$

The strain deformations appear on the diagonal, and the shear deformations are the off-diagonal elements. Rotating, for instance, the horizontal coordinate plane by 45° would swap the meaning of the horizontal shear and strain deformations (see Figure 7.4). But unexpectedly, this purely mathematical change of the coordinate system leads to a change in the amount of deformation by the factor 2/3, which should not be the case from a physical point of view. Therefore, the traditional Navier–Stokes stress tensor is questionable. The additional requirement of invariance to rotations in one plane enforces the factors 2/3 to be discarded in (7.19). The resulting form allows also for a meaningful distinction between stresses in horizontal and vertical planes with the coefficients K_h^m and K_v^m for anisotropic turbulence and coarse-grained numerics

$$\underline{\tau} = -\bar{\rho} \begin{pmatrix} K_h^m \hat{E} + K_v^m \hat{I} & K_h^m \hat{F} & K_v^m \hat{G} \\ K_h^m \hat{F} & -K_h^m \hat{E} + K_v^m \hat{J} & K_v^m \hat{H} \\ K_v^m \hat{G} & K_v^m \hat{H} & -K_v^m \hat{I} - K_v^m \hat{J} \end{pmatrix}. \quad (7.22)$$

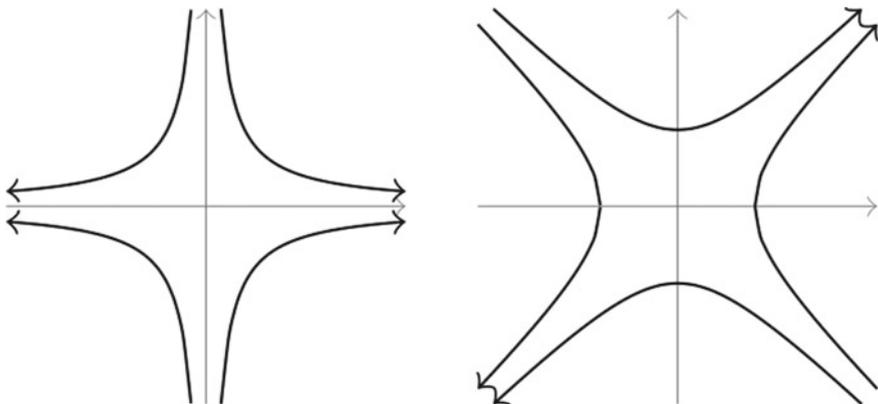


Fig. 7.4 Strain (left) and shear (right) deformations

The different diffusion coefficients are confined to the plane in which the deformations are measured. In numerical modeling of the atmosphere at resolutions typical for climate models, the horizontal diffusion coefficients are of $\mathcal{O}(10^5)$ m²/s and the vertical diffusion coefficients are of $\mathcal{O}(10)$ m²/s. From (7.22), it is also obvious that the incompressibility condition does not play any simplifying role in the formulation of the tensor, just unlike as in the traditional form (7.18). The form (7.22) coincides with the concepts described in Smagorinsky (1993) for horizontal diffusion but diverges for vertical diffusion.

The dynamic Smagorinsky model determines the strength of the diffusion coefficients in dependence on the modulus of the deformation stresses and a dynamically determined mixing length which depends on the resolution scale and the actual flow state. The careful determination of such diffusion coefficients is a current and future challenge for any fluid dynamical modeling. Physical constraints which step in at this place are scale independence and stochastic aspects associated with the Fluctuation Theorem.

As a last remark, it should be stressed that the discrete equivalent of the product rule $-\hat{\mathbf{v}} \cdot \nabla \cdot \underline{\boldsymbol{\tau}} - \underline{\boldsymbol{\tau}} \cdot \nabla \hat{\mathbf{v}} = -\nabla \cdot (\underline{\boldsymbol{\tau}} \cdot \hat{\mathbf{v}})$ has to be guaranteed. This is not a trivial endeavor if non-traditional meshes are used. One example is the hexagonal C-grid of the ICON-IAP model (Gassmann 2013). Recognizing that Gauss and Stokes theorems can be adapted for the evaluation of strain and shear deformations helps to define meaningful numerical schemes. In case of the hexagonal C-grid as it is used in the ICON-IAP model, the correct imitation of the mentioned product rule requires the shear deformation to be defined on a set of three kites just like the vorticity and the strain deformation to be defined on the hexagons just like the divergence (see Figure 7.5). For the ICON model with its triangular C-grid, the meanings are just swapped (Wan et al. 2013). Then, however, it becomes clear that the divergence has to be computed on a triple set of kites, too. This implies again that there are too many degrees of freedom in the mass field which cannot be slaved by intelligent numerics (Gassmann 2011). For either ICON mesh, the positivity of the frictional heating $-\underline{\boldsymbol{\tau}} \cdot \nabla \hat{\mathbf{v}} \geq 0$ is currently only proven for equilateral grids. This poses a problem

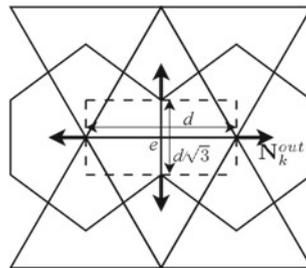


Fig. 7.5 The ICON-IAP model uses normal velocity components with respect to the hexagon edges. Shear deformations are computed for each double-triangle (kite) touching a central triangle. Strain deformations are computed for each hexagon. The divergence of the horizontal stress tensor is evaluated using the Gauss theorem for the dashed rectangle

especially in the vicinity of the 12 pentagons constituting the initial dodecahedron. As the ICON model on the triangular C-grid needs further numerically introduced divergence smoothing, the actual amount of friction cannot be attributed to the stress tensor alone. Therefore, for this model, a consistent energy or entropy budget equation may presumably not be formulated. The pathway of the energy flow remains thus to some extent obscure.

7.6 Fluctuation Theorem

Any heat, constituent or momentum flux has some probability density function. It might be quite broad and also contain nonzero small probabilities of negative entropy production. Then, the energy cascade is obstructed or even reversed. The latter means energy backscatter.

In small systems and on short timescales, the Fluctuation Theorem (FT) predicts a finite probability of negative entropy production (Gallavotti and Cohen 1995). The FT shows how such a violation of the second law becomes exponentially unlikely when the time period increases. The FT, which has been observed to hold in experiments and numerical simulations of out-of-equilibrium statistical mechanical systems, is a large deviation result that relates the ratio of the probabilities for positive to negative entropy production rates with the absolute value of entropy production and the growth time τ . For large τ , negative values become exponentially less likely and for infinitely long times τ the probability of negative entropy production vanishes.

The Fluctuation Theorem is not a single unique theorem, but a set of closely related theorems; for reviews, see Evans and Searles (2002) and Sevick et al. (2008). Common to all is that a FT quantifies the probability of observing second law violations in small systems observed for a short time. The ratio of probabilities for positive and negative values of time-averaged internal entropy production $\bar{\Sigma}_t$ satisfies in the long-term limit

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{P(\bar{\Sigma}_t = A)}{P(\bar{\Sigma}_t = -A)} = \bar{\Sigma}_t. \quad (7.23)$$

Prominent versions of the FT are: (i) Steady State Fluctuation Theorems (SSFT), which are valid in the long time limit (Evans et al. 1993), (ii) Transient Fluctuation Theorems (TFT), and (iii) the Crooks relation which involves work and free energy difference for forward and backward processes. Furthermore, local FTs are valid and FTs can be derived for dissipative systems under thermodynamic constraints (e.g., thermostatted conditions, Gallavotti and Lucarini (2014)). The FT has been derived by a maximum entropy condition (Dewar 2005).

Numerical models are not able to resolve all dynamical scales. Processes below the resolution of a numerical model are parameterized, i.e., they are retrieved from resolved variables. A grid box in a model represents a subsystem with unresolved degrees of freedom. In hydro-thermo-dynamic models, these fluctuations are coined as unresolved eddies. Here we assume that such a subsystem has a finite number

of degrees of freedom. In this case, the fluctuations are relevant. The properties of fluctuations are determined by the Fluctuation Theorem (FT) which is valid for a forced-dissipative system out of equilibrium. The FT is relevant for a system with a finite size or when the timescale is finite, i.e., below the thermodynamic limit. In a large system, fluctuations are negligible and the FT cannot be detected.

Gallavotti (1999) and Michel and Searles (2013) determined the modification of the FT for a small open subsystem embedded in a large system. Michel and Searles (2013) considered highly correlated systems; therefore, the results therein are applicable to turbulent hydro-thermo-dynamic systems. The authors find a correction term $1 + \alpha$ in the large system Fluctuation Theorem

$$\log \left(\frac{p(\bar{\Sigma}_{\ell,t} = A)}{p(\bar{\Sigma}_{\ell,t} = -A)} \right) = (1 + \alpha) \bar{\Sigma}_{\ell,t} \quad (7.24)$$

and coined the result the Local Fluctuation Theorem (LFT). The dissipation function $\bar{\Sigma}_{\ell,t}$ is measured over a time interval t and describes the irreversibility of the process. The index ℓ denotes the length scale of the subsystem (grid box). The probability $p(\bar{\Sigma}_{\ell,t} = A)$ means that $\bar{\Sigma}_{\ell,t}$ is in the range $(A - dA, A + dA)$, and the ratio is determined for small dA . Thus, the general form of the FT is still valid but with a correction term α . The correction is due to correlations of the subsystem with the surroundings and depends on the size and geometry of the subsystem. We assume that a model grid box is an open subsystem where this LFT is valid. Note that the probabilities on the left-hand side (7.24) can be written for any phase space function instead of the dissipation function $\bar{\Sigma}_{\ell,t}$. For the dissipation function, the standard form of the FT is obtained. The dissipation function can be determined for an external thermodynamic force \mathbf{F}_e and a dissipative flux \mathbf{J}_t

$$\bar{\Sigma}_{\ell,t} = 1/(k_B T) \mathbf{F}_e \cdot \mathbf{J}_t \quad (7.25)$$

where T is the temperature and k_B is the Boltzmann constant guaranteeing a non-dimensional shape of the dissipation function.

Michel and Searles (2013) determined the dependency of α on the size of the subsystem. For large size ℓ , $\alpha \sim \ell_0/\ell$, with the molecular mean free path ℓ_0 . In hydrodynamic applications, ℓ_0 is estimated by the size of the unresolved eddies. Hence, α should be inverse to the grid box size. For large grid boxes with many eddies, α is small and the FT is valid instead of the LFT. For higher resolutions, i.e., with smaller ℓ , fluctuations become more important. In general, the parameter α depends on the specific experimental conditions and needs to be estimated in numerical experiments or data.

Due to the fluctuations, the subsystem can behave as a sink or source for available energy. Sinks are parameterized in terms of diffusion-like eddy viscosities for positive diffusion coefficients. When the subscale system forces the large system, the process is described as backscatter. Energy is made available again. This backscatter can be transient or permanent. The LFT yields a guidance for stochastic backscatter parameterizations. In a stochastic model, the probability of a negative dissipation

function can be used as the probability for positive backscatter events at individual grid points. Out of these thoughts, it becomes clear that diffusion and backscatter parameterizations should have the same algebraic form in numerical models. A probability function for the diffusion coefficient describes the direction of the energy transfer. The occurrence of negative dissipation values due to parameterization of subscale fluxes has never been explicitly related to the possibility of observing deviations from positivity in the entropy production associated with the FT. A prerequisite to simulate and study transient negative dissipation is the formulation of the subscale flux terms such that in the long timescale limit, positive dissipation rates are guaranteed. Then, using the probability distribution of positive and negative diffusion coefficients derived from FT becomes feasible in a stochastic parametrization.

Shang et al. (2005) analyzed Rayleigh–Bénard convection in an upright cylindrical cell filled with water. The system is far from equilibrium. The authors determine fluctuations of the local vertical convective sensible heat transport $J_{s,z}$. The entropy production is given by the heat flux and the thermodynamic force which is the gradient of the inverse temperature,

$$\sigma \propto J_{s,z} \partial_z(1/T). \quad (7.26)$$

The Fluctuation Theorem determines the ratio of the probabilities for positive vs. negative heat transports

$$\log \left(\frac{P(+\bar{J}_{s,z})}{P(-\bar{J}_{s,z})} \right) = c \bar{J}_{s,z} \partial_z(1/T) \tau \quad (7.27)$$

where $\bar{J}_{s,z}$ is the running mean of the local flux $J_{s,z}$ in a time interval τ ; for the constant c , see Shang et al. (2005). The prediction of the linear relation (7.27) for the logarithm of the probability ratio versus $\bar{J}_{s,z}$ and τ was verified. The relation (7.27) provides a basis for the analysis of turbulent heat fluxes in the Planetary Boundary Layer simulated with the ICON-IAP model.

7.7 Applicability of the Fluctuation Theorem in Geophysical Flows

The validity of large deviation results on the Lyapunov exponents of geophysical fluid models has been recently discussed in Schalge et al. (2013). In this study, the entropy production rate is approximated by the local largest Lyapunov exponent (LLE). The largest Lyapunov exponent is determined by the instability of the state space trajectory for individual time steps (hence the exponent is not long-term averaged). Figure 7.6 shows a Gaussian fit of the LLE in the model PUMA with different resolutions which are typically used in many long-term climate integrations. The prediction of the Fluctuation Theorem

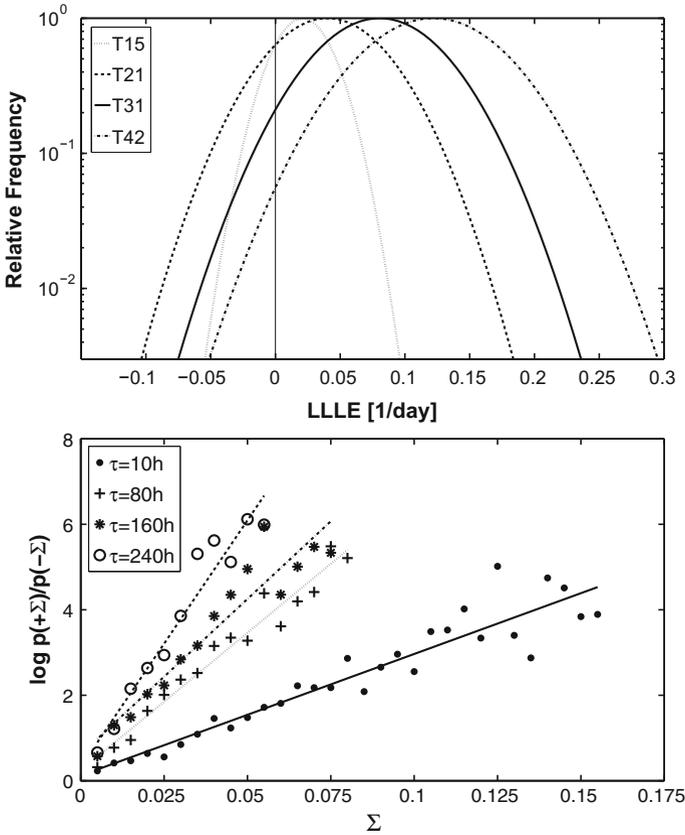


Fig. 7.6 Frequency distribution of the Local largest Lyapunov exponent (LLLE) in the model PUMA with different resolutions (top). Logarithm of the ratio of opposite sign probabilities versus entropy production for different time intervals in PUMA (bottom)

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \log \frac{P(+\Sigma)}{P(-\Sigma)} = \Sigma \tag{7.28}$$

was verified for different time windows τ for the linear relationship versus Σ (see Figure 7.6, bottom).

Negative values of the local largest Lyapunov exponent are due to a reduced instability for the global atmospheric flow. The results explain variations in predictability, in particular the so-called return of skill (Patil et al. 2001) which might be associated with negative local Lyapunov exponents. A further finding is that numerical models in geophysical fluid dynamics with $O(10,000)$ degrees of freedom can be considered as mesoscopic systems. Systems of this size have a transient non-negligible probability for being in a state with negative entropy production.

Seiffert et al. (2006) determined the subscale forcing of large scales in a global atmospheric primitive equation model to understand large-scale eddy diffusivities and backscatter. An aim of this study was to constrain stochastic forcings in weather forecast models. The approach used a T42 simulation of the model PUMA (Portable University Model of the Atmosphere) and retrieved the forcing of the embedded T21 modes. The residuals are projected on large-scale modes to determine parameterizations, and unexplained parts are interpreted as stochastic forcing. The analysis is performed in spectral space and allows an assessment of diffusion. The subscale forcing is largest in the mid-tropospheric mid-latitudes, and the decay timescale of unexplained residuals is of the order of one day. Simulations in T21 with additive stochastic forcing were compared.

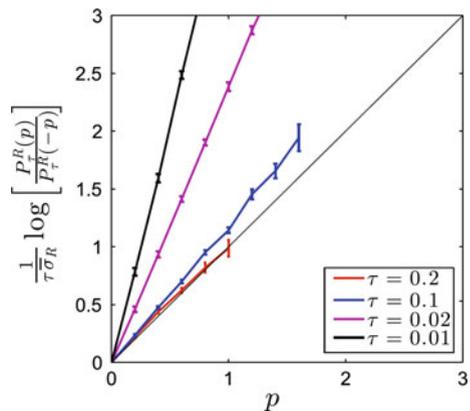
The experiments demonstrate that it is possible to derive subscale forcings for large scales and to interpret these as either stationary parameterizations or stochastic backscatter. Later on, the method has been used to determine stochastic forcings by the ECMWF (Hermanson et al. 2009).

A recent result by Gallavotti and Lucarini (2014) demonstrates that if the friction in a non-equilibrium system is adjusted appropriately to a so-called thermostatted system, dynamical properties are similar to an equilibrium system. The authors used the Lorenz-96 model with viscosity adjusted to preserve energy. The main property is that the dynamics is fully reversible. The average contraction rate of the phase space of the time-reversible model agrees with that of the irreversible model. The phase space contraction rate obeys the Fluctuation Theorem. The Fluctuation Theorem predicts for the τ -averaged phase space contraction rate

$$\frac{1}{\tau \bar{\sigma}_R} \log \frac{P_\tau^R(p)}{P_\tau^R(-p)} = p + O(\tau^{-1}) \tag{7.29}$$

where R is the constant forcing. Here $P_\tau^R(p)dp$ is the probability for observing the τ -averaged phase space contraction rate σ (normalized by the time average) in the interval $[p, p + dp]$. Figure 7.7 shows (7.29) for $R = 512$ versus p . For large τ , the correction term $\sim 1/\tau$ vanishes and the slope approaches 1.

Fig. 7.7 Fluctuation Theorem in the thermostatted Lorenz-96 model



The current proposal is an application of concepts developed in statistical mechanics to geophysical flows (Lucarini and Sarno 2011). These flows are forced and dissipative systems, where the asymptotic evolution is restricted to a strange attractor in a compact manifold Ruelle (2009). On the basis of the Chaotic Hypothesis of Gallavotti and Cohen (1995), it is assumed that the attractor is hyperbolic and stable against small stochastic perturbations. Time and phase space averages of measurable observables are identical.

To determine the response of non-equilibrium flows to perturbations, the Mori–Zwanzig formalism has been combined with the extension by Ruelle (Lucarini 2012; Lucarini and Colangeli 2012; Wouters and Lucarini 2012). Ruelle’s response theory is used to determine stochastic forcing in a multi-level system (Wouters and Lucarini 2012, 2013). In this analysis, stochastic terms with given spectral properties are obtained at second order in a perturbation analysis. A summary of mathematical and statistical mechanics approaches to geophysical fluid dynamics can be found in the review by Lucarini et al. (2014).

References

- Akmaev, R.A.: On the energetics of maximum-entropy temperature profiles. *Q. J. R. Meteorol. Soc.* **134**, 187–197 (2008)
- Arakawa, A.: Computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow. *J. Comput. Phys.* **1**, 119–143 (1966)
- Arakawa, A., Lamb, V.R.: A potential enstrophy and energy conserving scheme for the shallow water equations. *Mon. Wea. Rev.* **109**, 18–36 (1981)
- de Groot, S.R., Mazur, P.: *Nonequilibrium Thermodynamics*. Dover (1984)
- Dewar, R.C.: Maximum entropy production and the fluctuation theorem. *J. Phys. A Math. Gen.* **38**, 371 (2005)
- Egger, J.: Numerical generation of entropies. *Mon. Wea. Rev.* **127**, 2211–2216 (1999)
- Evans, D.J., Cohen, E.D.G., Morriss, G.P.: Probability of second law violations in shearing steady states. *Phys. Rev. Lett.* **71**, 2401 (1993)
- Evans, D.J., Searles, D.: The fluctuation theorem. *Adv. Phys.* **51**, 1529 (2002)
- Gallavotti, G.: A local fluctuation theorem. *Phys. A* **263**, 39–50 (1999)
- Gallavotti, G., Cohen, E.G.D.: Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.* **74**, 2694–2697 (1995)
- Gallavotti, G., Lucarini, V.: Equivalence of non-equilibrium ensembles and representation of friction in turbulent flows: the lorenz 96 model. *J. Stat. Phys.* **156**, 1027–1065 (2014)
- Gassmann, A.: Inspection of hexagonal and triangular C-grid discretizations of the shallow water equations. *J. Comput. Phys.* **230**, 2706–2721 (2011)
- Gassmann, A.: A global hexagonal C-grid non-hydrostatic dynamical core (ICON-IAP) designed for energetic consistency. *Q. J. R. Meteorol. Soc.* **139**, 152–175 (2013)
- Gassmann, A.: Entropy production due to subgrid-scale thermal fluxes in breaking gravity waves. *J. Atmos. Sci.*, subm.:0–0 (2016)
- Gassmann, A., Herzog, H.-J.: How is local material entropy production represented in a numerical model? *Q. J. R. Meteorol. Soc.* **141**, 854–869 (2015)
- Goody, R.: Sources and sinks of climate entropy. *Q. J. R. Meteorol. Soc.* **126**, 1953–1970 (2000)
- Hermanson, L., Hoskins, B., Palmer, T.: A comparative method to evaluate and validate stochastic parametrizations. *Q. J. R. Meteorol. Soc.* **135**, 1095–1103 (2009)

- Hollingsworth, A., Källberg, P., Renner, V., Burridge, D.M.: An internal symmetric computational instability. *Q. J. R. Meteorol. Soc.* **109**, 417–428 (1983)
- Liu, H.-L., Meriwether, J.W.: Analysis of a temperature inversion event in the lower mesosphere. *J. Geophys. Res.* **109**, D02S07 (2004)
- Lorenz, E. N. (1967). The nature and theory of the general circulation of the atmosphere. Technical Report, WMO—No. 218. TP. 115
- Lucarini, V.: Stochastic perturbations to dynamical systems: a response theory approach. *J. Stat. Phys.* **146**, 774–786 (2012)
- Lucarini, V., Blender, R., Herbert, C., Ragone, F., Pascale, S., Wouters, J.: Mathematical and physical ideas for climate science. *Rev. Geophys.* **52**, 809–859 (2014)
- Lucarini, V., Colangeli, M.: Beyond the linear fluctuation-dissipation theorem: the role of causality. *J. Stat. Mech. Theor. Exp.* **2012**, P05013 (2012)
- Lucarini, V., Sarno, S.: A statistical mechanical approach for the computation of the climatic response to general forcings. *Nonlinear Process. Geophys.* **18**, 7–28 (2011)
- Michel, G., Searles, D.J.: Local fluctuation theorem for large systems. *Phys. Rev. Lett.* **110**, 260602 (2013)
- Patil, D.J., Hunt, B.R., Kalnay, E., Yorke, J.A., Ott, E.: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.* **86**, 5878 (2001)
- Pauluis, O., Held, I.M.: Entropy budget of an atmosphere in radiative-convective equilibrium. Part I: maximum work and frictional dissipation. *J. Atmos. Sci.* **59**, 125–139 (2002a)
- Pauluis, O., Held, I.M.: Entropy budget of an atmosphere in radiative-convective equilibrium. Part II: Latent heat transport and moist processes. *J. Atmos. Sci.* **59**, 140–149 (2002b)
- Phillips, N.A.: An example of non-linear computational instability. In: *The Atmosphere and the Sea in Motion*, pp. 501–504. Rockefeller Institute Press, New York (1959)
- Priestley, C.H.B., Swinbank, W.C.: Vertical transport of heat by turbulence in the atmosphere. *Proc. R. Soc. Lond. A* **189**, 543–561 (1947)
- Raymond, D.J.: Sources and sinks of entropy in the atmosphere. *JAMES* **5**, 755–763 (2013)
- Romps, D.M.: The dry entropy budget of a moist atmosphere. *J. Atmos. Sci.* **65**, 3779–3799 (2008)
- Ruelle, D.: A review of linear response theory for general differentiable dynamical systems. *Non-linearity* **22**, 855 (2009)
- Schalge, B., Blender, R., Wouters, J., Fraedrich, K., Lunkeit, F.: Evidence for a fluctuation theorem in an atmospheric circulation model. *Phys. Rev. E* **87**, 052113 (2013)
- Seiffert, R., Blender, R., Fraedrich, K.: Subscale forcing in a global atmospheric circulation model and stochastic parametrization. *Q. J. R. Meteorol. Soc.* **132**, 1627–1643 (2006)
- Sevick, E.M., Prabhakar, R., Williams, S.R., Searles, D.J.: Fluctuation theorems. *Annu. Rev. Phys. Chem.* **59**, 603 (2008)
- Shang, X.-D., Tong, P., Xia, K.-Q.: Test of steady-state fluctuation theorem in turbulent rayleigh-bénard convection. *Phys. Rev. E* **72**, 015301 (2005)
- Smagorinsky, J.: Some historical remarks on the use of nonlinear viscosities. In: *Large Eddy Simulation of Complex Engineering and Geophysical Flows*, pp. 3–36. Cambridge University Press (1993)
- Verkley, W.T.M., Gerkema, T.: On maximum entropy profiles. *J. Atmos. Sci.* **61**, 931–936 (2004)
- Wan, H., Giorgetta, M.A., Zängl, G., Restelli, M., Majewski, D., Bonaventura, L., Fröhlich, K., Reinert, D., Rípodas, P., Kornbluh, L.: The ICON-1.2 hydrostatic atmospheric dynamical core on triangular grids, Part I: formulation and performance of the baseline version. *Geosci. Model. Dev.* **6**, 663–693 (2013)
- Woolings, T., Thurn, J.: Entropy sources in a dynamical core atmosphere model. *Q. J. R. Meteorol. Soc.* **132**, 43–59 (2006)
- Wouters, J., Lucarini, V.: Disentangling multi-level systems: averaging, correlations and memory. *J. Stat. Mech. Theor. Exp.* **2012**, P03003 (2012)
- Wouters, J., Lucarini, V.: Multi-level dynamical systems: connecting the ruelle response theory and the Mori-Zwanzig approach. *J. Stat. Phys.* **151**, 850–860 (2013)

Chapter 8

Reducing Spurious Diapycnal Mixing in Ocean Models



Knut Klingbeil, Hans Burchard, Sergey Danilov, Claus Goetz and Armin Iske

Abstract Transport algorithms of numerical ocean circulation models are frequently exhibiting truncation errors leading to spurious diapycnal mixing of water masses. This chapter discusses methods that might be useful in diagnosing spurious diapycnal mixing and describes some approaches that might be helpful for its reduction. The first one is related to the use of the Arbitrary Lagrangian Eulerian (ALE) vertical coordinate which allows the implementation of vertically moving meshes that may partly follow the isopycnals even if the basic vertical coordinate differs from isopycnal. The second approach relies on modified advection schemes with the dissipative part of the transport operators directed isopycnally. Finally the third approach deals with new efficient and stable advection algorithms of arbitrary high order based on the WENO-ADER method, which can be applied to both structured and unstructured meshes. While practical benefits of using the reviewed approaches depend on applications, there are indications that equipping present state-of-the-art ocean circulation models with them would lead to reduced spurious transformations.

K. Klingbeil · H. Burchard
Leibniz Institute for Baltic Sea Research Warnemünde (IOW),
Rostock-Warnemünde, Germany
e-mail: knut.klingbeil@io-warnemuende.de

H. Burchard
e-mail: hans.burchard@io-warnemuende.de

S. Danilov
Alfred Wegener Institute for Polar and Marine Research (AWI),
Bremerhaven, Germany
e-mail: sergey.danilov@awi.de

S. Danilov
Jacobs University, Bremen, Germany

C. Goetz · A. Iske (✉)
Fachbereich Mathematik, Universität Hamburg, Hamburg, Germany
e-mail: armin.iske@uni-hamburg.de

C. Goetz
e-mail: claus.goetz@uni-hamburg.de

8.1 Introduction

Spurious mixing of water masses in ocean circulation models based on a z - or generalised terrain-following vertical coordinate is a numerical artefact intrinsic to transport algorithms that use upstream flux estimates or employ limiters to avoid the development of artificial extreme values and negative concentrations. The interior of the ocean is nearly adiabatic, which implies that water parcels generally move along surfaces of neutral density. Since these surfaces are intersected by coordinate levels, advection of a water parcel by a (high-order) upwind or total variance diminishing (TVD) transport algorithm is accompanied by spurious diapycnal mixing caused by truncation errors. Numerous studies (see, e.g., Urakawa and Hasumi 2014) demonstrate the effect of mixing, and there is ongoing research aimed at quantifying spurious mixing in ocean circulation models, understanding its causes, and designing the technology that allows one to minimise its adverse effects. In addition to water mass modification, spurious mixing also impairs the energetic consistency. Indeed the increase in the flow's potential energy accompanying spurious mixing happens without any available energy sources, which is physically incorrect and prevents one from formulating a closed energy balance. Estimates in Ilicak et al. (2012) show that spurious diapycnal mixing may imply energy fluxes making up to a quarter of energy input by tides and winds. As a consequence, the way toward formulation of energetically consistent ocean circulation models demands solving the questions on spurious diapycnal mixing. These concerns are not common for atmospheric modelling. In contrast to the ocean, which is largely adiabatic in the interior outside mixed and boundary layers and some vicinity of topographic features, the atmosphere is a forced-dissipative system where spurious effects are masked. Our focus is therefore on the ocean where spurious effects are of more significance than in the atmosphere.

The following sections of this chapter describe a selection of approaches that can contribute to the reduction of spurious mixing in ocean circulation models. The rest of this section explains briefly why these approaches have been selected, with intention to help the reader to see the general framework.

The question on diagnosing spurious mixing is a subject of numerous papers. Griffies et al. (2000), Burchard and Rennau (2008), Getzlaff et al. (2010, 2012), Hill et al. (2012), Klingbeil et al. (2014), Ilicak (2016) and Gibson et al. (2017) introduce methods to quantify either diapycnal or full spurious mixing associated with transport schemes. While the diagnostics of Griffies et al. (2000) and Gibson et al. (2017), based in essence on the concept of reference potential energy (RPE), is global, the quantification in Getzlaff et al. (2010, 2012) and Hill et al. (2012) is based on the analysis of tracer spreading and could be sensitive to the initial structure of the tracer field. The concept of RPE has certain caveats as applied to the real ocean with a non-linear equation of state, so that using it implies some approximations. Ilicak (2016) introduced the temporal derivative of RPE density as a local measure of diapycnal mixing. The variance decay approach of Burchard and Rennau (2008) and its further developments by Klingbeil et al. (2014) allows one to obtain a three-dimensional view of the total spurious dissipation, see Figure 8.1.

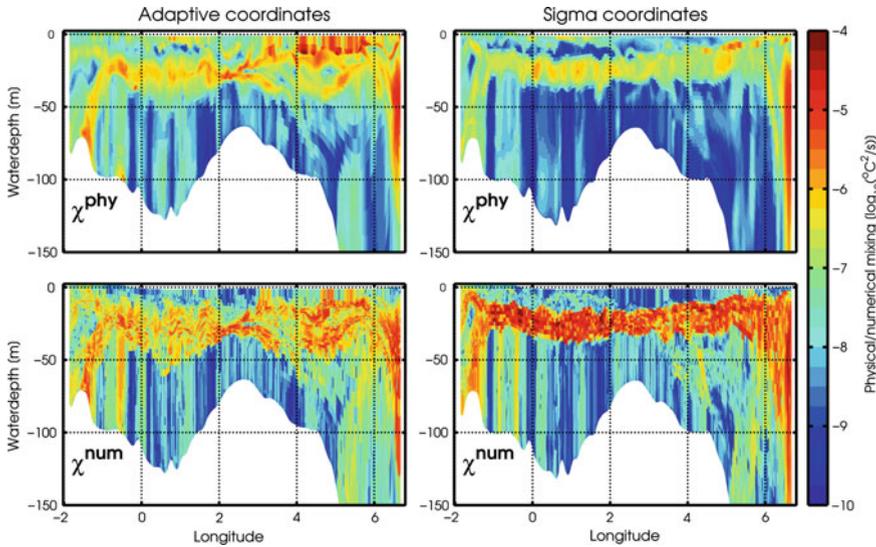


Fig. 8.1 Physical mixing (upper panels) and numerical mixing (lower panels) of temperature along a transect across the North Sea simulated with GETM using adaptive coordinates (left) and σ coordinates (right). A reduction of numerical mixing and an according increase of physical mixing when using adaptive coordinates is clearly seen. This figure is adapted from Gräwe et al. (2015)

Further applications of this approach and the RPE analysis by Mohammadi-Aragh et al. (2015) to a selection of transport schemes show that the amount of mixing is related to the type of scheme (dissipative or neutral), and that certain limiters may lead to spurious unmixing, which is as undesirable as spurious mixing, for it still creates new water masses. Schemes providing highest accuracy are found to show minimum spurious effects, in agreement with the conclusion by Hill et al. (2012).

These results indicate that the use of more accurate transport schemes in ocean models presents a promising direction despite the long history of research and the availability of numerous accurate algorithmic solutions. Additional motivation comes from the fact that at present some groups work with or develop ocean circulation models formulated on unstructured meshes (see Danilov 2013). Climate system models using the ocean component fitting in this class are already available and new ones will appear soon (see, e.g., Wan et al. 2013; Ringler et al. 2013; Wang et al. 2014; Sidorenko et al. 2015). Accurate and cost effective transport schemes for unstructured meshes and low-order spatial discretisations are actively discussed, and current practical solutions are equivalent to the third-order upwind or fourth-order centered differences if the mesh is locally uniform (see, e.g., Skamarock and Gassmann 2011; Danilov 2013). The algorithms based on direct flux estimates involving high-order centered polynomial reconstruction do not generally show significant improvements compared to the methods analogous to that by Skamarock and Gassmann (2011). Further developments in this direction augmented with practical estimates of the effect on spurious mixing present a challenging direction of research.

Toro et al. (2001) proposed an explicit one-step finite volume scheme, termed **ADER**, which is of **A**rbitrary high order, using high-order **DER**ivatives of polynomials. The finite volume discretisation of Toro et al. (2001) combines high-order polynomial reconstruction from cell averages with high-order flux evaluation. The latter is done by solving generalised Riemann problems across the cell interfaces, i.e., boundaries of adjacent control volumes. Therefore, the finite volume ADER scheme of the seminal work Toro et al. (2001) (and from their subsequent developments Toro and Titarev 2001; Titarev and Toro 2002) can be viewed as a generalisation of the classical first-order Godunov scheme to arbitrary high orders.

ADER schemes have since then gained considerable popularity in applications to gas and aerodynamics (see e.g. the text book by Toro 2009), and they were shown to be of particular high performance in combination with WENO (Weighted Essentially Non-Oscillatory) reconstructions. In fact, the resulting ADER-WENO method yields a computational method for the numerical simulation of fluid flow of arbitrary high order (see, e.g., Dumbser and Käser 2007) on both structured and unstructured meshes.

WENO reconstruction is a basic ingredient for a class of finite volume methods, used to avoid spurious oscillations in the solution of nonlinear hyperbolic problems. The WENO method works with a weighted sum (convex combination) of recovery functions from scattered samples, where the samples, for each reconstruction, are taken from stencils. The construction of the stencils requires particular care, as well as the selection of a suitable oscillation indicator (to determine the weights in the convex combination). Kernel-based methods provide natural choices of oscillation indicators, unlike traditional WENO schemes that are relying on polynomial reconstructions. The kernel-based reconstruction method is mesh-free, optimal (in the sense of splines), very flexible and inherently supports adaptivity. This was first demonstrated in the work by Aboiyar et al. (2010), where kernel-based WENO reconstructions were introduced first.

Alongside with the research on increasing accuracy of transport algorithms there are promising results related to the measures helping to minimise spurious effects accompanying transport. They have been proposed by the ROMS (Regional Ocean Modeling System) community and consist in modifying the high-order upwind-biased advection schemes by splitting them into a centered, non-diffusive part (dispersive truncation error) and a diffusive (biharmonic) operator, with the diffusion coefficient depending on the amplitude of velocity and mesh spacing. The biharmonic operator is further modified by directing it to mix along the isoneutral surfaces (Lemarié et al. 2012a, b). Only small residual mixing, related to numerical stabilisation, remains in this case. The technology is most rewarding for models formulated on terrain-following coordinates, since slopes between isopycnals and model levels may be substantial in this case in many places. Indeed, this is the main motivation behind the studies by Lemarié et al. (2012a, b). Although the terrain-following meshes are not common in global ocean circulation modelling, mixed meshes, combining the geopotential and terrain-following levels, are beginning to be used to better resolve processes on the continental shelf (see, e.g., Nakayama et al. 2014) or in ice cavities (see Timmermann et al. 2012), where accurate rendering of local topography is of

primary importance. In these situations the reduction in spurious mixing is a desired feature of transport algorithms.

Spurious mixing depends on the choice of vertical coordinates because this choice defines the angles between isoneutral surfaces and model levels as well as the grid-related vertical velocity. Additionally, spurious mixing is coupled to sub-grid momentum dissipation (Ilicak et al. 2012) and, more broadly, to the structure (smoothness) of the velocity field at smallest resolvable scales, which also depends on the mesh vertical motion. The aspects of vertical discretisation and mesh motion can be handled in the framework of Arbitrary Lagrangian Eulerian (ALE) vertical coordinates. The ALE method is well known in computational fluid dynamics (for a brief review see Donea et al. 2004) and presents a reformulation of motion equations to moving coordinates. One gets a Lagrangian representation if the mesh follows fluid particles, an Eulerian representation if it is fixed and any arbitrary combination in between, hence the name. In ocean modelling the ALE vertical coordinate was in fact used by models with generalized vertical coordinates. However, the notion of ALE was hardly mentioned before the analysis of Adcroft and Hallberg (2006).

In the ALE implementation proposed by Adcroft and Hallberg (2006) coordinate surfaces are moved with the vertical flow in a Lagrangian manner (thus without any spurious mixing) and then relocated using a remapping algorithm to take into account watermass transformation or any other mesh tendency. Alternatively, the ALE vertical coordinate can be implemented as the adjustment of model levels according to a selected law, such as tendency towards isopycnals or locally increased layer number in places with sharp density transitions. Through its ability to take into account the isopycnal (Lagrangian) tendency, the ALE vertical coordinate may lead to a substantial reduction in spurious mixing if paired with appropriate algorithms of layer motion. This motivates the use of this coordinate in new ocean model developments such as MPAS-ocean (Model for Prediction Across Scales, Los Alamos National Laboratory, Ringler et al. 2013) or MOM6 (Modular Ocean Model, Geophysical Fluid Dynamics Laboratory). Needless to say that the ALE vertical coordinate is in essence the basis of the Hybrid Coordinate Ocean Model (HYCOM) and numerous coastal models including the General Estuarine Transport Model (GETM) and many others. However, the ALE method is just the way of writing motion equations, and the degree of spurious mixing reduction in setups that are based on geopotential or terrain-following coordinate depends on the particular algorithm of mesh motion. Two particular implementations deserve mentioning.

The first one involves the non-uniform adaptation of vertical terrain-following grids as proposed by Hofmeister et al. (2010), see Figure 8.1. It aims at adjusting the layers thicknesses depending on a goal function that favours increased resolution and the alignment of level surfaces with isopycnals in places where the density gradients are strong, which results in a significant reduction in spurious effects (Gräwe et al. 2015). The latter functionality is, in fact, available in GETM (see, e.g., Hofmeister et al. 2010).

The second one is the so-called \tilde{z} -coordinate (Leclair and Madec 2011) which aims to follow high-frequency baroclinic motions, so that transport operators “feel” only the low-pass filtered vertical exchange. The intention is to reduce spurious

mixing in regimes when internal wave contributions are substantial. The experiments by Petersen et al. (2015) show that it is indeed possible. Except for these examples the question of optimal algorithms for level motion in function of basic vertical coordinate (z or terrain-following) and dynamical regimes remains largely unexplored.

Below we provide a brief review of the developments mentioned above. Many aspects remain a subject of research and may need further adjustments. Other measures of reducing spurious mixing may exist. The selection here picks up the measures that reflect the authors' interests and are practically affordable for the current global ocean circulation models. Furthermore, there are indications of their efficiency in the available literature. Learning about their actual contributions in the reduction of spurious mixing is a challenging program for future research.

The plan of the rest of this chapter is as follows. We start from the methods of diagnosing spurious mixing and present examples showing its behaviour in test cases and in practice (Section 8.2). We continue in Section 8.3 with the description of the ALE framework and details of the mesh adjustment technology. The description of the approaches by Lemarié et al. (2012a, b) will be presented in Section 8.4 and ADER-WENO algorithms will be reviewed in Section 8.5. In the Conclusion we briefly outline some of the open questions that could be addressed in future research.

8.2 Diagnosing Spurious Mixing

The local quantification of spurious mixing helps to detect hotspots of spurious mixing and offers the comparison with physically induced mixing. This is essential for the development and assessment of new numerical techniques like advection schemes and algorithms for mesh adaptation.

8.2.1 *An Analytical Example*

Spurious mixing is induced by truncation errors due to the discretisation of the original differential equation. Because of these truncation errors the solution of a discrete difference scheme satisfies not the original differential equation but a *modified equation* (Hirt 1968; Warming and Hyett 1974). In principle the modified equation can be derived analytically by Taylor series expansion of the discrete quantities within the difference scheme.

For example, for the 1D advection equation,

$$\frac{\partial \varphi}{\partial t} + \frac{\partial}{\partial x} \{u\varphi\} = 0, \quad (8.1)$$

the discretisation on an equidistant grid with the explicit First-Order Upstream (FOU) scheme for a constant velocity $u > 0$ is given by

$$\frac{\varphi_i^{(n+1)} - \varphi_i^{(n)}}{\Delta t} + u \frac{\varphi_i^{(n)} - \varphi_{i-1}^{(n)}}{\Delta x} = 0. \quad (8.2)$$

Taylor series expansion around $\varphi_i^{(n)} = \varphi(x = i\Delta x, t = n\Delta t)$ yields the modified equation

$$\frac{\partial \varphi}{\partial t} + u \frac{\partial \varphi}{\partial x} + \frac{1}{2} \left(\Delta t \frac{\partial^2 \varphi}{\partial t^2} - u \Delta x \frac{\partial^2 \varphi}{\partial x^2} \right) + \frac{1}{6} \left(\Delta t^2 \frac{\partial^3 \varphi}{\partial t^3} + u \Delta x^2 \frac{\partial^3 \varphi}{\partial x^3} \right) + \text{HOT} = 0, \quad (8.3)$$

with higher-order terms lumped into HOT. Successive replacement of higher-order temporal derivatives and mixed derivatives in (8.3) by the modified equation itself results in

$$\begin{aligned} 0 = & \frac{\partial \varphi}{\partial t} + u \frac{\partial \varphi}{\partial x} \\ & - \frac{1}{2} C (1 - C) \frac{\Delta x^2}{\Delta t} \frac{\partial^2 \varphi}{\partial x^2} \\ & + \frac{1}{6} C (1 - C) (1 - 2C) \frac{\Delta x^3}{\Delta t} \frac{\partial^3 \varphi}{\partial x^3} + \text{HOT}, \end{aligned} \quad (8.4)$$

with the Courant number

$$C = \frac{u \Delta t}{\Delta x}. \quad (8.5)$$

Thus, instead of the original advection equation (first row in (8.4)), a modified equation including a diffusion term (second row) and a dispersion term (third row) as well as further higher-order terms is solved by the solution obtained from the explicit FOU-scheme (8.2). It is this erroneous diffusion term that causes spurious mixing. The associated numerical diffusivity of the explicit FOU-scheme,

$$\nu^{\text{FOU}} = \frac{1}{2} C (1 - C) \frac{\Delta x^2}{\Delta t}, \quad (8.6)$$

consists of contributions from temporal truncation errors (first-order expansion around $^{(n)}$) and spatial truncation errors (upstream approximation of advective fluxes).

The analytical derivation of the diffusion term in the modified equation is impractical for the more advanced higher-order non-linear schemes on non-equidistant grids used in state-of-the-art numerical models. An alternative, general diagnostic method to quantify numerical mixing is presented in the remainder of this section.

8.2.2 Variance Decay as a Measure for Mixing and Dissipation

An alternative to the direct evaluation of diffusion terms is the indirect quantification of the effects of diffusion. As will be shown below, this can be done in terms of variance decay. Starting with the 1D diffusion equation,

$$\frac{\partial \varphi}{\partial t} + \frac{\partial}{\partial x} \left\{ -\nu \frac{\partial \varphi}{\partial x} \right\} = 0, \quad (8.7)$$

a diffusion equation for the second moment can be derived by multiplication of (8.7) with 2φ :

$$\frac{\partial (\varphi^2)}{\partial t} + \frac{\partial}{\partial x} \left\{ -\nu \frac{\partial (\varphi^2)}{\partial x} \right\} = -2\nu \left(\frac{\partial \varphi}{\partial x} \right)^2 = -\chi^{\text{ana}}(\varphi) \leq 0. \quad (8.8)$$

The term $\chi^{\text{ana}}(\varphi)$ on the right-hand side represents the local variance decay rate associated with the diffusion of φ . This variance decay rate quantifies the smoothing of gradients, i.e. the mixing of tracers or the dissipation of kinetic energy associated with the diffusion of velocity ($\varphi = u$).

Since there is no analytical variance decay associated with advection (derived from the advection equation (8.1)),

$$\frac{\partial (\varphi^2)}{\partial t} + \frac{\partial}{\partial x} \{ u (\varphi^2) \} = 0, \quad (8.9)$$

the variance decay rate can be established as a reasonable indirect measure for diffusion and the associated mixing and dissipation.

8.2.3 Discrete Variance Decay

Although no variance decay is associated with the continuous advection equation, see (8.9), numerical advection schemes cause spurious *discrete variance decay* (DVD) due to truncation errors as demonstrated in Section 8.2.1.

A Finite-Volume (FV) discretisation of the advection equation (8.1), that facilitates a later extension to 3D, is given by

$$\frac{V_i^{(n+1)} \varphi_i^{(n+1)} - V_i^{(n)} \varphi_i^{(n)}}{\Delta t} + [A_{i'} u_{i'} \tilde{\varphi}_{i'}]_{i'=i-1/2}^{i'=i+1/2} = 0, \quad (8.10)$$

with V_i being the volume of the i -th FV-cell and with the discrete first moment $\varphi_i^{(n)} = \frac{1}{V_i^{(n)}} \int_{V_i^{(n)}} \varphi dV$. Rectangular brackets $[\psi_{i'}]_{i'=i-1/2}^{i'=i+1/2} = \psi_{i+1/2} - \psi_{i-1/2}$ will be used as

shorthand notation for differences. Quantities with half indices are located at the corresponding interface between two FV-cells, e.g. the area of the interface $A_{i'}$ and the interfacial values $\tilde{\varphi}_{i'}$ approximated according to the applied advection scheme.

In order to quantify the associated variance decay, it is obvious that the discrete second moment equation must not be obtained by an independent discretisation of the analytical second moment equation (8.9), but should be derived from the original prognostic equation (8.10). This also guarantees the consistency of discrete moments, i.e. $(\varphi^2)_i^{(n+1)} \equiv (\varphi_i^{(n+1)})^2$, and actually turns the discrete second moment equation into a diagnostic equation for the DVD rate. The derivation of a straight-forward expression for the DVD rate is only possible for trivial schemes. Nevertheless, a general derivation of the discrete second moment equation is possible, as discussed in Klingbeil et al. (2014). Formally (8.10) can be split into two steps, the first describing the decomposition of the old FV-cell $^{(n)}$ into subvolumes advecting out to neighbouring cells and into the portion $^{(*)}$ remaining in the cell, and the second representing the recombination of the remained portion and the subvolumes advected into the cell (see Prather 1986),

$$V_i^{(*)} \varphi_i^{(*)} = V_i^{(n)} \varphi_i^{(n)} - \Delta t \left[A_{i+1/2} u_{i+1/2}^+ \tilde{\varphi}_{i+1/2} - A_{i-1/2} u_{i-1/2}^- \tilde{\varphi}_{i-1/2} \right], \quad (8.11a)$$

$$V_i^{(n+1)} \varphi_i^{(n+1)} = V_i^{(*)} \varphi_i^{(*)} - \Delta t \left[A_{i+1/2} u_{i+1/2}^- \tilde{\varphi}_{i+1/2} - A_{i-1/2} u_{i-1/2}^+ \tilde{\varphi}_{i-1/2} \right], \quad (8.11b)$$

with $u_{i'}^+ = \max\{0, u_{i'}\}$ and $u_{i'}^- = \min\{u_{i'}, 0\}$. A graphical example for this decomposition and recombination is shown in Figure 8.2. The existence of subvolumes offers a mathematically sound definition of variance in a single FV-cell (Morales Maqueda and Holloway 2006). The variance gain during the decomposition $(\Delta\sigma^2)^d$ and the variance loss during the recombination $(\Delta\sigma^2)^r$ can be calculated as

$$\begin{aligned} & (\Delta\sigma^2)^{d,(n)} \\ &= \left(\frac{V_i^{(*)} (\varphi_i^{(*)})^2 + \Delta t A_{i+1/2} u_{i+1/2}^+ (\tilde{\varphi}_{i+1/2})^2 - \Delta t A_{i-1/2} u_{i-1/2}^- (\tilde{\varphi}_{i-1/2})^2}{V_i^{(n)}} - (\varphi_i^{(n)})^2 \right) - 0, \end{aligned} \quad (8.12a)$$

$$\begin{aligned} & (\Delta\sigma^2)^{r,(n+1)} \\ &= 0 - \left(\frac{V_i^{(*)} (\varphi_i^{(*)})^2 - \Delta t A_{i+1/2} u_{i+1/2}^- (\tilde{\varphi}_{i+1/2})^2 + \Delta t A_{i-1/2} u_{i-1/2}^+ (\tilde{\varphi}_{i-1/2})^2}{V_i^{(n+1)}} - (\varphi_i^{(n+1)})^2 \right). \end{aligned} \quad (8.12b)$$

The net DVD rate can then be defined as

$$\chi_i^{\text{adv.},(n+1)} = -\frac{1}{\Delta t V_i^{(n+1)}} \left(V_i^{(n)} (\Delta\sigma^2)_i^{d,(n)} + V_i^{(n+1)} (\Delta\sigma^2)_i^{r,(n+1)} \right). \quad (8.13)$$

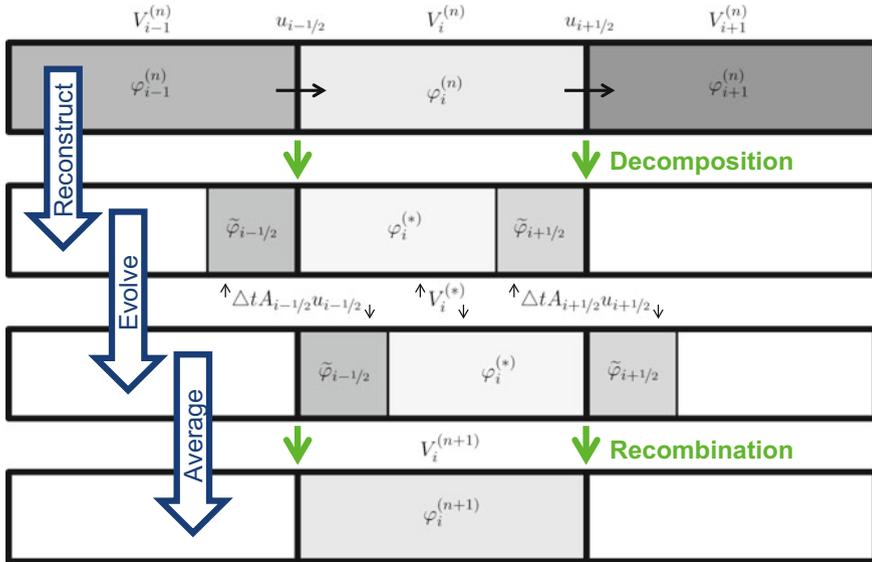


Fig. 8.2 Advection in a FV-framework (taken from Klingbeil et al. 2014): The schema sketches the exemplary decomposition of FV-cells into subvolumes and their recombination on a fixed 1D grid for uniform flow. V , A , u , φ and $\tilde{\varphi}_{i\pm 1/2}$ denote the volumes and interfacial areas of the FV-cells, the velocity, the quantity to be advected and its approximated interfacial values. During the decomposition subvolumes and variance are created inside a single FV-cell, whereas during the recombination subvolumes and variance are destroyed. The associated variance gain and loss determine the net variance decay during one timestep and can be calculated according to (8.12a)–(8.13)

For anti-dissipative advection schemes, which tend to steepen gradients, the net DVD rate can be negative. Combination of (8.12a), (8.12b) and (8.13) finally yields the discrete second moment equation, from which the DVD rate can be diagnosed:

$$\frac{V_i^{(n+1)} \left(\varphi_i^{(n+1)} \right)^2 - V_i^{(n)} \left(\varphi_i^{(n)} \right)^2}{\Delta t} + [A_{i'} u_{i'} \tilde{\varphi}_{i'}^2]_{i'=i-1/2}^{i'=i+1/2} = -V_i^{(n+1)} \chi_i^{\text{adv},(n+1)}. \tag{8.14}$$

The extension of (8.14) to 3D is straightforward by simply considering the additional fluxes through the other interfaces of the FV-cell.

For the 1D explicit FOU-scheme (8.14) can be derived directly by multiplication of (8.10) with $(\varphi_i^{(n)} + \varphi_i^{(n+1)})$ and further reformulation with the help of the continuity equation ((8.10) with $\varphi = 1$), see Morales Maqueda and Holloway (2006) and Burchard (2012). Interestingly, for $u_{i'} > 0$ the derived DVD rate coincides with a straightforward discretisation of the analytical variance decay rate derived in (8.9)

for a diffusion equation with the numerical diffusivity of the FOU scheme derived in (8.6):

$$\chi_i^{\text{FOU},(n+1)} = 2v_i^{\text{FOU},(n+1)} \left(\frac{\varphi_i^{(n)} - \varphi_{i-1}^{(n)}}{\Delta x_i} \right)^2. \quad (8.15)$$

The numerical diffusivity in (8.15),

$$v_i^{\text{FOU},(n)} = \frac{1}{2} C_i^{(n)} \left(1 - C_i^{(n)} \right) \frac{\Delta x_i^2}{\Delta t}, \quad (8.16)$$

depends on a generalised Courant number $C_i^{(n)} = \Delta t A_{i-1/2} u_{i-1/2} / V_i^{(n)}$ which coincides with (8.5) for equally-sized FV-cells.

8.2.4 Applications

The diagnostic method (8.14), outlined in Section 8.2.3, was implemented into the General Estuarine Transport Model (GETM, Burchard and Bolding 2002; Hofmeister et al. 2010; Klingbeil and Burchard 2013). Here results from its application to a realistic Western Baltic Sea model are presented. In the Western Baltic Sea saline bottom inflows from the Kattegat are subject to mixing induced by various sources. This mixing determines the final salinity of the plumes and thus their interleaving at the corresponding density levels in the downstream basins of the Baltic proper. Therefore, the accurate simulation of the mixing in the Western Baltic Sea is essential for realistic model results. The effective mixing within a model combines the well-calibrated physically induced mixing due to turbulence closure schemes and the spurious mixing induced by the numerics. In Figure 8.3 physically and numerically induced mixing of salinity in the Western Baltic Sea are juxtaposed. As is obvious from (8.8), any kind of mixing requires both, sharp gradients and high diffusivities. Therefore, significant mixing of salinity usually takes place in the vicinity of the halocline (see Figure 8.3c, d). High-turbulent regions with increased diffusivities can be detected as hotspots of physically induced mixing, e.g. boundary layers. Coarse vertical resolution, strong advective fluxes through the cell interfaces and the non-alignment of model layers with isohalines can be identified as sources for increased numerically induced salt mixing.

It is noteworthy that even in today's realistic simulations with state-of-the-art ocean models, numerically induced spurious mixing is still of the same order of magnitude as the physically induced counterpart. In Section 8.3 adaptive meshing techniques will be outlined that (i) refine the spatial resolution in the vicinity of gradients to decrease truncation errors, (ii) move with the fluid to minimise the advective transports through the mesh, and (iii) align the model layers with isopycnals to facilitate a reasonable physical mixing.

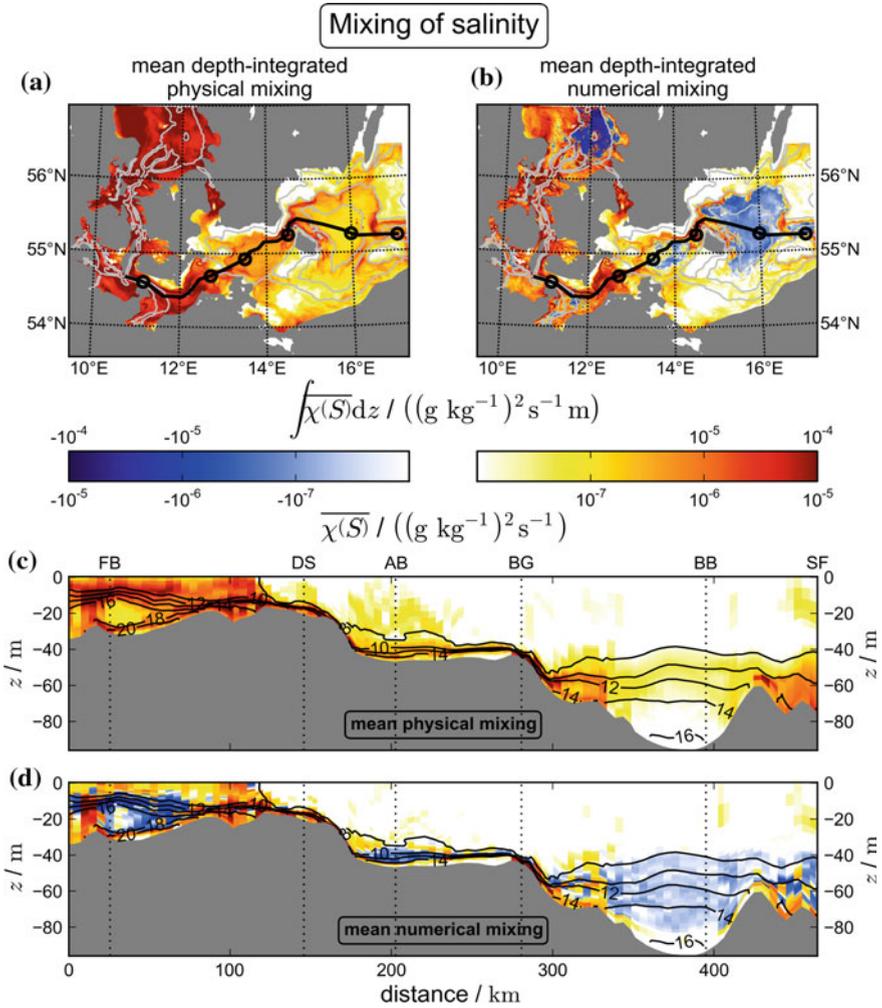


Fig. 8.3 Mixing of salinity in a realistic Western Baltic Sea model (taken from Klingbeil et al. 2014): Depth-integrated (**a**, **b**) and depth-resolved (**c**, **d**) mixing rates averaged over 19–28 Sep 2008. Ticks above the colorbars belong to a, b and ticks below to c, d. Grey contour lines in a, b show isobaths with an interval of 25 m. The black line with markers depicts the transect along the stations Fehmarn Belt (FB), Darss Sill (DS), Arkona Basin (AB), Bornholmsgat (BG), Bornholm Basin (BB) and Stupsk Furrow (SF), shown in c, d. In c, d isohalines with an interval of 2 g kg^{-1} are drawn as contour lines. It is obvious that numerically induced spurious mixing can be of the same order as the physical mixing contribution. Negative numerical mixing indicates spurious steepening of gradients due to anti-dissipative advection schemes

8.3 Arbitrary Lagrangian Eulerian Vertical Coordinate

The Arbitrary Lagrangian Eulerian (ALE) framework is commonly used in computational fluid dynamics in situations when dynamical equations have to be considered on moving meshes. As such, it is a subject of textbooks, and its brief exposition can be found, for example in Donea and Huerta (2003) and Donea et al. (2004). The essence of the method reduces to the observation that for a scalar quantity φ the material derivative (for a fixed fluid particle Lagrangian coordinate \mathbf{X}) on a moving mesh can be written as

$$\frac{D\varphi}{Dt} = \frac{\partial\varphi}{\partial t}|_{\mathbf{x}} = \frac{\partial\varphi}{\partial t}|_{\boldsymbol{\chi}} + \mathbf{c} \frac{\partial\varphi}{\partial \mathbf{x}}, \quad (8.17)$$

where $\boldsymbol{\chi}$ are the coordinates related to the moving mesh points and \mathbf{x} are the original (fixed) Eulerian coordinates. In this formula the time derivative is computed following the mesh and $\mathbf{c} = \mathbf{v} - \mathbf{v}_m = D\mathbf{x}/Dt - \partial\mathbf{x}/\partial t|_{\boldsymbol{\chi}}$ is the so called ‘‘convection’’ velocity, i.e. the difference of the Eulerian fluid velocity (the first term) and mesh velocity (the second term), or in still other words, the transport velocity through the mesh surfaces. Introducing fluid particle velocity in $\boldsymbol{\chi}$ coordinates, $\mathbf{w} = \partial\boldsymbol{\chi}/\partial t|_{\mathbf{x}}$, we have $\mathbf{c} = \mathbf{w}(\partial\mathbf{x}/\partial\boldsymbol{\chi})$, i.e., they are connected through the Jacobian. Note that the gradient term in (8.17) is expressed with respect to the original coordinates \mathbf{x} . The same modification concerns all other full time derivatives. Perhaps a more handy way of expressing this, especially in the view of finite-volume discretisation used in practice, is to use the Reynolds transport theorem, writing for a conserved quantity φ (like a mass fraction)

$$\frac{d}{dt} \int_{V(t)} \varphi dV + \int_{\partial V(t)} \varphi \mathbf{c} \mathbf{n} dA = 0, \quad (8.18)$$

where ∂V is the moving boundary of volume V and \mathbf{n} is the outer normal to the surface. Thus, the ALE procedure reduces in essence to using the velocity of motion relative to the moving mesh to describe the exchange between the interacting volumes. Note that the mesh need not necessarily to move, it can be fixed, with arbitrarily drawn (but not intersecting) surfaces. Whatever the case one will be working with the transport velocity through the faces of computational cells on this mesh.

The oceanographic community came to the use of ALE framework through the concept of a generalized vertical coordinate $r = r(x, y, z, t)$, a special case of $\boldsymbol{\chi}$ above. Transforming from the (x, y, z, t) to (x, y, r, t) representation one gets the basic set of governing equations in the form

$$\frac{D\mathbf{u}}{Dt} + f\mathbf{k} \times \mathbf{u} + (1/\rho_0)(\nabla_r p + g\rho\nabla_r z) = \mathbf{F}, \quad (8.19)$$

$$\partial_r p + g\rho\partial_r z = 0, \quad (8.20)$$

$$\partial_t h + \nabla_r(h\mathbf{u}) + \partial_r(hr\dot{r}) = 0, \quad (8.21)$$

$$\partial_t(h\varphi) + \nabla_r(h\mathbf{u}\varphi) + \partial_r(hr\dot{r}\varphi) = Q_\varphi, \quad (8.22)$$

$$\partial_t \eta + \nabla_r \int \mathbf{u} h dr = P. \quad (8.23)$$

Here $h = \partial_r z$ is the Jacobian of transform from z to r coordinate, which will be called thickness in what follows, φ is the temperature T or salinity S , \mathbf{F} and Q_φ are the “sources” in the respective equations, the operator ∇ is the horizontal one. The index r with ∇ implies that computations are along the surfaces of constant r , while the same index with ∂ points to the derivative with respect to r . The rest of notations is standard, with \mathbf{u} the horizontal velocity, g the acceleration due to gravity, p the pressure, ρ_0 and ρ the reference density and the deviation from it respectively, f the Coriolis parameter, \mathbf{k} the vertical unit vector, η the elevation and P the freshwater added to the ocean. The density ρ obeys the equation of state $\rho = \rho(T, S, p_0)$ ($p_0 = -\rho_0 g z$).

This set of equations includes the cases of isopycnal coordinate when r is the potential density, or generalized terrain following coordinate (including the standard sigma-coordinate) if r is defined appropriately. By construction, the equations above are already written in the ALE framework, restricted to the case of only vertical mesh transforms. Indeed, the velocity \dot{r} entering these equations is the velocity relative to the mesh (in analogy to \mathbf{w} above and to \mathbf{c} if taken together with the Jacobian), and not the Eulerian velocity relative to the original set of coordinates (x, y, z, t) . The difference in the form of these equations compared to the ALE equation (8.17) stems from the fact that they are written with respect to the transformed coordinates, while it has not been done above.

The extent the algorithm is Eulerian or Lagrangian depends on how equation (8.21) is treated. Adcroft and Hallberg (2006) introduce the terminology of Eulerian and Lagrangian vertical direction, EVD and LVD respectively. One gets the EVD if the construction of r is fully prescribed, as is the case for the standard sigma-coordinate. In this case

$$r = \sigma = \frac{z - \eta}{H + \eta},$$

where H is the ocean depth and the Jacobian $h = \partial_r z = H + \eta$. As a result, $\partial_t h = \partial_t \eta$ in equation (8.21) is known from the evolution of η , which is solved together with the momentum equation. In this case equation (8.21) is used to diagnose the quasi-vertical transport velocities $h\dot{r}$, beginning from the bottom. More interesting cases follow when $r = f(x, y, \sigma)\sigma$, in which case one may select the function f so that the discrete values of h , denoted by the index k become $h_k = h_{0k}(1 + \eta/H)$, where h_{0k} are the thicknesses of layers in the quiescent ocean, i.e, the Jacobian becomes the actual layer thickness.

One gets the LVD if Equation (8.21) is purely prognostic. An obvious case is that of adiabatic vertical coordinate $r = \rho$ when $\dot{r} = 0$ (no mass flux through the isopycnal surfaces). In this case there is no property exchange between the layers, and advection becomes horizontal. Adcroft and Hallberg (2006) consider the ALE framework as a method associated to the LVD and serving to allow some mass exchange (non-zero \dot{r}) needed in practice to simulate physical vertical mixing. In reality, (i) the ALE framework is more general than LVD or EVD for it contains them as special cases and (ii) it does not enforce one to first integrate in a truly Lagrangian manner and then remap. This is only one interesting possibility. Perhaps,

cases that are sufficiently close to the isopycnal vertical coordinate deserve special attention. The alternative is to split the tendency $\partial_t h$ into a prescribed and adapting parts, as would be more appropriate in models relying on quasi-geopotential or terrain following layers. Indeed, the so-called \tilde{z} vertical coordinate proposed by Leclair and Madec (2011) and also explored by Petersen et al. (2015) follows this direction, without remapping.

Although the approach relying on the transform to the generalized vertical coordinate is popular, the transform can be avoided, for one may directly derive equations for flow in vertically moving layers by integrating the governing equations in the Eulerian representation vertically over these layers. This approach is selected by MPAS (Ringler et al. 2013) and is followed by FESOM2 (Danilov et al. 2017). The appendix of Ringler et al. (2013) presents a succinct derivation. In this case h is the physical layer thickness. The difference from the transformed equations above is that terms like $\partial_r(h\dot{r})$ are replaced by the difference in quasi-vertical transports through the top and bottom surfaces of each layer. This approach is more transparent, and technically easier to introduce in cases where the number of layers is variable (the geopotential basis coordinate). Adding it to existing z -coordinate codes is relatively straightforward.

We thus see that equipping ocean model codes with ALE vertical coordinate is a relatively straightforward task and many terrain-following models routinely have it without explicitly stating. Its availability opens up new possibilities allowing one to use general distributions of layer thicknesses (such as a combination of z - and terrain following layers). Gaining real benefits from the ALE framework as concerns spurious mixing requires a strategy of adapting the layer thicknesses. Indeed, if layer thicknesses were moving isopycnally, spurious water mass transformations would be eliminated altogether. However, depending on applications, physical aspects of the dynamics to be simulated may require other basic vertical coordinate (for example, it is terrain-following in most coastal applications; see Klingbeil et al. 2018a). In this case the question is how to introduce elements of isopycnal behaviour or how to locally refine vertical resolution in order to minimize spurious smoothing. We will present two examples below.

The implementation may raise additional questions such as accurate time stepping and numerical stability, the need in reducing pressure gradient errors, or in adapting parametrisations that might be sensitive to the choice of vertical coordinate. Another important aspect is the correct interpretation of residual dynamics in a particular coordinate system (Klingbeil et al. 2018b).

8.3.1 \tilde{z} -Vertical Coordinate and its Effect on Spurious Mixing

Spurious mixing is significantly reduced in Lagrangian vertical coordinates because of the missing advection through the coordinate surfaces. However, a pure Lagrangian

movement of model layers following arbitrary motions is prone to strong mesh distortion and thus impractical for numerical models. The ALE approach offers to consider Lagrangian tendencies up to a desired level.

A trivial example are separable terrain-following coordinates of the form $z - \eta = S(r)D(x, y, t)$ with the total water depth $D = H + \eta$. These coordinates treat the barotropic portion of motion in a Lagrangian way (Shchepetkin and McWilliams 2009). This can easily be shown by deriving the prognostic equation for h from the transformation function above. Using (8.23) with $P = 0$ and the barotropic velocity (here approximated by the depth-averaged velocity) $\mathbf{U} = \int \mathbf{u} h dr / D$ yields:

$$\partial_t h = (\partial_r S) \partial_t \eta = -(\partial_r S) \nabla_r (D\mathbf{U}) = -\nabla_r ((\partial_r S) D\mathbf{U}) = -\nabla_r (h\mathbf{U}). \quad (8.24)$$

Combination with (8.21), split into a prognostic part for h and a diagnostic part for \dot{r} according to

$$\underbrace{\partial_t h + X}_{=0} + \underbrace{\nabla_r (h\mathbf{u}) - X + \partial_r (h\dot{r})}_{=0} = 0, \quad (8.25)$$

shows that $X = \nabla_r (h\mathbf{U})$ and thus $\dot{r} = 0$ for purely barotropic motions with $\mathbf{u} = \mathbf{U}$.

Recently, Leclair and Madec (2011) proposed to also treat the fast oscillating baroclinic portion of motion in a Lagrangian way by introducing the so-called \tilde{z} -coordinates. These coordinates are not defined by an explicit transformation function, but by a prescribed X in (8.25),

$$X = \frac{h}{D} \nabla_r (D\mathbf{U}) - \partial_t \langle \tilde{h} \rangle_{\text{HF}, \tau_h}, \quad (8.26)$$

with the simple high-pass filter $\partial_t \langle \psi \rangle_{\text{HF}, \tau} = \partial_t \psi - \langle \psi \rangle_{\text{HF}, \tau} / \tau$ and the high-frequency displacements of coordinate surfaces obtained in terms of the high-frequency part of the baroclinic flow divergence $\partial_t \tilde{h} = -\langle \nabla_r (h\mathbf{u}) - h/D \nabla_r (D\mathbf{U}) \rangle_{\text{HF}, \tau_h}$. In order to avoid grid distortion due to a long-term drift, the prognostic equation for the high-frequency displacements contains a term restoring the displacements back to zero, which is equivalent to the secondary high-pass filter $\langle \tilde{h} \rangle_{\text{HF}, \tau_h}$. Additionally, the optional inclusion of a lateral thickness diffusion term was suggested to suppress grid-scale noise.

The performance of \tilde{z} -coordinates was compared against that of z -level, z^* , σ and isopycnal coordinates by Petersen et al. (2015), who repeated the idealised experiments of Ilicak et al. (2012) and confirmed reduced diapycnal mixing. Furthermore, for a global spin-up simulation significantly reduced vertical cross-coordinate transports were demonstrated. Despite the advantages of \tilde{z} -coordinates in regimes dominated by (internal) waves, Petersen et al. (2015) also experienced stability issues with \tilde{z} coordinates that need to be investigated. In addition, the definition of X in (8.26) needs further research, because only $X = \nabla_r (h\mathbf{U}) + \nabla_r (h \langle \mathbf{u} - \mathbf{U} \rangle_{\text{HF}})$ treats barotropic and high-frequency baroclinic motions with $\mathbf{u} = \mathbf{U} + \langle \mathbf{u} - \mathbf{U} \rangle_{\text{HF}}$ in a Lagrangian way.

8.3.2 Additional Techniques for Adaptive Vertical Model Layers

In addition to adjustable Lagrangian tendencies, the adaptive coordinates presented by Hofmeister et al. (2010) also offer adjustable isopycnal tendencies. From a truncated Taylor series for a prescribed target density $\hat{\rho} = \rho(\hat{z})$, the corresponding target position of the layer is estimated as

$$\hat{z} = z + (\partial_z \rho)^{-1} (\hat{\rho} - \rho(z)). \quad (8.27)$$

The tendency to move the layer to this target position can be adjusted and determines the isopycnal tendency. Clipping towards prescribed positive minimum layer thicknesses avoids invalid meshes. Furthermore, a final stretching of all layers is required to guarantee that the sum of layer heights coincides with the total water depth. The resulting isopycnal alignment of layers reduces the spurious diapycnal mixing associated with lateral advection.

In order to reduce the spurious mixing caused by truncation errors associated with vertical advection, Burchard and Beckers (2004) developed an optimisation technique for the vertical resolution, i.e. the layer heights within a water column. The desired zooming of layers towards boundaries, stratification and shear is based on a vertical diffusion equation for the interface positions $z_{\kappa=k\pm 1/2}$ with $\kappa \in \mathbb{R}$ and $k \in [1, k_{\max}]$ denoting the layer number,

$$\partial_t z - \partial_\kappa (D^{\text{grid}} \partial_\kappa z) = 0, \quad (8.28)$$

with boundary conditions $z_{\kappa=1/2} = -H$ and $z_{\kappa=k_{\max}+1/2} = \eta$. Diffusion coefficients D^{grid} that are uniform in the vertical cause the layer heights to become uniform as well, i.e. a tendency towards equidistant layers. In contrast, a locally increased diffusion coefficient, e.g. proportional to stratification or shear, tends to decrease the local layer height, i.e. refines the vertical resolution in the vicinity of gradients. The resulting layer distribution in adjacent water columns can be homogenised by lateral diffusion of layer heights and interface positions (Hofmeister et al. 2010), which must be combined with a subsequent stretching of layer heights as described above.

Figure 8.1 shows the reduced spurious mixing for adaptive coordinates compared to σ -coordinates in an exemplary realistic coastal ocean model simulation.

8.4 Advection Algorithms Stabilized with Isonutral Mixing

Many popular high-order transport schemes rely on upwind-biased reconstructions in order to warrant numerical stability. Along with stability, upwinding also introduces spurious numerical mixing. However, if one assumes that the velocity field is

uniform, the dissipative part of such transport schemes can be explicitly identified. For example, the widely used 3rd-order algorithm (see, e.g., Webb et al. 1998; Shchepetkin and McWilliams 1998 for structured meshes and Skamarock and Gassmann 2011; Abalakin et al. 2002 for unstructured meshes), forming the basis of many advection schemes, can be written as a sum of the fourth-order centered method and a dissipative correction, which reduces the order to the third (see below). This correction resembles biharmonic diffusion with a diffusivity coefficient that is proportional to the amplitude of velocity and cube of mesh element size. The centered estimate of fluxes introduces a residual error of dispersive nature (scaling with odd derivative). It does not lead to mixing on its own, for the residual terms with odd derivatives do not contribute to domain-mean variance decay. Yet it cannot work without the dissipative part, since it is prone to instabilities caused by small-scale dispersive errors.

This consideration underlies the analysis of Lemarié et al. (2012a) who propose to use this explicit splitting into high-order centered part and velocity-dependent dissipative correction to reduce spurious mixing accompanying numerical advection. It will be clarified below based on 1D example. Let discrete values of tracer be T_i at location indexed by the index i . Let the velocity u be positive (in the direction of i increasing). The upwind gradient-based reconstruction leading to the third-order scheme is in this case

$$T_{i+1/2}^- = T_i + (1/2)\Delta x G_{i+1/2}^-,$$

where the gradient G is the combination of central and upwind estimates

$$G_{i+1/2}^- = (2/3)G_{i+1/2}^c + (1/3)G_{i+1/2}^u,$$

with $G_{i+1/2}^c = (T_{i+1} - T_i)/\Delta x$, $G_{i+1/2}^u = (T_i - T_{i-1})/\Delta x$, and Δx the mesh horizontal increment. Computing the flux leaving cell i at location $i + 1/2$ as $F_{i+1/2}^{(3)} = uT_{i+1/2}^-$ gives the third order advection scheme. Introducing the reconstruction for the opposite (downwind) direction,

$$T_{i+1/2}^+ = T_{i+1} - (1/2)\Delta x G_{i+1/2}^+,$$

with

$$G_{i+1/2}^+ = (2/3)G_{i+1/2}^c + (1/3)G_{i+1/2}^d,$$

where $G_{i+1/2}^d = (T_{i+2} - T_{i+1})/\Delta x$, and combining it with the previous one gives the centered flux estimate $F_{i+1/2}^{(4)} = u(T_{i+1/2}^+ + T_{i+1/2}^-)/2$ and the method of the fourth order which has a dispersive leading error and needs to be augmented with diffusion in practical applications. As a rule, the third-order method is too dissipative but its combination with the fourth-order method $F = \gamma F^{(4)} + (1 - \gamma)F^{(3)}$ with $\gamma = 0.75$ or higher works invariably well in practice (cf. Skamarock and Gassmann 2011). Obviously the difference of two estimates,

$$F_{i+1/2,diss} = F_{i+1/2}^{(3)} - F_{i+1/2}^{(4)}$$

corresponds to a dissipative diffusive contribution stabilizing the method. It can be easily shown that it leads to the residual term $\partial_x(|\mathbf{u}|\Delta x^3/12)\partial_{xxx}T$ in the flux divergence, which is biharmonic diffusion (cf. Lemarié et al. 2012a). It will enter with the additional factor $(1 - \gamma)$ for the combined flux estimate. In the case of three dimensions the flux \mathbf{F} becomes a three-dimensional vector and an analogous dissipative term will appear. It will cause mixing which will necessarily have a cross-isopycnal (diapycnal) component. One can introduce an isoneutral projection operator \mathbf{K} which eliminates the diapycnal component of the flux. Introducing the slope vector

$$\mathbf{s} = -\nabla_h \rho / \partial_z \rho, \tag{8.29}$$

where the index h implies quasi-horizontal (along layers) directions and an agreement is made that ρ should be referenced to local pressure in estimation of derivatives, in the small-slope approximation the operator reads

$$\mathbf{K} = \begin{pmatrix} 1 & 0 & s_x \\ 0 & 1 & s_y \\ s_x & s_y & s_x^2 + s_y^2 \end{pmatrix}.$$

Indeed, the flux $\mathbf{K}\mathbf{F}_{diss}$ has no projection on the diapycnal vector $(-s_x, -s_y, 1)$. However, because of \mathbf{K} , it also ceases to be purely dissipative in a general case and may limit the time step if the parameter $S = |\mathbf{s}|\Delta x/\Delta z$ exceeds one, with Δz the vertical mesh increment. It remains to see whether this simple approach is useful.

Lemarié et al. (2012a) propose to substitute the dissipative term by the appropriately scaled modified biharmonic diffusion operator. They write this operator as a combination

$$\mathcal{D}_4 T = \mathcal{D}_2 \mathcal{D}_2 T,$$

where

$$\mathcal{D}_2 T = \nabla \kappa \mathbf{K} \nabla T,$$

and the diffusivity κ is computed as $\kappa = (|\mathbf{u}|\Delta x^3/12)^{1/2}$ (the attenuation factor $(1 - \gamma)$ may be added in general case). It also needs stabilization carried out with a variant of the method of stabilizing corrections (MSC). Applying stabilization requires care, because the projection operator \mathbf{K} appears twice, effectively squaring the contributions from mixed (vertical-horizontal) derivatives which cause the problem. Although the use of two isoneutral harmonic operators implies certain code simplification, computations involving \mathbf{K} are relatively expensive.

Two remarks are due here. First, functioning of the proposed approach on unstructured triangular meshes may pose additional challenges with respect of numerical stability, leaving aside the aspects of a consistent implementation of the isoneutral projection operator (see the discussion in Lemarié et al. 2012b).

Second, the algorithm presents a special interest for meshes with generalized level surfaces that follow the bottom topography at some locations. They may deviate stronger from isopycnals than the geopotential surfaces do, and spurious mixing will

be increased there. Many practical tasks involving interactions of ocean shelf with the deep ocean would benefit from using such locally terrain following meshes. In this case the slope entering the expression for \mathbf{K} is with respect to the level surfaces if horizontal derivatives are computed along the level surfaces. As should be clear from foregoing the strongest time step limitations will be in the places where the product of mesh aspect ratio with slope is large, $S = |\mathbf{s}| \Delta x / \Delta z > 1$, which can easily happen. The implication is that practical utility of the algorithm proposed is inseparable from the issues related to mesh geometry and should be studied together. This links the question considered in this section to the implementation of ALE vertical coordinate.

While the measures based on the suggestion of Lemarié et al. (2012a) are valuable, transport schemes may include limiters, in which case it is impractical to explicitly write the operator responsible for dissipation. A question arising in the context of present discussion is whether the accompanying spurious dissipation can be minimized or controlled. The answer should depend on the nature of the limiter used. For example, the flux corrected transport (FCT) algorithm, which is based on comparing a low-order, but monotonic, solution with the high-order, but non-monotone solution, limits antidiffusive fluxes which are the difference between the high-order and low-order fluxes. The difference between the limited and full antidiffusive fluxes should be analogous to the diffusive flux considered above, and one may try to reduce its diapycnal component (although this may interfere with monotonicity). The ideas like this one are purely hypothetical at present, but research in this direction is worthwhile.

The extent to which these measures lead to the reduction in spurious mixing needs to be quantified in realistic applications. The change in advection scheme also implies some change in dynamics, which may influence spurious effects. Furthermore, spurious mixing may interfere with physical mixing (the latter may be triggered by small-scale dispersive errors which in theory do not contribute to mixing on their own). The task is therefore, alongside with mathematical developments, to learn about accompanying practical details. This presents a major task for future research.

8.5 ADER High Order Flux Evaluation and WENO Reconstruction

In this section, we explain how to combine ADER high order flux evaluation with WENO reconstruction. The resulting ADER-WENO method yields a high performance method for the numerical simulation of fluid flow of arbitrary high order (see, e.g., Dumbser and Käser 2007). The ADER scheme of the seminal work by Toro et al. (2001) can be viewed as a generalisation of the classical first-order Godunov scheme to arbitrary high orders. In this generalisation, the ADER method provides the numerical solution of the generalized Riemann problem (GRP), as discussed in Section 8.5.1. Then, in Section 8.5.2, we explain how to design WENO (Weighted Essentially Non-Oscillatory) reconstructions for the recovery step of finite volume methods, where particular emphasis is placed on kernel-based reconstructions for unstructured data, as proposed in Aboiyar et al. (2010).

8.5.1 The Generalized Riemann Problem

The (classical) Riemann problem is the initial value problem for an evolutionary PDE with piecewise constant but discontinuous initial data. Interest in such problems originates from the study of *shock-tube problems* in gas dynamics, where a high pressure part of a gas in a tube is separated from a low pressure part by a membrane. When this membrane is removed very quickly, high and low pressure parts interact instantaneously and several interesting wave phenomena can occur. A similar problem arises in hydrodynamics when we consider a *dam-break problem* where large water heights interact with low water heights after the instantaneous removal of an obstacle. For a general conservation law in 1D, the Riemann problem has the form

$$\frac{\partial \mathbf{u}(x, t)}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u}(x, t))}{\partial x} = 0. \quad (x, t) \in \mathbb{R} \times (0, \infty), \quad (8.30)$$

$$\mathbf{u}(x, 0) = \begin{cases} \bar{\mathbf{u}}_L, & x < 0, \\ \bar{\mathbf{u}}_R, & x > 0, \end{cases} \quad (8.31)$$

with constant states $\bar{\mathbf{u}}_L, \bar{\mathbf{u}}_R$.

The Riemann problem is an important building block, not only for modelling and computing, but also in the analytical treatment of conservation laws. Modern results concerning existence and uniqueness of solutions and the continuous dependence on the initial data can be achieved by carefully piecing together local Riemann problems, see Bressan (2000). The analytical solution of Riemann problems, together with several examples, is discussed, e.g., in the textbooks by Dafermos (2010) and Smoller (1994).

From a computational point of view, the Riemann problem naturally occurs in Finite Volume schemes: When the discrete solution of the PDE is represented by its cell-averages, our interpretation is that we have a piecewise constant function that is constant in each computation cell and has jumps at the boundaries of the cells. In this sense, a Finite Volume Method (approximately) solves Riemann problems at each time-step at each cell boundary. For a survey of numerical techniques for solving the Riemann problem numerically, we refer to Toro (2009, 2016)

For *hyperbolic conservation laws*, the Riemann problem has been studied thoroughly and for many systems of practical interest it can be either solved exactly or robust, efficient and accurate methods for approximately solving the Riemann problem are available. Thus, we can use these Riemann solvers directly to compute numerical fluxes in our Finite Volume update. This idea dates back to Godunov (1959) and the method that employs the exact Riemann solver has become famous as the Godunov scheme.

High-order accurate one-step finite volume methods can then be constructed by allowing spatial variation of the discrete solution in each cell. Instead of a piecewise constant representation of the discrete solution, we seek a discrete representation such that in each cell we have a smooth function (say, a polynomial of a prescribed

degree), but in order to deal with shock waves we still allow for discontinuities at the cell boundaries. In order to construct a method that is of *generalized Godunov type*, we then want to solve the *generalized Riemann problem (GRP)*, that is, the initial value problem with piecewise smooth, but discontinuous data. In fact, any high order method that represents the discrete solution by a piecewise polynomial, be it a WENO reconstruction or a polynomial from a discontinuous Galerkin finite element space, in some way or the other has to deal with the generalized Riemann problem.

A generalized Godunov method can be understood in the REA-framework (Reconstruct—Evolve—Average) described by LeVeque (2002). We need three components (compare Figure 8.4):

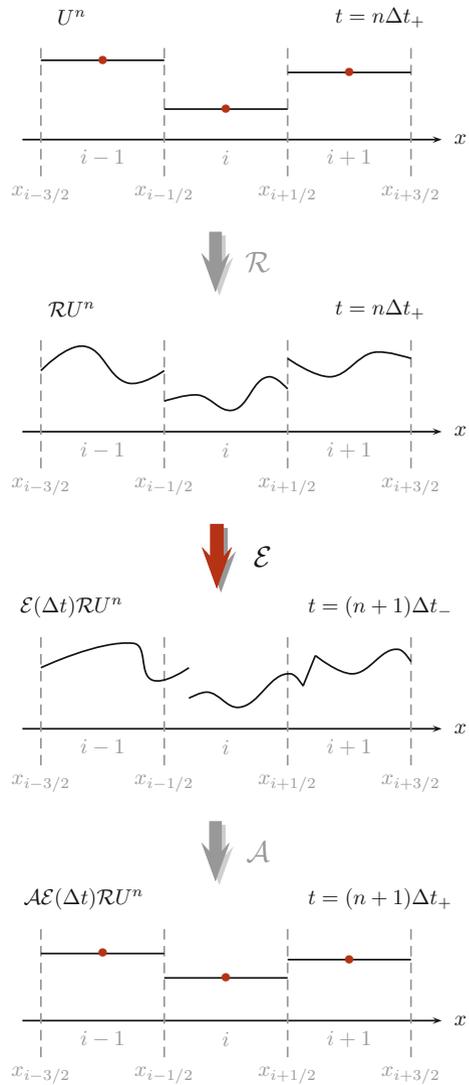
- \mathcal{R} : A conservative reconstruction operator \mathcal{R} . That is, a method that computes a higher order accurate representation \mathbf{w} from cell averages, such that the integral of \mathbf{w} in each cell is the cell average in that cell. How this can be done with WENO reconstruction is described in the next section.
- \mathcal{E} : An approximate evolution operator, that describes the exact (or at least, well approximated) evolution of the initial value problem given by the data from the reconstruction step. In this section, we focus on the construction of approximate evolution operators by solving the generalized Riemann problem.
- \mathcal{A} : Computation of cell averages of the evolved data. In fact, evolution and averaging can be done in one step if we use a conservative finite volume method.

Generally speaking, there are two strategies for solving the GRP in a high order Godunov type scheme: The first is the *evolution in the small* approach, in which the data is evolved in each cell and then Riemann problems at times $t > 0$ are solved to compute the interaction between different cells. The second approach is the *instantaneous interaction* variant, in which we first compute an interaction between neighbouring cells at $t = 0$ and then do the evolution after that.

Our focus here lies on instantaneous interaction solvers, but let us briefly discuss the first (evolution in the small) ansatz. Early methods following this line of thought are van Leer's MUSCL scheme (van Leer 1979), the PPM method of Colella and Woodward (1984), and the Harten-Engquist-Osher-Chakravarthy method originally proposed in the context of ENO schemes (Harten et al. 1987). A recent variant of this approach is the use *local space-time discontinuous Galerkin predictors*, put forward by Balsara et al. (2013), Dumbser et al. (2008a, b), in which a weak formulation of the PDE is used for the evolution of the data in each cell. A similar space-time expansion approach is due to Gassner et al. (2011, 2008), Lörcher et al. (2007). This modern interpretation of ADER schemes currently is a very popular research topic and the approach has been applied successfully to a broad range of complex flow problems, see for example Dumbser et al. (2016), Zanotti and Dumbser (2016), Zanotti et al. (2015) and references therein. For an application of the ADER approach in a Lagrangian or arbitrary Eulerian-Lagrangian framework see Boscheri et al. (2014a), Boscheri and Dumbser (2013, 2014), Boscheri et al. (2014b).

Now let us get back to instantaneous interaction methods, in which we use a direct solution strategy for the GRP as a building block in the numerical scheme. The

Fig. 8.4 Schematic description of a generalized Godunov scheme



solution of the generalized Riemann problem is generally not available in a closed, analytic form. We therefore need to construct approximate solutions, typically in the form of space-time polynomial expansions. The idea of turning a direct solution strategy for the GRP into a numerical method was introduced by Ben-Artzi and Falcovitz (1984, 2003) for piecewise linear initial data and consequentially this method has become known as the GRP method. We will, however, use the term more broadly, allowing any kind of piecewise smooth initial data. Convergence of the GRP method for scalar problems was proven in Ben-Artzi et al. (2009), recent versions of the GRP scheme are presented in Ben-Artzi et al. (2006) or Yang et al. (2011).

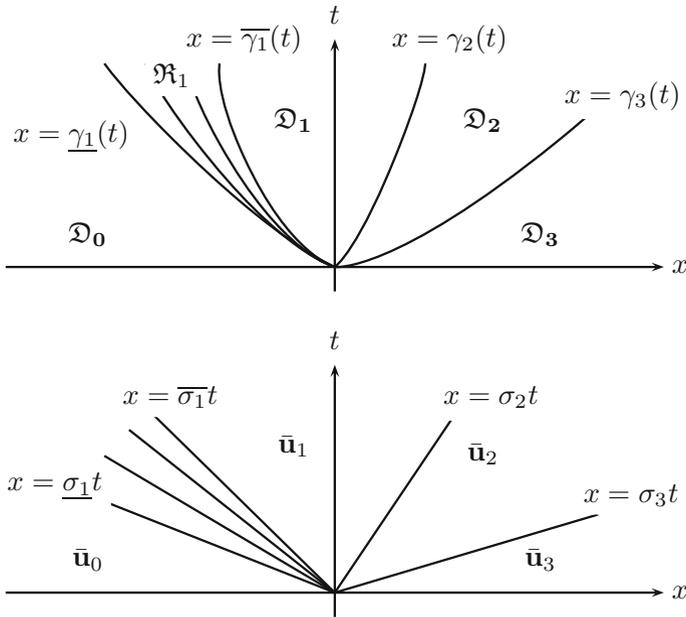


Fig. 8.5 Corresponding wave pattern for a generalized Riemann problem and the associated classical Riemann problem for a 3×3 system

Different techniques for solving the GRP numerically are available. They are usually rooted in a deep analytical examination of the problem. For example, the treatment of the GRP in terms of Riemann invariants by Ben-Artzi and Li (2007) was used to construct a third-order solver by Qian et al. (2014). Goetz and Dumbser (2016) constructed a GRP solver motivated by the analysis via asymptotic series expansion carried out by Bourgeade et al. (1989), LeFloch and Raviart (1988). Castro and Toro (2008) and Montecinos et al. (2012) ran extensive numerical tests, comparing different solvers for the GRP. A novel and conceptually different GRP solver is the recent *implicit* solver of Toro and Montecinos (2015).

The fundamental observation that allows us to construct approximate GRP solutions is the fact that the GRP inherits its wave structure from the so-called *associated classical Riemann problem*, see Figures 8.5 and 8.6. We describe this connection more precisely below, but at first let us recall a few well-known facts about classical Riemann problems. Consider the Riemann problem for an $m \times m$ system. We give a very short outline of what the solution of a Riemann problem looks like and refer to Dafermos (2010), Smoller (1994) for a detailed discussion. Under some mild regularity assumptions (see Lax 1957), the solution of the classical Riemann problem consists of $m + 1$ states separated by m waves.

$$\bar{\mathbf{u}}_L = \bar{\mathbf{u}}_0, \bar{\mathbf{u}}_1, \dots, \bar{\mathbf{u}}_m = \bar{\mathbf{u}}_R.$$

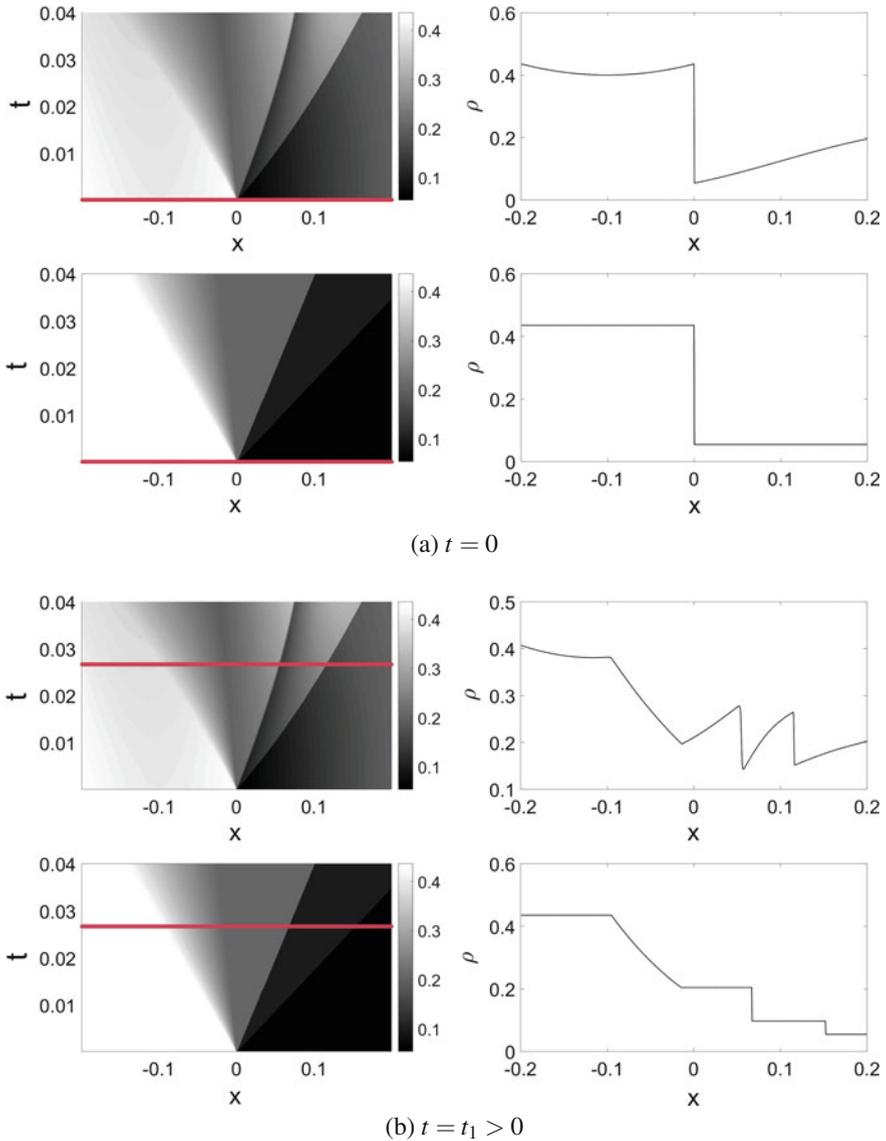


Fig. 8.6 Solution of a generalized Riemann problem for a 1D shock tube problem together with the associated classical Riemann problem

These constant states are separated by different kinds of waves: *Rarefaction waves*, which describe a continuous transition between two states; and *shock waves*, or *contact discontinuities*, which correspond to jumps in the solution.

An i -contact discontinuity between states $\bar{\mathbf{u}}_{i-1}$, $\bar{\mathbf{u}}_i$ travels with speed given by the i th eigenvalue λ_i of the Jacobian $\mathbf{A}(\bar{\mathbf{u}}_i) = \mathbf{Df}(\bar{\mathbf{u}}_i)$:

$$\sigma_i = \lambda_i(\bar{\mathbf{u}}_{i-1}) = \lambda_i(\bar{\mathbf{u}}_i).$$

In the case of an i -shock wave, $\bar{\mathbf{u}}_{i-1}$ and $\bar{\mathbf{u}}_i$ are connected by a jump discontinuity with speed σ_i , such that the *Lax entropy condition* holds:

$$\lambda_i(\bar{\mathbf{u}}_{i-1}) \geq \sigma_i \geq \lambda_i(\bar{\mathbf{u}}_i).$$

An i -rarefaction wave describes the continuous transition from $\bar{\mathbf{u}}_{i-1}$ to $\bar{\mathbf{u}}_i$. This transition takes place over a zone whose boundaries travel with speeds $\underline{\sigma}_i < \bar{\sigma}_i$. These speeds are given by

$$\underline{\sigma}_i = \lambda_i(\bar{\mathbf{u}}_{i-1}), \quad \bar{\sigma}_i = \lambda_i(\bar{\mathbf{u}}_i).$$

If the i -wave in the solution is a jump discontinuity, let us denote $\underline{\sigma}_i = \bar{\sigma}_i = \sigma_i$. For convenience, let us also write $\bar{\sigma}_0 = -\infty$, $\underline{\sigma}_{m+1} = +\infty$. The the solution of the Riemann problem satisfies

$$\mathbf{u}(x, t) = \bar{\mathbf{u}}_i^0 \quad \text{for } \bar{\sigma}_i < \frac{x}{t} < \underline{\sigma}_{i+1}.$$

In particular, there exist an i_* such that $\bar{\sigma}_{i_*} < 0$ and $\underline{\sigma}_{i_*+1} > 0$, which means that the domain

$$\mathcal{D}_{i_*}^0 = \left\{ (x, t) \in \mathbb{R} \times [0, \infty) \mid \bar{\sigma}_{i_*} < \frac{x}{t} < \underline{\sigma}_{i_*+1} \right\}$$

contains the line $\{x = 0\} \times [0, \infty)$. We call the corresponding state $\bar{\mathbf{u}}_{i_*}$ the *Godunov state* of the Riemann problem and denote it simply by $\bar{\mathbf{u}}_*$.

For the generalized Riemann problem, assume that the initial data

$$\mathbf{u}(x, 0) = \begin{cases} \mathbf{u}_L(x), & \text{if } x < 0, \\ \mathbf{u}_R(x), & \text{if } x > 0, \end{cases} \tag{8.32}$$

with $\mathbf{u}_L, \mathbf{u}_R : \mathbb{R} \rightarrow \Omega$, is piecewise smooth but discontinuous at $x = 0$.

It can be shown that (8.30), (8.32) has a unique entropy admissible weak solution in a neighbourhood of the origin, provided that the initial states $\bar{\mathbf{u}}_L^0 = \mathbf{u}_L(0)$ and $\bar{\mathbf{u}}_R^0 = \mathbf{u}_R(0)$ are sufficiently close, see Li and Yu (1985), Tatsien and Libin (2009).

Similar to the wave pattern for the classical Riemann problem, we can give a geometric description of the solution in the $x - t$ -plane: For sufficiently small $T > 0$, the strip $\mathbb{R} \times [0, T)$ can be decomposed into $m + 1$ open *domains of smoothness* \mathcal{D}_i , $0 \leq i \leq m$. These domains are separated by smooth curves $\gamma_j(t)$ passing through the origin, or by rarefaction zones with boundaries $\underline{\gamma}_j(t)$, $\bar{\gamma}_j(t)$, where $\underline{\gamma}_j(t)$, $\bar{\gamma}_j(t)$ are smooth characteristic curves passing through the origin. More precisely: We have curves $\gamma_j(t)$ and *rarefaction zones*

$$\mathfrak{R}_j = \left\{ (x, t) \in \mathbb{R} \times [0, T) \mid \underline{\gamma}_j(t) < x < \overline{\gamma}_j(t) \right\}.$$

For $\gamma_j(t)$, we let $\underline{\gamma}_j(t) = \overline{\gamma}_j(t) = \gamma_j(t)$ for all $t \in [0, T)$. Then, we can write

$$\begin{aligned} \mathfrak{D}_0 &= \left\{ (x, t) \in \mathbb{R} \times [0, T) \mid x < \underline{\gamma}_1(t) \right\}, & \mathfrak{D}_m &= \left\{ (x, t) \in \mathbb{R} \times [0, T) \mid \overline{\gamma}_m(t) < x \right\}, \\ \mathfrak{D}_i &= \left\{ (x, t) \in \mathbb{R} \times [0, T) \mid \overline{\gamma}_i(t) < x < \underline{\gamma}_{i+1}(t) \right\}, & 1 \leq i \leq m-1. \end{aligned}$$

The solution \mathbf{u} is smooth inside each domain \mathfrak{D}_i and inside each rarefaction zone \mathfrak{R}_j . Moreover, \mathbf{u} has a shock or contact discontinuity across each curve $x = \gamma_j(t)$ and is continuous across the characteristic curves $x = \underline{\gamma}_j(t)$, $x = \overline{\gamma}_j(t)$.

Locally we can think of the GRP as perturbation of a classical Riemann problem. It is therefore natural that the solution of the generalized Riemann problem and has a wave structure determined by the solution of the *associated classical Riemann problem* with the initial states $\mathbf{u}_L^0 = \mathbf{u}_L(0)$ and $\mathbf{u}_R^0 = \mathbf{u}_R(0)$. For example, if the j -wave in the solution of the associated Riemann problem is a shock wave, the corresponding j -wave in the generalized Riemann problem is also a shock wave. The same hold for contact discontinuities and rarefaction waves. A typical configuration of corresponding wave patterns for a 3×3 system is shown in Figure 8.5.

We denote the constant states in the solution of the associated Riemann problem with initial data \mathbf{u}_L^0 , \mathbf{u}_R^0 by \mathbf{u}_i^0 , for $i = 0, \dots, m$, and the wave speeds by $\underline{\sigma}_j^0$, $\overline{\sigma}_j^0$, $j = 1, \dots, m$. Here we have set $\underline{\sigma}_j^0 = \overline{\sigma}_j^0 = \sigma_j^0$ if the j -wave is a shock or contact discontinuity. Then the slope of the curves $\underline{\gamma}_j$, $\overline{\gamma}_j$ for $t \rightarrow 0$ is given by the corresponding wave speed in the associated Riemann problem. That is, for $j = 1, \dots, m$ we have

$$\underline{\gamma}_j(0) = 0, \quad \overline{\gamma}_j(0) = 0 \quad \text{and} \quad \lim_{t \rightarrow 0} \dot{\underline{\gamma}}_j(t) = \underline{\sigma}_j^0, \quad \lim_{t \rightarrow 0} \dot{\overline{\gamma}}_j(t) = \overline{\sigma}_j^0.$$

Inside each domain of smoothness \mathfrak{D}_i , the solution \mathbf{u} of the generalized Riemann problem is given in the limit towards the origin by the corresponding state in the associated Riemann problem:

$$\lim_{\substack{(x,t) \rightarrow (0,0) \\ (x,t) \in \mathfrak{D}_i}} \mathbf{u}(x, t) = \mathbf{u}_i^0 \quad \text{for } i = 0, \dots, m.$$

With this knowledge about the structure of GRP solution is mind, let us now consider techniques for constructing approximate GRP solutions. Our presentation of the topic is intended to be a rather broad introduction, so we will only discuss the original solver of Titarev and Toro (2002), Toro et al. (2001), Toro and Titarev (2001, 2002, 2006) to highlight some basic concepts. Approximate GRP solver do involve large amount of technical work, so to keep the presentation of the material concise and accessible, we will only discuss the 1D case. With this in mind, the initial value problem we are interested in is

$$\frac{\partial \mathbf{u}(x, t)}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{u}(x, t))}{\partial x} = 0. \quad (x, t) \in \mathbb{R} \times (t^n, \infty), \quad (8.33)$$

$$\mathbf{u}(x, t^n) = \begin{cases} \mathbf{w}_i(x), & x < x_{i+1/2}, \\ \mathbf{w}_{i+1}(x), & x > x_{i+1/2}, \end{cases} \quad (8.34)$$

for each cell boundary $x_{i+1/2}$ and each time step t^n . To use this as a building block of a high order method, we need three components: First, a *conservative reconstruction method*, that computes a high order accurate piecewise polynomial representation \mathbf{w}_i for the given cell averages \mathbf{U}_i^n . How this can be done with WENO reconstruction is explained in the next second. Second, we need a GRP solver. That is a computational method that finds an approximate solution $\mathcal{U}(x, t)$ of problem (8.33), (8.34). Finally, we need a way to compute a numerical flux from the approximate solution:

$$\mathbf{f}_{i+1/2} \approx \frac{1}{\Delta t} \int_0^{\Delta t} \mathbf{f}(\mathcal{U}(x_{i+1/2}, t)) dt.$$

Assume that we can evaluate the approximate solution $\mathcal{U}(x_{i+1/2}, t)$ for all times $0 < t < \Delta t$. Then one simple option for computing the numerical flux is to replace the time integral by some Gaussian quadrature, i.e.,

$$\mathbf{f}_{i+1/2} = \frac{1}{\Delta t} \sum_{p=1}^G \omega_p \mathbf{f}(\mathcal{U}(x_{i+1/2}, t_p)),$$

where ω_p, t_p are Gaussian weights and nodes and the number G of nodes is chosen according to the desired accuracy.

Let us assume for simplicity that $x_{i+1/2} = 0$ and $t^n = 0$ and denote $\mathbf{u}_L = \mathbf{w}_i, \mathbf{u}_R = \mathbf{w}_{i+1}$. The most direct approach is to take $\mathcal{U}(x, t)$ as a space-time Taylor approximation of the exact solution \mathbf{u} . Denote by $D^{(\ell,k)}$ the differential

$$D^{(\ell,k)} \mathbf{u}(x, t) = \frac{\partial^\ell}{\partial x^\ell} \frac{\partial^{k-\ell}}{\partial t^{k-\ell}} \mathbf{u}(x, t), \quad (\ell, k) \in \mathbb{N}_0 \times \mathbb{N}_0, \quad k \geq \ell.$$

We choose a $N > 0$ (typically, the degree of the reconstruction polynomial \mathbf{w}) and approximate \mathbf{u} by

$$\begin{aligned} \mathbf{u}(x, t) &\approx \sum_{k=0}^N \sum_{\ell=1}^k D^{(\ell,k)} \mathbf{u}(0, 0^+) \frac{x^\ell t^{k-\ell}}{\ell!(k-\ell)!} \\ &= \mathbf{u}(0, 0^+) + \sum_{k=0}^N t^k \sum_{\ell=0}^k \frac{D^{(\ell,k)} \mathbf{u}(0, 0^+)}{\ell!(k-\ell)!} \left(\frac{x}{t}\right)^\ell. \end{aligned}$$

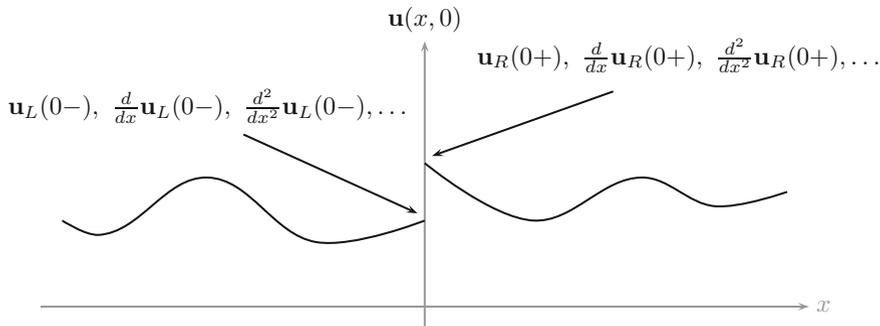


Fig. 8.7 Spatial derivatives at $t = 0$ are known from the initial data

where $D^{(\ell,k)}\mathbf{u}(0, 0^+)$ denotes the limit

$$D^{(\ell,k)}\mathbf{u}(0, 0^+) = \lim_{\substack{(x,t) \rightarrow (0,0) \\ (x,t) \in \mathcal{D}_i}} D^{(\ell,k)}\mathbf{u}(x, t).$$

Clearly, in order to compute this expansion, we need time and mixed space-time derivatives of the solution at times $t > 0$. What we have is information about spatial derivatives at $t = 0$, compare Figure 8.7.

The Cauchy-Kovalevskaya theorem is a classical result in the theory of PDEs. Roughly speaking, applied in our context, it states that when the flux \mathbf{f} and the initial data $\mathbf{u}(x, 0)$ are analytic at the point $(0, 0)$ then the solution \mathbf{u} of (8.30) is analytic in a neighbourhood of the origin. Recall that a (real) analytic function is a function that can be expanded locally into a convergent power series.

This does imply that time derivatives and mixed space-time derivatives can be computed from spatial derivatives up to the same order, which is called the *Cauchy-Kovalevskaya procedure*. It was shown by Harabetian (1986) that for a generalized Riemann problem with piecewise analytic data a local Cauchy-Kovalevskaya procedure can be used inside each domain of smoothness. For $r \geq s \geq 0$ we can construct a function $\mathcal{E}^{r,s}$ such that

$$\frac{\partial^r \mathbf{u}}{\partial t^{r-s} \partial x^s} = \mathcal{E}^{r,s} \left(\mathbf{u}, \frac{\partial \mathbf{u}}{\partial x}, \dots, \frac{\partial^r \mathbf{u}}{\partial x^r} \right), \quad \mathcal{E}^{0,0}(\mathbf{u}) = \mathbf{u}.$$

Let us write \mathcal{E}^r for $\mathcal{E}^{r,0}$, i.e. for the map that gives pure time derivatives from spatial derivatives. In particular, we want to compute the evolution of \mathbf{u} at the cell interface $x = 0$ during one time step, so we are interested in

$$\begin{aligned} \mathbf{u}(0, t) &\approx \mathbf{u}(0, 0^+) + \sum_{k=1}^N \frac{\partial^k \mathbf{u}(0, 0^+)}{\partial t^k} \frac{t^k}{k!} \\ &= \mathbf{u}(0, 0^+) + \sum_{k=1}^N \mathcal{E}^k \left(\mathbf{u}, \frac{\partial \mathbf{u}}{\partial x}, \dots, \frac{\partial^k \mathbf{u}}{\partial x^k} \right) (0, 0^+) \frac{t^k}{k!}. \end{aligned}$$

Thus, if we can give a meaning to the *spatial derivatives* $\frac{\partial \mathbf{u}}{\partial x}(0, 0^+), \dots, \frac{\partial^N \mathbf{u}}{\partial x^N}(0, 0^+)$, we can construct the *time expansion*. The Cauchy-Kowalewskaya procedure, however, can be computationally challenging and is indeed tied to rather cumbersome symbolic manipulations. As an alternative, it was suggested by Goetz and Dumbser (2016) to use a local space-time discontinuous Galerkin ansatz similar to the *evolution in the small* method of Dumbser et al. (2008b). The local space-time DG approach also requires only knowledge about spatial derivatives as initial data.

One popular strategy for computing spatial derivatives at the origin is the *Toro-Titarev solver*, see Titarev and Toro (2002), Toro et al. (2001) and Toro and Titarev (2001, 2002, 2006). The basic idea is to compute an approximate solution of the GRP via a sequence of classical Riemann problems.

At first, in the Toro-Titarev solver we solve the (nonlinear) associated Riemann problem to find the Godunov state $\mathbf{u}(0, 0^+) = \bar{\mathbf{u}}_*^0$. Next, denote by

$$\mathbf{u}^k(x, t) = \frac{\partial^k}{\partial x^k} \mathbf{u}(x, t)$$

the k th spatial derivative of the solution. Wherever the solution \mathbf{u} is smooth, it is straightforward to show that all \mathbf{u}^k satisfy a semilinear hyperbolic PDE of the form

$$\frac{\partial}{\partial t} \mathbf{u}^k(x, t) + \mathbf{A}(\mathbf{u}(x, t)) \frac{\partial}{\partial x} \mathbf{u}^k(x, t) = \mathbf{H}^k(\mathbf{u}(x, t), \mathbf{u}^1(x, t), \dots, \mathbf{u}^k(x, t)). \tag{8.35}$$

Denote the the *one-sided derivatives* of the initial data at the cell interface by

$$\bar{\mathbf{u}}_L^k = \lim_{x \rightarrow 0^-} \frac{d^k}{dx^k} \mathbf{u}_L(x), \quad \bar{\mathbf{u}}_R^k = \lim_{x \rightarrow 0^+} \frac{d^k}{dx^k} \mathbf{u}_R(x), \quad k = 1, \dots, N. \tag{8.36}$$

Toro and Titarev suggested to use equation (8.35) with Riemann data provided by (8.36) to determine the values $\frac{\partial^k \mathbf{u}}{\partial x^k}(0, 0^+)$. Moreover, they linearised (8.35) around the Godunov state $\bar{\mathbf{u}}_*^0$ and neglected the source terms \mathbf{H}^k . In this approach, spatial derivatives are found by solving the classical, linear Riemann problem

$$\frac{\partial}{\partial t} \mathbf{u}^k(x, t) + \mathbf{A}(\bar{\mathbf{u}}_*^0) \frac{\partial}{\partial x} \mathbf{u}^k(x, t) = \mathbf{0}, \tag{8.37}$$

$$\mathbf{u}^k(x, 0) = \begin{cases} \bar{\mathbf{u}}_L^k, & \text{if } x < 0, \\ \bar{\mathbf{u}}_R^k, & \text{if } x > 0. \end{cases} \tag{8.38}$$

The sought value $\mathbf{u}^k(0, 0^+)$ is then given by the Godunov state $\bar{\mathbf{u}}_*^k$ of (8.37), (8.38).

This approach is in fact an *acoustical approximation*, which means that we assume the leading term is continuous and only derivatives have jumps. It was observed that this can lead to poor results if the initial data contains large jumps in the state, see Castro and Toro (2008), Montecinos et al. (2012). However, the test cases with very large jumps in the initial data are often of limited importance for practical applications. Wherever the solutions is smooth, the assumptions of the Toro-Titarev

solver are justified and good numerical results are achieved. A theoretical discussion of the Toro-Titarev solver and analysis how jumps in the states relate to numerical errors introduced by the solver was done by Goetz and Iske (2016).

8.5.2 Kernel-Based WENO Reconstruction

To explain the WENO reconstruction scheme, let us first discuss the basic concepts of the *Finite Volume Particle Method (FVPM)* (for further details, we refer to Iske 2013). In the prototypical approach of FVPM, scattered data approximation algorithms are required in the recovery step of the WENO reconstruction.

To explain basic features of the FVPM, let $\mathcal{E} = \{\xi_1, \dots, \xi_n\} \subset \Omega$ denote a finite point set of particles (i.e., particle positions). Moreover, for any particle $\xi \in \mathcal{E}$ we denote its *influence area* by $V_{\mathcal{E}}(\xi) \subset \Omega$. To make a rather straight forward example, the particles' influence areas may, for instance, be given by the Voronoi tiles

$$V_{\mathcal{E}}(\xi) = \left\{ x \in \Omega : \|x - \xi\| = \min_{v \in \mathcal{E}} \|x - v\| \right\} \subset \Omega \quad \text{for } \xi \in \mathcal{E}$$

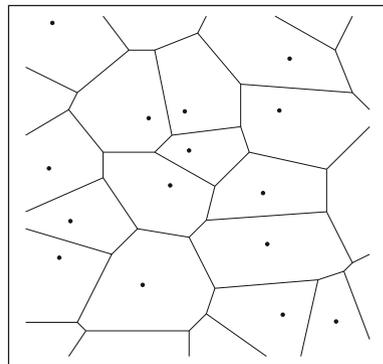
of the Voronoi diagram $\mathcal{V}_{\mathcal{E}} = \{V_{\mathcal{E}}(\xi)\}_{\xi \in \mathcal{E}}$ for \mathcal{E} , in which case $\mathcal{V}_{\mathcal{E}}$ yields a decomposition of Ω into convex and closed subdomains $V_{\mathcal{E}}(\xi) \subset \Omega$ with pairwise disjoint interior, see Figure 8.8 for illustration.

Now, for any particle located at $\xi \in \mathcal{E}$ at time t , its *particle average* is defined by

$$\bar{\mathbf{u}}_{\xi}(t) = \frac{1}{|V_{\mathcal{E}}(\xi)|} \int_{V_{\mathcal{E}}(\xi)} \mathbf{u}(x, t) dx \quad \text{for } \xi \in \mathcal{E} \text{ and } t \in [0, T].$$

Note that the particle average $\bar{\mathbf{u}}_{\xi}$ may be viewed as *cell average*, according to the classical concept of finite volume methods (FVM), see (LeVeque 2002), where $V_{\mathcal{E}}(\xi)$ is referred to as *control volume*.

Fig. 8.8 Finite volume particle method (FVPM). A finite set \mathcal{E} of scattered particles ξ (displayed as ●) and their influence areas, here given by their Voronoi tiles $V_{\mathcal{E}}(\xi)$, are shown



Further according to the concept of FVM, for each $\xi \in \mathcal{E}$ the average value $\bar{\mathbf{u}}_\xi(t)$ is, at time step $t \rightarrow t + \tau$, updated by an explicit numerical method of the form

$$\bar{\mathbf{u}}_\xi(t + \tau) = \bar{\mathbf{u}}_\xi(t) - \frac{\tau}{|V_{\mathcal{E}}(\xi)|} \sum_{\nu} \mathbf{f}_{\xi,\nu}, \quad (8.39)$$

where $\mathbf{f}_{\xi,\nu}$ is the *numerical flux* between particle ξ and a neighbouring $\nu \in \mathcal{E} \setminus \xi$.

The required exchange of information between neighbouring particles is modelled via a generic numerical flux function, which may be implemented by any suitable finite volume flux evaluation scheme, such as by the generalized Godunov approach of high order ADER flux evaluation, as explained in Section 8.5.1.

The following algorithm reflects one basic time step of the FVPM.

Algorithm 1 Finite Volume Particle Method (FVPM).

INPUT: Time step $\tau > 0$, particles \mathcal{E} , particle averages $\{\bar{\mathbf{u}}_\xi(t)\}_{\xi \in \mathcal{E}}$ at time t .

FOR each $\xi \in \mathcal{E}$ **DO**

- (a) Determine set $N_\xi \subset \mathcal{E} \setminus \xi$ of neighbouring particles around ξ ;
- (b) Compute numerical flux $\mathbf{f}_{\xi,\nu}$ for each $\nu \in N_\xi$;
- (c) Update particle average $\bar{\mathbf{u}}_\xi$ for ξ by (8.39).

OUTPUT: Particle averages $\{\bar{\mathbf{u}}_\xi(t + \tau)\}_{\xi \in \mathcal{E}}$ at time $t + \tau$.

Now let us turn to *essentially non-oscillatory* (ENO) and *weighted ENO* (WENO) schemes. To this end, let us view the influence area $V_{\mathcal{E}}(\xi)$ of any particle $\xi \in \mathcal{E}$ as the *control volume* of ξ , where the control volume $V_{\mathcal{E}}(\xi)$ is uniquely represented by ξ .

The basic idea of the ENO method is to first select, for each particle $\xi \in \mathcal{E}$, a small set $\{S_i\}_{i=1}^k$ of k *stencils*, where any stencil $S_i \subset \mathcal{E}$ is given by a set of particles lying in the neighbourhood of ξ . Then, for each stencil S_i , $1 \leq i \leq k$, a reconstruction $s_i \equiv s_{S_i}$ is computed, which interpolates the given particle averages $\{\bar{\mathbf{u}}_\nu(t) : \nu \in S_i\}$ over the control volumes $\{V_{\mathcal{E}}(\nu)\}_{\nu \in S_i}$ of the stencil S_i .

Among the k different reconstructions s_i , $1 \leq i \leq k$, for the k different stencils S_i , the *smoothest* (i.e., the least oscillatory) reconstruction is selected, which constitutes the numerical solution over the control volume $V_{\mathcal{E}}(\xi)$. The selection of the smoothest s_i among the k reconstructions is done by using a suitable *oscillation indicator* \mathcal{I} .

In a WENO reconstruction, *all* reconstructions s_i , $1 \leq i \leq k$, are used to construct, for a corresponding control volume $V_{\mathcal{E}}(\xi)$, a *weighted* sum of the form

$$s(x) = \sum_{i=1}^k \omega_i s_i(x) \quad \text{with} \quad \sum_{i=1}^k \omega_i = 1,$$

where the weights $\omega_i = \tilde{\omega}_i / \sum_{j=1}^k \tilde{\omega}_j$, with $\tilde{\omega}_i = (\epsilon + \mathcal{I}(s_i))^{-\rho}$ for some $\epsilon, \rho > 0$, are determined by using the aforementioned oscillation indicator \mathcal{I} .

We remark that commonly used ENO/WENO schemes work with *polynomial* reconstruction, which, however, may lead to numerical instabilities, especially for anisotropic distributions of particles, see Abgrall (1994). This severe drawback has motivated Aboiyar et al. (2010) to construct a numerically stable reconstruction method of arbitrary high order, which essentially avoids (plain) polynomial reconstruction. The utilized reconstruction relies on a variational formulation, which also provides a natural choice for an efficient oscillation indicator \mathcal{I} , as proposed in (8.44).

Next we explain how to employ kernel-based reconstruction (rather than polynomial reconstruction) in particle flow simulations, in particular in the FVPM. In a generic formulation of particle methods (Iske 2007), we are essentially concerned with the reconstruction of a numerical solution $\mathbf{u} \equiv \mathbf{u}(t, \cdot)$, for fixed time $t \in [0, T]$, from its discrete values

$$\mathbf{u}_{\mathcal{E}} = (\mathbf{u}(\xi_1), \dots, \mathbf{u}(\xi_n))^T \in \mathbb{R}^n,$$

taken at a scattered set $\mathcal{E} = \{\xi_1, \dots, \xi_n\} \subset \mathbb{R}^d$ of particles, cf. Figure 8.9.

Numerical particle flow simulations usually require flexible reconstruction methods from multivariate scattered data approximation to establish, at any time t , the coupling between the discrete model for the numerical solution \mathbf{u} and its continuous output in the recovery step. To this end, (conditionally) positive definite kernel functions are popular tools.

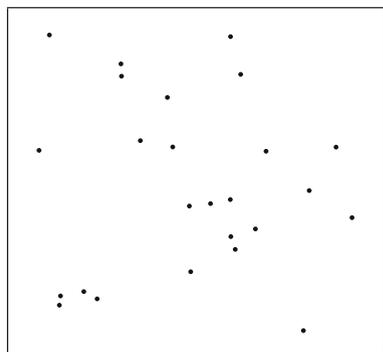
To explain the basic features of such kernel-based reconstructions, we restrict ourselves to the special case of interpolation, where we seek to compute a suitable interpolant $s : \Omega \rightarrow \mathbb{R}$ satisfying $\mathbf{u}_{\mathcal{E}} = s_{\mathcal{E}}$, i.e.,

$$\mathbf{u}(\xi_k) = s(\xi_k) \quad \text{for all } k = 1, \dots, n. \tag{8.40}$$

According to the general formulation of kernel-based interpolation, we assume that the reconstruction s has the form

$$s(x) = \sum_{j=1}^n c_j \varphi(x - \xi_j) + p(x) \quad \text{for } p \in \mathcal{P}_m^d, \tag{8.41}$$

Fig. 8.9 A finite scattered set $\mathcal{E} = \{\xi_1, \dots, \xi_n\} \subset \mathbb{R}^2$ of particles. Each particle ξ (displayed as \bullet) bears a scalar function value $\mathbf{u}(\xi) \equiv \mathbf{u}(\xi, t)$ of the numerical solution \mathbf{u} at time $t \in [0, T]$



for some coefficients $c_1, \dots, c_n \in \mathbb{R}$, where $\varphi : \Omega \rightarrow \mathbb{R}$ is a fixed (conditionally positive definite) *kernel* function and \mathcal{P}_m^d is the linear space of all d -variate polynomials of a specific order $m \in \mathbb{N}_0$. The required order m in (8.41) is determined by the choice of φ . If $m = 0$, then the polynomial part in (8.41) is empty, in which case the reconstruction s has the form

$$s(x) = \sum_{j=1}^n c_j \varphi(x - \xi_j). \tag{8.42}$$

Rather than dwelling much on explaining (conditionally) positive definite kernel functions, we refer to the text books (Buhmann 2003; Iske 2018; Wendland 2005). For the following of our discussion, it is sufficient to say that scattered data interpolation by positive definite kernels (where $m = 0$) leads to a unique reconstruction of the form (8.42). Moreover, for conditionally positive definite kernels of order $m \in \mathbb{N}$, we obtain under vanishing moment conditions

$$\sum_{j=1}^n c_j p(\xi_j) = 0 \quad \text{for all } p \in \mathcal{P}_m^d \tag{8.43}$$

a reconstruction s of the form (8.41), where s is unique, if any polynomial $p \in \mathcal{P}_m^d$ can uniquely be reconstructed from its values at the points \mathcal{E} , i.e., $p_{\mathcal{E}} = 0$ implies $p \equiv 0$.

Let us make examples of commonly used radial kernel functions $\varphi(x) = \phi(\|x\|)$, along with their orders $m \equiv m(\phi)$, where $r = \|x\| \in [0, \infty)$ is, for $x \in \mathbb{R}^d$, the radial variable w.r.t. the Euclidean norm $\|\cdot\|$ on \mathbb{R}^d .

Example 8.1 The positive definite *Gaussian* function

$$\phi(r) = e^{-r^2} \quad \text{for } r \in [0, \infty)$$

is a radial kernel of order $m = 0$, so that the reconstruction s has the form (8.42).

Example 8.2 The *multiquadric*

$$\phi(r) = (1 + r^2)^\beta \quad \text{for } \beta > 0 \text{ and } \beta \notin \mathbb{N}$$

is a conditionally positive definite kernel of order $m = \lceil \beta \rceil$.

The *inverse multiquadric*

$$\phi(r) = (1 + r^2)^\beta \quad \text{for } \beta < 0$$

is positive definite, i.e., $m = 0$. In this case, the reconstruction s has the form (8.42).

Example 8.3 The compactly supported *radial characteristic functions* (Askey 1973)

$$\phi(r) = (1-r)_+^\beta = \begin{cases} (1-r)^\beta & \text{for } r < 1 \\ 0 & \text{for } r \geq 1 \end{cases}$$

are for $d \geq 2$ positive definite on \mathbb{R}^d , provided that $\beta \geq (d+1)/2$. In this case, $m = 0$, and so the reconstruction s has the form (8.42).

Next, we add *polyharmonic splines* to the list of our examples. Polyharmonic splines are extraordinarily useful radial kernels. In fact, we will give a strong recommendation in favour of polyharmonic splines, where our arguments are based on their superior numerical stability at arbitrary high local approximation order.

Polyharmonic splines, due to Duchon (1977), are traditional tools for Lagrange interpolation from multivariate scattered data. According to the polyharmonic spline interpolation scheme, the reconstruction s has the form (8.41), where the radial polyharmonic spline kernel $\varphi(x) = \phi_{d,m}(r)$, for $r = \|x\|$, is given as

$$\phi_{d,m}(r) = \begin{cases} r^{2m-d} \log(r) & \text{for } d \text{ even} \\ r^{2m-d} & \text{for } d \text{ odd} \end{cases} \quad \text{for } 2m > d,$$

with m being the order of the kernel $\phi_{d,m}$, i.e., m is the order of the polynomial in (8.41).

According to Duchon (1977), scattered data interpolation by polyharmonic splines is *optimal* in its native reproducing kernel Hilbert space, as given by the *Beppo Levi space*

$$\text{BL}^m(\mathbb{R}^d) = \{ \mathbf{u} : D^\alpha \mathbf{u} \in L^2(\mathbb{R}^d) \text{ for all } |\alpha| = m \} \subset \mathcal{C}(\mathbb{R}^d),$$

being equipped with the semi-norm

$$|\mathbf{u}|_{\text{BL}^m}^2 = \sum_{|\alpha|=m} \binom{m}{\alpha} \|D^\alpha \mathbf{u}\|_{L^2(\mathbb{R}^d)}^2.$$

In other words, the reconstruction s in (8.41) minimizes the Beppo Levi energy functional $|\cdot|_{\text{BL}^m}$ among all recovery functions \mathbf{u} in $\text{BL}^m(\mathbb{R}^d)$, i.e.,

$$|s|_{\text{BL}^m} \leq |\mathbf{u}|_{\text{BL}^m}, \quad \text{for all } \mathbf{u} \in \text{BL}^m(\mathbb{R}^d) \text{ with } \mathbf{u}_{\mathcal{E}} = s_{\mathcal{E}}.$$

Therefore, the energy functional $|\cdot|_{\text{BL}^m}$ is a natural choice for the oscillation indicator \mathcal{I} required in the WENO reconstruction (cf. Aboiyar et al. 2010). Hence, we let

$$\mathcal{I}(\mathbf{u}) := |\mathbf{u}|_{\text{BL}^m} \tag{8.44}$$

for the oscillation indicator of the utilized WENO reconstruction. We remark that the semi-norm $|s|_{\text{BL}^m}$ of the polyharmonic spline reconstruction s is readily available by the quadratic form

$$|s|_{\text{BL}^m}^2 = \sum_{j,k=1}^n c_j c_k \phi_{d,m}(\|\xi_k - \xi_j\|),$$

whose coefficient vector $c = (c_1, \dots, c_n)^T \in \mathbb{R}^n$ is determined by the solution of the interpolation problem's resulting linear equation system. This allows efficient evaluations of the reconstruction's oscillation indicator $\mathcal{I}(s)$, giving the above choice in (8.44) yet another advantage.

We finally remark that numerical properties of polyharmonic splines in kernel-based WENO reconstruction were first investigated in (Aboiyar et al. 2010) and further explained in (Iske 2013). The following advantages in favour of the kernel-based polyharmonic WENO reconstruction were identified (see Iske 2013).

Well-posedness. Polyharmonic spline interpolation yields a well-posed WENO reconstruction method, which works for arbitrary distributions of scattered particles $\mathcal{E} \subset \mathbb{R}^d$, for arbitrary average values $\bar{u}_{\mathcal{E}}$, and in arbitrary space dimensions $d \geq 1$.

Optimality. Polyharmonic spline reconstruction is *optimal* in its Beppo-Levi space $\text{BL}^m(\mathbb{R}^d)$. Note that the *optimal recovery space* $\text{BL}^m(\mathbb{R}^d)$ is the Sobolev space $\mathcal{H}^m(\mathbb{R}^d)$, which is a relevant function space in nonlinear hyperbolic problems.

Meshfree reconstruction and high flexibility. The reconstruction scheme of polyharmonic splines is meshfree and therefore very flexible, especially when it comes to adaptively modifying the set of moving particles. This property is particularly important for problems with solutions of rapid variation or singularities, or for problems with free or complicated boundaries as well as for various other challenging problems, where the simulation of multiscale phenomena is an issue.

Efficient implementation. The implementation of the polyharmonic spline reconstruction scheme merely requires solving a square linear system, which is small, if the number of particles in \mathcal{E} is small. But the linear system may be ill-conditioned, if the *separation* $q_{\mathcal{E}} := \min_{\xi, \nu \in \mathcal{E}, \xi \neq \nu} \|\xi - \nu\|$ of the particles in \mathcal{E} is small.

Numerical stability. The polyharmonic spline reconstruction scheme allows for numerically stable evaluations of the interpolant, by using the preconditioner proposed in Iske (2013).

Arbitrary local approximation order. Polyharmonic spline reconstruction has local approximation order m with respect to \mathcal{C}^m functions, for the kernel $\phi_{d,m}$.

Adaptivity. Polyharmonic spline reconstruction supports the implementation of effective adaption rules. This is due to a customized error indicator, which comes with the design of the reconstruction scheme. In fact, the error indicator provides suitable rules for the coarsening and refinement of particles (see Iske 2013).

We conclude our discussion by giving a strong recommendation in favour of polyharmonic splines, especially when they are used in combination with adaptive particle methods for numerical flow simulation, where highly flexible and sufficiently robust reconstructions of arbitrary order are needed. This recommendation is supported by our numerical experiments (Aboiyar et al. 2010; Iske 2013).

8.6 Discussion and Conclusions

While the methods reviewed above may contribute to the reduction of spurious mixing, the extent to which they do so remains unexplored and is expected to depend on applications. It is therefore important that they are analysed in more detail, together with accompanying questions of numerical stability and efficiency. The focus of this chapter was on measures largely involving tracer dynamics, however the question of spurious mixing involves not only tracers, but also the small-scale structure of the simulated velocity field. By analysing the behaviour of RPE in a series of test cases Ilicak et al. (2012) show that spurious mixing is strongly sensitive to the grid-scale Reynolds number and demonstrate that the amount of diapycnal mixing in a suite of ocean models with different high-order advection schemes can indeed be controlled by keeping this number around two via the Smagorinsky viscosity. Coupling of spurious mixing to functioning of other popular operators of sub-grid momentum mixing and also to the momentum advection implementation (flux or vector-invariant, its spatial accuracy), remains to be explored. In many unstructured-mesh discretisations the sizes of velocity and scalar spaces are inconsistent (Danilov 2013), which may have additional impact on the diapycnal mixing. Commonly the velocity space is too large, resulting in spurious inertial oscillations generating spurious vertical velocities in eddying regimes. In contrast, a triangular C-grid approach has too many pressure points, leading to a mode in the divergence of horizontal velocity, but finally in the vertical velocity too. The related questions on the implications of discretisation detail, mesh geometry and the implementation of momentum advection need to be explored. One more question is related to the mere effect of mesh unstructuredness for models formulated on unstructured meshes. All these questions present an interesting direction of research which will contribute to improving consistency of ocean circulation models.

To conclude, this chapter reviewed a selection of methods that might be useful in reducing spurious mixing in ocean circulation models. Most of them are known from previous research, yet their adaptation to new appearing models as well as learning about the full extent of their effect remains a pressing topic.

References

- Abalakin, I., Dervieux, A., Kozubskaya, T.: A vertex-centered high-order muscl scheme applying to linearized euler acoustics. *Rapport de recherche* **4459**, INRIA (2002)
- Abgrall, R.: On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation. *J. Comput. Phys.* **144**, 45–58 (1994)
- Aboiyar, T., Georgoulis, E., Iske, A.: Adaptive ADER methods using kernel-based polyharmonic spline WENO reconstruction. *SIAM J. Sci. Comput.* **32**(6), 3251–3277 (2010)
- Adcroft, A., Hallberg, R.: On methods for solving the oceanic equations of motion in generalized vertical coordinates. *Ocean Modell.* **11**, 224–233 (2006)
- Askey, R.: Radial characteristic functions. Technical Report TSR # 1262, University of Wisconsin, Madison, USA (1973)

- Balsara, D., Meyer, C., Dumbser, M., Du, H., Xu, Z.: Efficient implementation of ADER schemes for Euler and magnetohydrodynamical flows on structured meshes—speed comparisons with runge-kutta methods. *J. Comput. Phys.* **235**, 934–969 (2013)
- Ben-Artzi, M., Falcovitz, J.: A second-order Godunov type scheme for compressible fluid dynamics. *J. Comput. Phys.* **55**, 1–32 (1984)
- Ben-Artzi, M., Falcovitz, J.: *Generalized Riemann Problems in Computational Fluid Dynamics*. Cambridge University Press, Cambridge (2003)
- Ben-Artzi, M., Falcovitz, J., Li, J.: The convergence of the GRP scheme. *Discrete Contin. Dyn. Syst.* **23**, 1–27 (2009)
- Ben-Artzi, M., Li, J.: Hyperbolic balance laws: Riemann invariants and the generalized Riemann problem. *Numer. Math.* **106**, 369–425 (2007)
- Ben-Artzi, M., Li, J., Warnecke, G.: A direct Eulerian GRP scheme for compressible fluid flow. *J. Comput. Phys.* **218**, 19–43 (2006)
- Boscheri, W., Balsara, D., Dumbser, M.: Lagrangian ADER-WENO finite volume schemes on unstructured triangular meshes based on genuinely multidimensional HLL Riemann solvers. *J. Comput. Phys.* **267**, 112–138 (2014a)
- Boscheri, W., Dumbser: Arbitrary-Lagrangian-Eulerian one-step WENO finite volume schemes on unstructured triangular meshes. *Commun. Comput. Phys.* **14**, 1174–1206 (2013)
- Boscheri, W., Dumbser, M.: A direct arbitrary-Lagrangian-Eulerian ADER-WENO finite volume scheme on unstructured tetrahedral meshes for conservative and nonconservative hyperbolic systems in 3d. *J. Comput. Phys.* **275**, 484–523 (2014)
- Boscheri, W., Dumbser, M., Balsara, D.: High order Lagrangian ADER-WENO schemes on unstructured meshes—application of several node solvers to hydrodynamics and magnetohydrodynamics. *Int. J. Numer. Methods Fluids* **76**, 737–778 (2014b)
- Bourgeade, A., LeFloch, P., Raviart, P.: An asymptotic expansion for the solution of the generalized Riemann problem. Part 2: application to the equations of gas dynamics. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **6**, 437–480 (1989)
- Bressan, A.: *Hyperbolic Systems of Conservation Laws—The Onedimensional Cauchy Problem*. Oxford University Press, Oxford (2000)
- Buhmann, M.D.: *Radial Basis Functions: Theory and Implementations*. Cambridge University Press, Cambridge, UK (2003)
- Burchard, H.: Quantification of numerically induced mixing and dissipation in discretisations of shallow water equations. *Int. J. Geomath* **3**, 51–65 (2012)
- Burchard, H., Beckers, J.M.: Non-uniform adaptive vertical grids in one-dimensional numerical ocean models. *Ocean Modell.* **6**, 51–81 (2004)
- Burchard, H., Bolding, K.: GETM—a general estuarine transport model. Scientific documentation. Technical Report EUR 20253 EN, European Commission (2002)
- Burchard, H., Rennau, H.: Comparative quantification of physically and numerically induced mixing in ocean models. *Ocean Modell.* **20**, 293–311 (2008)
- Castro, C.E., Toro, E.F.: Solvers for the high-order Riemann problem for hyperbolic balance laws. *J. Comput. Phys.* **227**, 2481–2513 (2008)
- Colella, P., Woodward, P.R.: The piecewise parabolic method (PPM) for gas-dynamical simulations. *J. Comput. Phys.* **54** (1984)
- Dafermos, C.M.: *Hyperbolic Conservation Laws in Continuum Physics*. Springer, Heidelberg (2010)
- Danilov, S.: Ocean modeling on unstructured meshes. *Ocean Modell.* **69**, 195–210 (2013)
- Danilov, S., Sidorenko, D., Wang, Q., Jung, T.: The Finite-volume Sea ice-Ocean model (FESOM2). *Geosci. Mod. Dev.* **10**, 765–789 (2017)
- Donea, J., Huerta, A.: *Finite Element Methods for Flow Problems*. Wiley (2003)
- Donea, J., Huerta, A., Ponthot, J.-P., Rodríguez-Ferran, A.: Arbitrary lagrangian-eulerian methods. In: Stein, E., de Borst, R., Hughes, T.J.R. (eds.) *Encyclopedia of Computational Mechanics*, vol. 1, chapter 14, pp. 1–25. Wiley (2004)

- Duchon, J.: Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In: Schempp, W., Zeller, K. (eds.) *Constructive Theory of Functions of Several Variables*, pp. 85–100. Springer (1977)
- Dumbser, M., Balsara, D., Toro, E., Munz, C.-D.: A unified framework for the construction of one-step finite volume and discontinuous Galerkin schemes on unstructured meshes. *J. Comput. Phys.* **227**, 8209–8253 (2008a)
- Dumbser, M., Enaux, C., Toro, E.F.: Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws. *J. Comput. Phys.* **227**, 3971–4001 (2008b)
- Dumbser, M., Käser, M.: Arbitrary high-order non-oscillatory finite volume schemes on unstructured meshes for linear hyperbolic systems. *J. Comput. Phys.* **221**, 693–723 (2007)
- Dumbser, M., Peshkov, I., Romenski, E., Zanotti, O.: High order ADER schemes for a unified first order hyperbolic formulation of continuum mechanics: viscous heat-conducting fluids and elastic solids. *J. Comput. Phys.* **314**, 824–862 (2016)
- Gassner, G., Dumbser, M., Hindenlang, F., Munz, C.D.: Explicit one-step time discretizations for discontinuous Galerkin and finite volume schemes based on local predictors. *J. Comput. Phys.* **230**, 4232–4247 (2011)
- Gassner, G., Lörcher, F., Munz, C.D.: A discontinuous Galerkin scheme based on a space-time expansion ii. Viscous flow equations in multi dimensions. *J. Sci. Comput.* **34**, 260–286 (2008)
- Getzlaff, J., Nurser, G., Oschlies, A.: Diagnostics of diapycnal diffusivity in z -level ocean models. Part i: 1-dimensional case studies. *Ocean Modell.* **35**, 173–186 (2010)
- Getzlaff, J., Nurser, G., Oschlies, A.: Diagnostics of diapycnal diffusion in z -level ocean models. Part ii: 3-dimensional OGCM. *Ocean Modell.* **45–46**, 27–36 (2012)
- Gibson, A.H., Hogg, A.M., Kiss, A.E., Shakespeare, C.J., Adcroft, A.: Attribution of horizontal and vertical contributions to spurious mixing in an Arbitrary Lagrangian Eulerian ocean model. *Ocean Modell.* **119**, 45–56 (2017)
- Godunov, S.K.: A difference scheme for numerical solution of discontinuous solution of hydrodynamic equations (in Russian). *Math. Sbornik* **47**, 271–306 (1959)
- Goetz, C.R., Dumbser, M.: A novel solver for the generalized Riemann problem based on a simplified LeFloch-Raviart expansion and a local space-time discontinuous Galerkin formulation. *J. Sci. Comput.* **69**, 805–840 (2016)
- Goetz, C.R., Iske, A.: Approximate solutions of generalized Riemann problems for nonlinear systems of hyperbolic conservation laws. *Math. Comput.* **85**, 35–62 (2016)
- Gräwe, U., Holtermann, P., Klingbeil, K., Burchard, H.: Advantages of vertically adaptive coordinates in numerical models of stratified shelf seas. *Ocean Modell.* **92**, 56–68 (2015)
- Griffies, S.M., Pacanowski, R.C., Hallberg, R.: Spurious diapycnal mixing associated with advection in a z -coordinate ocean model. *Mon. Wea. Rev.* **128**, 538–564 (2000)
- Harabetian, E.: A convergent series expansion for hyperbolic systems of conservation laws. *Trans. Am. Math. Soc.* **294**, 383–424 (1986)
- Harten, A., Engquist, B., Osher, S., Chakravarthy, S.R.: Uniformly high order essentially non-oscillatory schemes. III. *J. Comput. Phys.* **77**, 231–303 (1987)
- Hill, C., Ferreira, D., Campin, J.-M., Marshall, J., Abernathy, R., Barrie, N.: Controlling spurious diapycnal mixing in eddy-resolving height-coordinate ocean models—insights from virtual deliberate tracer release experiments. *Ocean Modell.* **45–46**, 14–26 (2012)
- Hirt, C.W.: Heuristic stability theory for finite-difference equations. *J. Comput. Phys.* **2**, 339–355 (1968)
- Hofmeister, R., Burchard, H., Beckers, J.M.: Non-uniform adaptive vertical grids for 3D numerical ocean models. *Ocean Modell.* **33**, 70–86 (2010)
- Ilicak, M., Adcroft, A.J., Griffies, S.M., Hallberg, R.W.: Spurious diapycnal mixing and the role of momentum closure. *Ocean Modell.* **45**, 37–58 (2012)
- Ilicak, M.: Quantifying spatial distribution of spurious mixing in ocean models. *Ocean Modell.* **108**, 30–38 (2016)
- Iske, A.: Polyharmonic spline reconstruction in adaptive particle flow simulation. In: Iske, A., Levesley, J. (eds.) *Algorithms for Approximation*, pp. 83–102. Springer, Berlin (2007)

- Iske, A.: On the construction of kernel-based adaptive particle methods in numerical flow simulation. In: Ansonge, R, Bijl, H., Meister, A., Sonar, T. (eds.) *Recent Developments in the Numerics of Nonlinear Hyperbolic Conservation. Notes on Numerical Fluid Mechanics and Multidisciplinary Design (NNFM)*, pp. 197–221. Springer, Berlin (2013)
- Iske, A.: *Approximation*. Springer-Lehrbuch Masterclass, Springer Spektrum (2018)
- Klingbeil, K., Burchard, H.: Implementation of a direct nonhydrostatic pressure gradient discretisation into a layered ocean model. *Ocean Modell.* **65**, 64–77 (2013)
- Klingbeil, K., Mohammadi-Aragh, M., Gräwe, U., Burchard, H.: Quantification of spurious dissipation and mixing discrete variance decay in a finite-volume framework. *Ocean Modell.* **81**, 49–64 (2014)
- Klingbeil, K., Lemarié, F., Debreu, L., Burchard, H.: The numerics of hydrostatic structured-grid coastal ocean models: state of the art and future perspectives. *Ocean Modell.* **125**, 80–105 (2018a)
- Klingbeil, K., Becherer, J., Schulz, E., de Swart, H. E., Schuttelaars, H. M., Valle-Levinson, A., Burchard, H.: Thickness-weighted averaging in tidal estuaries and the vertical distribution of the Eulerian residual transport. *J. Phys. Oceanogr* (2018b) (revised)
- Lax, P.D.: Hyperbolic systems of conservation laws II. *Commun. Pure Appl. Math.* **10**(4), 537–566 (1957)
- Leclair, M., Madec, G.: \tilde{z} -coordinate, an Arbitrary Lagrangian-Eulerian coordinate separating high and low frequency motions. *Ocean Modell.* **37**, 139–152 (2011)
- LeFloch, P., Raviart, P.A.: An asymptotic expansion for the solution of the generalized Riemann problem. Part I: general theory. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **5**, 179–207 (1988)
- Lemarié, F., Debreu, L., Shchepetkin, A.F., McWilliams, J.C.: On the stability and accuracy of the harmonic and biharmonic isoneutral mixing operators in ocean models. *Ocean Modell.* **52–53**, 9–35 (2012a)
- Lemarié, F., Kurian, J., Shchepetkin, A.F., Molemaker, M.J., Colas, F., McWilliams, J.C.: Are there inescapable issues prohibiting the use of terrain-following coordinates in climate models? *Ocean Modell.* **42**, 57–79 (2012b)
- LeVeque, R.: *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, Cambridge, UK (2002)
- Li, T., Yu, W.: *Boundary Value Problems for Quasilinear Hyperbolic Systems*. Duke University Press, Durham (1985)
- Lörcher, F., Gassner, G., Munz, C.D.: A discontinuous Galerkin scheme based on a space-time expansion. I. Inviscid compressible flow in one space dimension. *J. Sci. Comput.* **32**, 175–199 (2007)
- Mohammadi-Aragh, M., Klingbeil, K., Brüggemann, N., Eden, C., Burchard, H.: The impact of advection schemes on restratification due to lateral shear and baroclinic instabilities. *Ocean Modell.* **94**, 112–127 (2015)
- Montecinos, G., Castro, C.E., Dumbser, M., Toro, E.F.: Comparison of solvers for the generalized Riemann problem for hyperbolic systems with source terms. *J. Comput. Phys.* **231**, 6472–6494 (2012)
- Morales Maqueda, M.A., Holloway, G.: Second-order moment advection scheme applied to Arctic Ocean simulation. *Ocean Modell.* **14**, 197–221 (2006)
- Nakayama, Y., Timmermann, R., Schröder, M., Hellmer, H.: On the difficulty of modeling circumpolar deep water intrusions onto the Amundsen Sea continental shelf. *Ocean Modell.* **84**, 26–34 (2014)
- Petersen, M.R., Jacobsen, D.W., Ringler, T.D., Hecht, M.W., Maltrud, M.E.: Evaluation of the arbitrary lagrangian-eulerian vertical coordinate method in the mpas-ocean model. *Ocean Modell.* **86**, 93–113 (2015)
- Prather, M.J.: Numerical advection by conservation of second-order moments. *J. Geophys. Res.* **91**(D6), 6671–6681 (1986)
- Qian, J., Li, J., Wang, S.: The generalized Riemann problem for compressible fluid flows: towards high order. *J. Comput. Phys.* **259**, 358–389 (2014)

- Ringler, T., Petersen, M., Higdon, R.L., Jacobsen, D., Jones, P.W., Maltrud, M.: A multi-resolution approach to global ocean modeling. *Ocean Modell.* **69**, 211–232 (2013)
- Shchepetkin, A.F., McWilliams, J.C.: Quasi-monotone advection schemes based on explicit locally adaptive dissipation. *Mon. Weather. Rev.* **126**, 1541–1580 (1998)
- Shchepetkin, A.F., McWilliams, J.C.: Correction and commentary for “Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the regional ocean modeling system” by Haidvogel et al. *Journal of Computational Physics*, vol. 227, pp. 3595–3624; vol. 228, pp. 8985–9000 (2009)
- Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Goessling, H.F., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schröter, J., Sein, D.V., Wang, Q.: Towards multi-resolution global climate modeling with echam6-fesom. Part i: model formulation and mean climate. *Clim. Dyn.* **44**(3), 757–780 (2015)
- Skamarock, W.C., Gassmann, A.: Conservative transport schemes for spherical geodesic grids: high-order flux operators for ode-based time integration. *Mon. Wea. Rev.* **139**, 2962–2975 (2011)
- Smoller, J.: *Shock Waves and Reaction-Diffusion Equations*. Springer, New York (1994)
- Tatsien, L., Libin, W.: *Global Propagation of Regular Nonlinear Hyperbolic Waves*. Birkhäuser, Boston (2009)
- Timmermann, R., Wang, Q., Hellmer, H.: Ice-shelf basal melting in a global finite-element sea-ice/ice-shelf/ocean model. *Ann. Glaciol.* **53**, 303–314 (2012)
- Titarev, V.A., Toro, E.F.: ADER: arbitrary high order Godunov approach. *J. Sci. Comput.* **17**, 609–618 (2002)
- Toro, E., Montecinos, G.: Implicit, semi-analytical solution of the generalized Riemann problem for stiff hyperbolic balance laws. *J. Comput. Phys.* **303**, 146–172 (2015)
- Toro, E. F.: *Riemann Solvers and Numerical Methods for Fluid Dynamics*, 3 edn. Springer (2009)
- Toro, E.F.: The Riemann problem: solvers and numerical fluxes. In: *Handbook of Numerical Analysis*, vol. 17, chapter 2, pp. 19–54. Elsevier (2016)
- Toro, E.F., Millington, R.C., Nejad, L.A.M.: Towards very high order Godunov schemes. In: Toro, E.F. (ed.) *Godunov Methods; Theory and Applications*, pp. 907–940, Oxford. Kluwer Academic Plenum Publishers. International Conference (2001)
- Toro, E.F., Millington, R.C., Nejad, L.A.M.: Towards very high order Godunov schemes. In: Toro, E.F. (ed.) *Godunov Methods; Theory and Applications*, pp. 907–940. Kluwer Academic Plenum Publishers, Oxford. International Conference (2001)
- Toro, E.F., Titarev, V.A.: Solution of the generalized Riemann problem for advection? Reaction equations. In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 458, pp. 271–281 (2002)
- Toro, E.F., Titarev, V.A.: Derivative Riemann solvers for systems of conservation laws and ADER methods. *J. Comput. Phys.* **212**, 150–165 (2006)
- Urakawa, L.S., Hasumi, H.: Effect of numerical diffusion on the water mass transformation in eddy-resolving models. *Ocean Modell.* **74**, 22–35 (2014)
- van Leer, B.: Towards the ultimate conservative difference scheme. v. a second-order sequel to Godunov’s method. *J. Comput. Phys.* **32**, 101–136 (1979)
- Wan, H., Giorgetta, M.A., Zängl, G., Restelli, M., Majewski, D., Bonaventura, L., Fröhlich, K., Reinert, D., Rípodas, P., Kornblueh, L., Förstner, J.: The ICON-1.2 hydrostatic atmospheric dynamical core on triangular grids—part 1: formulation and performance of the baseline version. *Geosci. Model Dev.* **6**(3), 735–763 (2013)
- Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., Jung, T., Schröter, J.: The finite element sea ice-ocean model (fesom) v.1.4: formulation of an ocean general circulation model. *Geosci. Model Dev.* **7**, 663–693 (2014)
- Warming, R.F., Hyett, B.J.: The modified equation approach to the stability and accuracy analysis of finite-difference methods. *J. Comput. Phys.* **14**, 159–179 (1974)
- Webb, D.J., de Cuevas, B.A., Richmond, C.S.: Improved advection schemes for ocean models. *J. Atmos. Ocean. Technol.* **15**(5), 1171–1187 (1998)
- Wendland, H.: *Scattered Data Approximation*. Cambridge University Press, Cambridge, UK (2005)

- Yang, Z., He, P., Tang, H.: A direct eulerian grp scheme for relativistic hydrodynamics: one-dimensional case. *J. Comput. Phys.* **230**, 7964–7987 (2011)
- Zanotti, O., Dumbser, M.: Efficient conservative ADER schemes based on WENO reconstruction and space-time predictor in primitive variables. *Comput. Astrophys. Cosmol.* (2016)
- Zanotti, O., Fambri, F., Dumbser, M.: Solving the relativistic magnetohydrodynamics equations with ADER discontinuous Galerkin methods, a posteriori subcell limiting and adaptive mesh refinement. *Mon. Not. R. Astron. Soc.* **452**, 3290 (2015)

Chapter 9

Diffuse Interface Approaches in Atmosphere and Ocean—Modeling and Numerical Implementation



Harald Garcke, Michael Hinze and Christian Kahle

Abstract We propose to model physical effects at the sharp density interface between atmosphere and ocean with the help of diffuse interface approaches for multiphase flows with variable densities. We use the *thermodynamical consistent* variable density model proposed in Abels et al. (Mathematical Models and Methods in Applied Sciences 22:1150013, 2012). This results in a Cahn–Hilliard/Navier–Stokes-type system which we complement with tangential Dirichlet boundary conditions to incorporate the effect of wind in the atmosphere. Wind is responsible for waves at the surface of the ocean, whose dynamics have an important impact on the CO_2 —exchange between ocean and atmosphere. We tackle this mathematical model numerically with fully adaptive and integrated numerical schemes tailored to the simulation of variable density multiphase flows governed by diffuse interface models. Here, *fully adaptive, integrated, efficient, and reliable* means that the mesh resolution is chosen by the numerical algorithm according to a prescribed error tolerance in the *a posteriori* error control on the basis of residual-based error indicators, which allow to estimate the true error from below (efficient) and from above (reliable). Our approach is based on the work of Hintermüller et al. (Journal of Computational Physics 235:810–827, 2013), Garcke et al. (Applied Numerical Mathematics 99:151–171, 2016), where a fully adaptive efficient and reliable numerical method for the simulation of two-dimensional multiphase flows with variable densities is developed. In a first step, we incorporate the stimulation of surface waves via appropriate volume forcing.

H. Garcke

Fakultät für Mathematik, Universität Regensburg, Regensburg, Germany
e-mail: harald.garcke@mathematik.uni-regensburg.de

M. Hinze (✉)

Fachbereich Mathematik, Universität Hamburg, Hamburg, Germany
e-mail: michael.hinze@uni-hamburg.de

C. Kahle

Zentrum Mathematik, Technische Universität München, Munich, Germany
e-mail: christian.kahle@ma.tum.de

© Springer Nature Switzerland AG 2019

C. Eden and A. Iske (eds.), *Energy Transfers in Atmosphere and Ocean*,
Mathematics of Planet Earth 1, https://doi.org/10.1007/978-3-030-05704-6_9

287

9.1 Introduction

The energy and momentum transfer from the atmosphere to the ocean is an essential ingredient for the accurate modeling of the energy cycle. In fact, the vast majority of the energy input to the ocean comes from the winds (~ 20 TW), with much smaller inputs from tides (~ 3.5 TW) and geothermal heating (~ 0.05 TW) (numbers taken from Wunsch and Ferrari 2004). It is therefore not surprising that the energy transfer from the wind to the surface wave field, and the ensuing energy dissipation through breaking waves, represents the largest transfer of energy in the oceans (Wunsch and Ferrari 2004). Despite the enormous importance of the processes of surface wave generation and dissipation, there are still fundamental gaps in our ability to conduct both process modeling and observational studies of these processes operating near the air–sea interface.

However, many major advances have recently been made through the use of powerful numerical simulations (see the recent review by Sullivan and McWilliams 2010). These simulations have shown that classical modeling and parameterization techniques, such as the use of the law-of-the-wall turbulence scaling, must be revised to account for the dynamics of wind waves. For a number of results which the geoscience community expects to be included if an accurate, and energy consistent, treatment of atmosphere–ocean interactions is to be accomplished, see, e.g., Sullivan et al. (2007), Drennan et al. (1998), Sutherland and Melville (2015), McWilliams et al. (1997), Sullivan et al. (2007), Polton et al. (2008).

Here, we propose an *holistic* approach to an accurate and energy consistent treatment of atmosphere–ocean interactions based on the thermodynamically consistent diffuse interface model suggested in Abels et al. (2012) to model the air–water interface between atmosphere and ocean. This model will be extended to produce a series of direct numerical simulations of wind generated waves. Only very few studies have examined the evolution of wind waves using such a fundamental approach (Kihara et al. 2007; Shen et al. 2003; Sullivan et al. 2000; Sullivan and McWilliams 2002; Tsai and Hung 2007; Tsai et al. 2015; Lubin and Glockner 2015). However, these studies often do not involve a proper coupling of the water surface and the air flow above. For example, the water surface is often replaced by another simpler boundary condition, such as an impermeable sinusoidal wall (Shen et al. 2003), or an uncoupled propagating water wave solution (Sullivan et al. 2000; Sullivan and McWilliams 2002). In these situations, the evolution of the air–water interface is not driven by wind or the actual flow field of the ocean. The recent study of Lubin and Glockner (2015) has shown how powerful a direct coupling of the air and water layers is for predicting turbulent air-entraining structures in the breaking of surface waves. Another study that utilizes a fully coupled treatment of the air and water layers is that of Tsai et al. (2015). They show the important result that turbulent water flows are generated even under the conditions of non-breaking surface waves. In all approaches above, a so-called sharp representation of the air–water interface is used. One major drawback of this formulation is model break down as, e.g., topology changes, which appears in the case of wave breaking. We believe that the diffuse interface methods developed for the Cahn–Hilliard/Navier–Stokes system will provide an improved

method to deal with the current shortcomings of simulating a direct coupling of the air–water interface. In diffuse interface models, the surface is assumed to have a positive thickness, where both air and water exist in a mixed state. Further, this surface is not explicitly given, but is an outcome of the model. The distribution of fluid and air is described by a smooth indicator function, called phase field variable, which is a smooth function taking the value $+1$ in the pure air phase and -1 in the pure water phase, with a rapid transition between the phases across the diffuse interface. The topology of the phases is not explicitly specified in the model so that topology changes like wave breaking are possible with this model. Another benefit is the mass-conserving property of the approach; see, e.g., Hintermüller et al. (2013). On the long run, it is planned to use the method to simulate wind-wave growth and compare the numerical results to laboratory experiments using the PIV technique to resolve the airflow and water surface elevation.

The diffuse interface method of treating the air–water interface using the Cahn–Hilliard/Navier–Stokes (CHNS) system forms a new approach to the model studies described so far. We note, however, that there are several contributions to numerical approaches to the simulation of multiphase flows in the *sharp* interface formulation. Here, we refer, e.g., to the book of Gross and Reusken (2011), the work of Ganesan and Tobiska (2009), and the works of Barrett et al. (2014) and Wan and Turek (2007). A benchmark for sharp interface approaches to the numerical simulation of rising bubble dynamics is proposed by Hysing et al. (2009), which is accomplished with diffuse interface simulations by Aland and Voigt (2012). A review of the development of phase field models and their numerical methods for multicomponent fluid flows with interfacial phenomena is given by Kim (2012). In the context of mechanical engineering and meteorological applications, phase field models for two-phase flows are often referred to as the *two-fluid formulation*; see, e.g., Mellado et al. (2010), and Druzhinin and Elghobashi (1998); compare also the related *volume-of-fluid schemes*; see, e.g., James and Lowengrub (2004), as well as the references cited therein.

Since the dynamics of multiphase flows essentially depend on the dynamics at the interfaces, it is important to resolve the interfacial region in diffuse interface models well. Here, adaptive numerical concepts are the method of choice. Concerning the existing literature on the solver development for the coupled CHNS system, we note that in Kay and Welford (2007) a robust (with respect to the interfacial width) non-linear multigrid method was introduced with a double-well homogeneous free energy density. We refer to Kay and Welford (2006) for the multigrid solver for the Cahn–Hilliard (CH) part only. Later, in Kay et al. (2008) error estimates for the coupled system were derived and numerically verified. Coupled CHNS systems were also considered in Boyer et al. (2010) with a double-well potential in the case of three-phase flows; see also (Boyer 1999, 2002; Boyer et al. 2004) and (Aland et al. 2010; Aland and Voigt 2012) as well as the references therein for rather qualitative studies of the behavior of multiphase and mixture flows.

Stable numerical schemes for the recently developed thermodynamically consistent diffuse interface model (Abels et al. 2012) are developed in Garcke et al. (2016), Grün et al. (2016), Grün and Klingbeil (2014). Concerning the numerical treatment of the sole CH system, many contributions can be found in the literature. For a

rather comprehensive discussion of available solvers, we refer to Hintermüller et al. (2011). In the latter work, a *fully integrated* adaptive finite element approach for the numerical treatment of the CH system with a non-smooth homogeneous free energy density was developed. The notion of a fully adaptive method relates to the fact that the local mesh adaptation is based on rigorous residual-based *a posteriori* error estimates (rather than heuristic techniques based on, e.g., thresholding the discrete concentration or the discrete concentration gradient). The concept of an integrated adaptation couples the adaptive cycle and the underlying solver as the latter might need to be equipped with additional stabilization methodologies such as the Moreau–Yosida regularization in the case of non-smooth homogeneous free energy densities for guaranteeing mesh independence. The latter is indeed obtained upon balancing regularization and discretization errors. When equipped with a multigrid scheme for solving the linear systems occurring in the underlying semi-smooth Newton iteration, an overall iterative scheme is obtained which is optimal in the sense that the computational effort grows only linearly in the number of degrees of freedom.

In Hintermüller et al. (2013), the approach of Hintermüller et al. (2011) is extended to a fully practical adaptive solver for the two-dimensional CHNS system with a double-obstacle potential according to Blowey and Elliott (1991). To the best of the applicant’s knowledge, the work of Hintermüller et al. (2013) contains the first rigorous approach to reliable and efficient residual-based *a posteriori* error analysis for multiphase flows governed by diffuse interface models. This approach is combined with a stable, energy-conserving time integration scheme in Garcke et al. (2016) to a fully reliable and efficient adaptive and energy-conserving *a posteriori* concept for the numerical treatment of variable density multiphase flows. This approach was very successfully validated against the existing sharp and diffuse interface rising benchmarks of Hysing et al. (2009) and Aland and Voigt (2012), respectively, in the field.

9.2 Diffuse Interface Approach

9.2.1 Notation

Let $\Omega \subset \mathbb{R}^n$, $n \in \{2, 3\}$ denote a bounded domain with boundary $\partial\Omega$ and unit outer normal ν_Ω . Let $I = (0, T]$ denote a time interval.

We use the conventional notation for Sobolev and Hilbert spaces; see, e.g., Adams and Fournier (2003). With $L^p(\Omega)$, $1 \leq p \leq \infty$, we denote the space of measurable functions on Ω , whose modulus to the power p is Lebesgue-integrable. $L^\infty(\Omega)$ denotes the space of measurable functions on Ω , which are essentially bounded. For $p = 2$, we denote by $L^2(\Omega)$ the space of square-integrable functions on Ω with inner product (\cdot, \cdot) and norm $\|\cdot\|$. For a subset $D \subset \Omega$ and functions $f, g \in L^2(\Omega)$, we by $(f, g)_D$ denote the inner product of f and g restricted to D , and by $\|f\|_D$ the respective norm. By $W^{k,p}(\Omega)$, $k \geq 1$, $1 \leq p \leq \infty$, we denote the Sobolev space of functions admitting weak derivatives up to order k in $L^p(\Omega)$. If $p = 2$, we write $H^k(\Omega)$. The subset $H_0^1(\Omega)$ denotes $H^1(\Omega)$ functions with vanishing boundary trace. We further set

$$L^2_{(0)}(\Omega) = \{v \in L^2(\Omega) \mid (v, 1) = 0\},$$

and with

$$H(\operatorname{div}, \Omega) = \{v \in H^1_0(\Omega)^n \mid (\operatorname{div}(v), q) = 0 \forall q \in L^2_{(0)}(\Omega)\}$$

we denote the space of all weakly solenoidal $H^1_0(\Omega)$ vector fields.

For vector fields $u \in L^q(\Omega)^n$, $q > n$, and $v, w \in H^1(\Omega)^n$, we introduce the trilinear form

$$a(u, v, w) = \frac{1}{2} \int_{\Omega} ((u \cdot \nabla) v) w \, dx - \frac{1}{2} \int_{\Omega} ((u \cdot \nabla) w) v \, dx.$$

Note that there holds $a(u, v, w) = -a(u, w, v)$, and especially $a(u, v, v) = 0$.

Further for matrices $A = (A_{ij})_{i,j=1}^n$, $B = (B_{ij})_{i,j=1}^n \in \mathbb{R}^{n \times n}$, we define $A : B := \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij}$.

9.2.2 The Mathematical Model

In the present work, we consider the following diffuse interface model for two-phase flows with variable densities proposed in Abels et al. (2012):

$$\rho \partial_t v + ((\rho v + J) \cdot \nabla) v - \operatorname{div}(2\eta Dv) + \nabla p = \mu \nabla \varphi + \rho g \quad \forall x \in \Omega, \forall t \in I, \quad (9.1)$$

$$\operatorname{div}(v) = 0 \quad \forall x \in \Omega, \forall t \in I, \quad (9.2)$$

$$\partial_t \varphi + v \cdot \nabla \varphi - \operatorname{div}(m \nabla \mu) = 0 \quad \forall x \in \Omega, \forall t \in I, \quad (9.3)$$

$$-\sigma \varepsilon \Delta \varphi + \frac{\sigma}{\varepsilon} F'(\varphi) - \mu = 0 \quad \forall x \in \Omega, \forall t \in I, \quad (9.4)$$

$$v(0, x) = v_0(x) \quad \forall x \in \Omega, \quad (9.5)$$

$$\varphi(0, x) = \varphi_0(x) \quad \forall x \in \Omega, \quad (9.6)$$

$$v(t, x) = 0 \quad \forall x \in \partial \Omega, \forall t \in I, \quad (9.7)$$

$$\nabla \mu(t, x) \cdot \nu_{\Omega} = \nabla \varphi(t, x) \cdot \nu_{\Omega} = 0 \quad \forall x \in \partial \Omega, \forall t \in I, \quad (9.8)$$

where $J = -\frac{d\rho}{d\varphi} m \nabla \mu$ describes an additional flux that appears due to the diffuse interface. Here, φ denotes the phase field, μ the chemical potential, v the volume averaged velocity, p the pressure, and $\rho = \rho(\varphi) = \frac{1}{2} ((\rho_2 - \rho_1)\varphi + (\rho_1 + \rho_2))$ the

mean density, where $0 < \rho_1 \leq \rho_2$ denote the densities of the involved fluids. The viscosity is denoted by $\eta = \eta(\varphi)$ and can be chosen as an arbitrary positive function fulfilling $\eta(-1) = \eta_1$ and $\eta(1) = \eta_2$, with individual fluid viscosities η_1, η_2 . The mobility is denoted by $m = m(\varphi)$. The gravitational force is denoted by g . Note that the material properties, i.e., the density and viscosity, are encoded by the phase field variable φ . The free energy density is denoted by F and is of double-well type with exactly two minima at ± 1 . By the structure of F , solutions to (9.1)–(9.8) fulfill that $\varphi \approx \pm 1$ in the air and water phase, but the model allows φ to take arbitrary values. For F , we use a splitting $F = F_+ + F_-$, where F_+ is convex and F_- is concave. Finally, by $Dv = \frac{1}{2}(\nabla v + (\nabla v)^t)$ we denote the symmetrized gradient. The scaled surface tension is denoted by σ and the interfacial width is proportional to ε .

The above model couples the Navier–Stokes equations (9.1)–(9.2) to the Cahn–Hilliard model (9.3)–(9.4) in a thermodynamically consistent way; i.e., a free energy inequality holds. In the Navier–Stokes equation, the term $\mu \nabla \varphi$ models the surface forces that appear due to surface tension. Further, an additional flux J appears due to the diffuse interface. The remaining part of the Navier–Stokes equations is standard, but note the spatially varying variables ρ and η . The Cahn–Hilliard equation contains a transport equation for the phase field, i.e., (9.3), while an additional flux $m \nabla \mu$ is present. This flux arises from the chemical potential μ and drives φ toward the desired phase field form. It is the main goal to introduce and analyze an (essentially) linear time discretization scheme for the numerical treatment of (9.1)–(9.8), which also on the discrete level fulfills the free energy inequality. This in conclusion leads to a stable scheme that is thermodynamically consistent on the discrete level.

Existence of weak solutions to system (9.1)–(9.8) for a specific class of free energies F is shown in Abels et al. (2013a, b). See also the work of Grün (2013), where the existence of weak solutions for a different class of free energies F is shown by passing to the limit in a numerical scheme. We refer to Aki et al. (2014), Boyer (2002), Ding et al. (2007), Lowengrub and Truskinovsky (1998), and the review Anderson et al. (1998) for other diffuse interface models for two-phase incompressible flow. Numerical approaches for different variants of the Navier–Stokes/Cahn–Hilliard system have been studied in Aland and Voigt (2012), Boyer (2002), Feng (2006), Grün (2013), Grün and Klingbeil (2014), Guo et al. (2014), Hintermüller et al. (2013), Grün et al. (2016), and Kay et al. (2008).

Our numerical treatment approach is based on the following weak formulation, which is proposed in Garcke et al. (2016).

Definition 9.1 (Garcke et al. 2016, Def. 1) We call v, φ, μ a weak solution to (9.1)–(9.8) if $v(0) = v_0, \varphi(0) = \varphi_0, v(t) \in H(\text{div}, \Omega)$ for a.e. $t \in I$ and

$$\begin{aligned} & \frac{1}{2} \int_{\Omega} (\partial_t(\rho v) + \rho \partial_t v) w \, dx + \int_{\Omega} 2\eta Dv : Dw \, dx \\ & + a(\rho v + J, v, w) = \int_{\Omega} \mu \nabla \varphi w + \rho g w \, dx \quad \forall w \in H(\text{div}, \Omega), \end{aligned} \tag{9.9}$$

$$\int_{\Omega} (\partial_t \varphi + v \cdot \nabla \varphi) \Phi \, dx + \int_{\Omega} m(\varphi) \nabla \mu \cdot \nabla \Phi \, dx = 0 \quad \forall \Phi \in H^1(\Omega), \tag{9.10}$$

$$\sigma \varepsilon \int_{\Omega} \nabla \varphi \cdot \nabla \Psi \, dx + \frac{\sigma}{\varepsilon} \int_{\Omega} F'(\varphi) \Psi \, dx - \int_{\Omega} \mu \Psi \, dx = 0 \quad \forall \Psi \in H^1(\Omega), \tag{9.11}$$

is satisfied for a.e. $t \in I$.

We note that we restrict to solenoidal velocity fields $v(t) \in H(\text{div}, \Omega)$ which allows us to skip the pressure. Given $v(t) \in H(\text{div}, \Omega)$, the pressure can be uniquely reconstructed by standard results; see, e.g., Constantin and Foias (1988).

For the assumptions on the data, we refer to Garcke et al. (2016). In the present work, we use the relaxed double-obstacle free energy given by

$$F(\varphi) = \frac{1}{2} (1 - \varphi^2 + s\lambda^2(\varphi)), \tag{9.12}$$

with

$$\lambda(\varphi) := \max(0, \varphi - 1) + \min(0, \varphi + 1),$$

where $s \gg 0$ denotes the relaxation parameter. F is introduced in Hintermüller et al. (2011) as Moreau–Yosida relaxation of the double-obstacle free energy

$$F^{obst}(\varphi) = \begin{cases} \frac{1}{2} (1 - \varphi^2) & \text{if } |\varphi| \leq 1, \\ 0 & \text{else,} \end{cases}$$

which is proposed in Blowey and Elliott (1991) to model phase separation.

Let v, φ, μ be a sufficiently smooth solution to (9.9)–(9.11). Then, we have from Garcke et al. (2016) the energy relation

$$\frac{d}{dt} \left(\int_{\Omega} \frac{\rho}{2} |v|^2 + \frac{\sigma \varepsilon}{2} |\nabla \varphi|^2 + \frac{\sigma}{\varepsilon} F(\varphi) \, dx \right) = - \int_{\Omega} 2\eta |Dv|^2 + m |\nabla \mu|^2 \, dx + \int_{\Omega} \rho g v \, dx. \tag{9.13}$$

9.3 Discretization

9.3.1 The Temporal Discretization

We now introduce a time discretization which mimics the energy inequality in (9.13) on the discrete level. Let $0 = t_0 < t_1 < \dots < t_{k-1} < t_k < t_{k+1} < \dots < t_M = T$

denote an equidistant subdivision of the interval $\bar{I} = [0, T]$ with $t_{k+1} - t_k = \tau$. From here onward, the superscript k denotes the corresponding variables at time instance t_k .

Time integration scheme (Garcke et al. 2016)

Let $\varphi_0 \in H^1(\Omega)$ and $v_0 \in H(\text{div}, \Omega)$.

Initialization for $k = 0$:

Set $\varphi^0 = \varphi_0$ and $v^0 = v_0$.

Find $\varphi^1 \in H^1(\Omega)$, $\mu^1 \in H^1(\Omega)$, $v^1 \in H(\text{div}, \Omega)$, such that for all $w \in H(\text{div}, \Omega)$, $\Phi \in H^1(\Omega)$, and $\Psi \in H^1(\Omega)$ it holds

$$\begin{aligned} \frac{1}{\tau} \int_{\Omega} \rho^1 (v^1 - v^0) w \, dx + \int_{\Omega} ((\rho^0 v^0 + J^1) \cdot \nabla) v^1 \cdot w \, dx \\ + \int_{\Omega} 2\eta^1 Dv^1 : Dw \, dx - \int_{\Omega} \mu^1 \nabla \varphi^1 w + \rho^1 g w \, dx = 0 \quad \forall w \in H(\text{div}, \Omega), \end{aligned} \quad (9.14)$$

$$\begin{aligned} \frac{1}{\tau} \int_{\Omega} (\varphi^1 - \varphi^0) \Phi \, dx + \int_{\Omega} (v^0 \cdot \nabla \varphi^0) \Phi \, dx \\ + \int_{\Omega} m(\varphi^0) \nabla \mu^1 \cdot \nabla \Phi \, dx = 0 \quad \forall \Phi \in H^1(\Omega), \end{aligned} \quad (9.15)$$

$$\begin{aligned} \sigma \varepsilon \int_{\Omega} \nabla \varphi^1 \cdot \nabla \Psi \, dx - \int_{\Omega} \mu^1 \Psi \, dx \\ + \frac{\sigma}{\varepsilon} \int_{\Omega} ((F_+)'(\varphi^1) + (F_-)'(\varphi^0)) \Psi \, dx = 0 \quad \forall \Psi \in H^1(\Omega), \end{aligned} \quad (9.16)$$

where $J^1 := -\frac{d\rho}{d\varphi}(\varphi^1) m^1 \nabla \mu^1$.

Two-step scheme for $k \geq 1$:

Given $\varphi^{k-1} \in H^1(\Omega)$, $\varphi^k \in H^1(\Omega)$, $\mu^k \in W^{1,q}(\Omega)$, $q > n$, $v^k \in H(\text{div}, \Omega)$, find $v^{k+1} \in H(\text{div}, \Omega)$, $\varphi^{k+1} \in H^1(\Omega)$, $\mu^{k+1} \in H^1(\Omega)$ satisfying

$$\begin{aligned} \frac{1}{2\tau} \int_{\Omega} (\rho^k v^{k+1} - \rho^{k-1} v^k) w + \rho^{k-1} (v^{k+1} - v^k) w \, dx \\ + a(\rho^k v^k + J^k, v^{k+1}, w) + \int_{\Omega} 2\eta^k Dv^{k+1} : Dw \, dx \\ - \int_{\Omega} \mu^{k+1} \nabla \varphi^k w - \rho^k g w \, dx = 0 \quad \forall w \in H(\text{div}, \Omega), \end{aligned} \quad (9.17)$$

$$\begin{aligned} \frac{1}{\tau} \int_{\Omega} (\varphi^{k+1} - \varphi^k) \Phi \, dx + \int_{\Omega} (v^{k+1} \cdot \nabla \varphi^k) \Phi \, dx \\ + \int_{\Omega} m(\varphi^k) \nabla \mu^{k+1} \cdot \nabla \Phi \, dx = 0 \quad \forall \Phi \in H^1(\Omega), \end{aligned} \quad (9.18)$$

$$\begin{aligned} \sigma \varepsilon \int_{\Omega} \nabla \varphi^{k+1} \cdot \nabla \Psi \, dx - \int_{\Omega} \mu^{k+1} \Psi \, dx \\ + \frac{\sigma}{\varepsilon} \int_{\Omega} ((F_+)'(\varphi^{k+1}) + (F_-)'(\varphi^k)) \Psi \, dx = 0 \quad \forall \Psi \in H^1(\Omega), \end{aligned} \quad (9.19)$$

where $J^k := -\frac{d\rho}{d\varphi}(\varphi^k) m^k \nabla \mu^k$.

We note that in (9.17)–(9.19) the only non-linearity arises from F'_+ , and thus, only the equation (9.19) is non-linear. For a discussion of this scheme, we refer to Garcke et al. (2016). Grün and Klingbeil (2014) propose a time discrete solver for (9.1)–(9.8) which leads to strongly coupled systems for v , φ , and p at every time step and requires a fully non-linear solver. For this scheme, Grün (2013) proves an energy inequality and the existence of so-called generalized solutions.

9.3.2 The Spatial Discretization and Energy Inequalities

For a numerical treatment, we next discretize the weak formulation (9.17)–(9.19) in space. We aim at an adaptive discretization of the domain Ω , and thus to have a different spatial discretization in every time step.

Let $\mathcal{T}^k = \bigcup_{i=1}^{NT} T_i$ denote a conforming triangulation of $\overline{\Omega}$ with closed simplices T_i , $i = 1, \dots, NT$ and edges E_i , $i = 1, \dots, NE$, $\mathcal{E}^k = \bigcup_{i=1}^{NE} E_i$. Here, k refers to the time instance t_k . On \mathcal{T}^k , we define the following finite element spaces:

$$\begin{aligned} \mathcal{V}^1(\mathcal{T}^k) &= \{v \in C(\mathcal{T}^k) \mid v|_T \in P^1(T) \forall T \in \mathcal{T}^k\} =: \text{span}\{\Phi^i\}_{i=1}^{NP}, \\ \mathcal{V}^2(\mathcal{T}^k) &= \{v \in C(\mathcal{T}^k) \mid v|_T \in P^2(T) \forall T \in \mathcal{T}^k\}, \end{aligned}$$

where $P^l(S)$ denotes the space of polynomials up to order l defined on S .

We introduce the discrete analogon to the space $H(\text{div}, \Omega)$:

$$\begin{aligned} H(\text{div}, \mathcal{T}^k) &= \{v \in \mathcal{V}^2(\mathcal{T}^k)^n \mid (\text{div} v, q) = 0 \forall q \in \mathcal{V}^1(\mathcal{T}^k) \cap L^2_{(0)}(\Omega), v|_{\partial\Omega} = 0\} \\ &:= \text{span}\{b^i\}_{i=1}^{NF}, \end{aligned}$$

We further introduce a H^1 -stable projection operator $\mathcal{P}^k : H^1(\Omega) \rightarrow \mathcal{V}^1(\mathcal{T}^k)$ satisfying

$$\|\mathcal{P}^k v\|_{L^p(\Omega)} \leq \|v\|_{L^p(\Omega)} \quad \text{and} \quad \|\nabla \mathcal{P}^k v\|_{L^r(\Omega)} \leq \|\nabla v\|_{L^r(\Omega)}$$

for $v \in H^1(\Omega)$ with $r \in [1, 2]$ and $p \in [1, 6]$ if $n = 3$, and $p \in [1, \infty)$ if $n = 2$. Possible choices are the H^1 -projection, the Clément operator (Clément 1975) or, by restricting the preimage to $C(\overline{\Omega}) \cap H^1(\Omega)$, the Lagrangian interpolation operator.

Using these spaces, we state the discrete counterpart of (9.17)–(9.19):

Let $k \geq 1$, given $\varphi^{k-1} \in \mathcal{V}^1(\mathcal{T}^{k-1})$, $\varphi^k \in \mathcal{V}^1(\mathcal{T}^k)$, $\mu^k \in \mathcal{V}^1(\mathcal{T}^k)$, $v^k \in H(\operatorname{div}, \mathcal{T}^k)$, find $v_h^{k+1} \in H(\operatorname{div}, \mathcal{T}^{k+1})$, $\varphi_h^{k+1} \in \mathcal{V}^1(\mathcal{T}^{k+1})$, $\mu_h^{k+1} \in \mathcal{V}^1(\mathcal{T}^{k+1})$ such that for all $w \in H(\operatorname{div}, \mathcal{T}^{k+1})$, $\Phi \in \mathcal{V}^1(\mathcal{T}^{k+1})$, $\Psi \in \mathcal{V}^1(\mathcal{T}^{k+1})$ there holds:

$$\begin{aligned} \frac{1}{2\tau}(\rho^k v_h^{k+1} - \rho^{k-1} v^k + \rho^{k-1}(v_h^{k+1} - v^k), w) + a(\rho^k v^k + J^k, v_h^{k+1}, w) \\ + (2\eta^k Dv_h^{k+1}, Dw) - (\mu_h^{k+1} \nabla \varphi^k + \rho^k g, w) = 0, \end{aligned} \quad (9.20)$$

$$\frac{1}{\tau}(\varphi_h^{k+1} - \mathcal{P}^{k+1} \varphi^k, \Phi) + (m(\varphi^k) \nabla \mu_h^{k+1}, \nabla \Phi) + (v_h^{k+1} \nabla \varphi^k, \Phi) = 0, \quad (9.21)$$

$$\sigma \varepsilon (\nabla \varphi_h^{k+1}, \nabla \Psi) + \frac{\sigma}{\varepsilon} (F'_+(\varphi_h^{k+1}) + F'_-(\mathcal{P}^{k+1} \varphi^k), \Psi) - (\mu_h^{k+1}, \Psi) = 0, \quad (9.22)$$

where $\varphi^0 = P\varphi_0$ denotes the L^2 projection of φ_0 in $\mathcal{V}^1(\mathcal{T}^0)$, $v^0 = P^L v_0$ denotes the Leray projection of v_0 in $H(\operatorname{div}, \mathcal{T}^0)$ (see Constantin and Foias 1988), and $\varphi_h^1, \mu_h^1, v_h^1$ are obtained from the fully discrete variant of (9.14)–(9.16).

We have from Garcke et al. (2016) that the fully discrete system (9.20)–(9.22) admits a unique solution, where the analysis crucially depends on an energy inequality for the solution $(\varphi_h^{k+1}, \mu_h^{k+1}, v_h^{k+1})$ of (9.20)–(9.22). The energy inequality for one time instance, which is also proven in Garcke et al. (2016), is given as:

For $k \geq 1$:

$$\begin{aligned} & \frac{1}{2} \int_{\Omega} \rho^k |v_h^{k+1}|^2 dx + \frac{\sigma \varepsilon}{2} \int_{\Omega} |\nabla \varphi_h^{k+1}|^2 dx + \frac{\sigma}{\varepsilon} \int_{\Omega} F(\varphi_h^{k+1}) dx \\ & + \frac{1}{2} \int_{\Omega} \rho^{k-1} |v_h^{k+1} - v^k|^2 dx + \frac{\sigma \varepsilon}{2} \int_{\Omega} |\nabla \varphi_h^{k+1} - \nabla \mathcal{P}^{k+1} \varphi^k|^2 dx \\ & + \tau \int_{\Omega} 2\eta^k |Dv_h^{k+1}|^2 dx + \tau \int_{\Omega} m^k |\nabla \mu_h^{k+1}|^2 dx \\ & \leq \frac{1}{2} \int_{\Omega} \rho^{k-1} |v^k|^2 dx + \frac{\sigma \varepsilon}{2} \int_{\Omega} |\nabla \mathcal{P}^{k+1} \varphi^k|^2 dx \\ & + \frac{\sigma}{\varepsilon} \int_{\Omega} F(\mathcal{P}^{k+1} \varphi^k) dx + \tau \int_{\Omega} \rho^k g v_h^{k+1}. \end{aligned} \quad (9.23)$$

In (9.23), the Ginzburg–Landau energy of the current phase field φ^{k+1} is estimated against the Ginzburg–Landau energy of the projection of the old phase field $\mathcal{P}^{k+1}(\varphi^k)$. Since our aim is to obtain global in time inequalities estimating the energy of the new phase field against the energy of the old phase field at each time step, we assume

Assumption 1 Let $\varphi^k \in \mathcal{V}^1(\mathcal{T}^k)$ denote the phase field at time instance t_k . Let $\mathcal{P}^{k+1}\varphi^k \in \mathcal{V}^1(\mathcal{T}^{k+1})$ denote the projection of φ^k in $\mathcal{V}^1(\mathcal{T}^{k+1})$. We assume that there holds

$$\frac{\sigma}{\varepsilon} F(\mathcal{P}^{k+1}\varphi^k) + \frac{1}{2}\sigma\varepsilon|\nabla\mathcal{P}^{k+1}\varphi^k|^2 \leq \frac{\sigma}{\varepsilon} F(\varphi^k) + \frac{1}{2}\sigma\varepsilon|\nabla\varphi^k|^2. \quad (9.24)$$

This assumption means that the Ginzburg–Landau energy is not increasing through projection. Thus, no energy is numerically produced.

Assumption 1 is in general not fulfilled for arbitrary sequences (\mathcal{T}^{k+1}) of triangulations. To ensure (9.24), a postprocessing step can be added to the adaptive space meshing; see Section 9.3.3.

With this assumption we immediately get

Theorem 9.1 (Garcke et al. 2016, Thm. 7) *Assume that for every $k = 0, 1, \dots$, Assumption (1) holds. Then, for every $1 \leq k < l$ we have*

$$\begin{aligned} & \frac{1}{2}(\rho_h^{k-1}v_h^k, v_h^k) + \frac{\sigma}{\varepsilon} \int_{\Omega} F(\varphi_h^k) dx + \frac{1}{2}\sigma\varepsilon(\nabla\varphi_h^k, \nabla\varphi_h^k) + \tau \sum_{m=k}^{l-1} (\rho^m g, v_h^{m+1}) \\ & \geq \frac{1}{2}(\rho^{l-1}v_h^l, v_h^l) + \frac{\sigma}{\varepsilon} \int_{\Omega} F(\varphi_h^l) dx + \frac{1}{2}\sigma\varepsilon(\nabla\varphi_h^l, \nabla\varphi_h^l) \\ & \quad + \sum_{m=k}^{l-1} (\rho^{m-1}(v_h^{m+1} - v_h^m), (v_h^{m+1} - v_h^m)) \\ & \quad + \tau \sum_{m=k}^{l-1} (2\eta^m Dv_h^{m+1}, Dv_h^{m+1}) \\ & \quad + \tau \sum_{m=k}^{l-1} (m(\varphi_h^m) \nabla\mu_h^{m+1}, \nabla\mu_h^{m+1}) \\ & \quad + \frac{1}{2}\sigma\varepsilon \sum_{m=k}^{l-1} (\nabla\varphi_h^{m+1} - \nabla\mathcal{P}^{m+1}\varphi_h^m, \nabla\varphi_h^{m+1} - \nabla\mathcal{P}^{m+1}\varphi_h^m). \end{aligned}$$

In our presentation, F denotes the relaxed double-obstacle free energy depending on the relaxation parameter s . Let $(v_s, \varphi_s, \mu_s)_{s \in \mathbb{R}}$ denote the sequence of solutions of (9.17)–(9.19) for a sequence $(s_l)_{l \in \mathbb{N}}$. Then, we are able to argue convergence to solutions of a limit system related to the double-obstacle free energy F^{obst} . More specifically, from the linearity of (9.17) and (Hintermüller et al. 2011, Prop. 4.2) we conclude that there exists a subsequence, still denoted by $(v_s, \varphi_s, \mu_s)_{s \in \mathbb{R}}$, such that

$$(v_s, \varphi_s, \mu_s)_{s \in \mathbb{R}} \rightarrow (v^*, \varphi^*, \mu^*) \quad \text{in } H^1(\Omega),$$

where (v^*, φ^*, μ^*) denotes the solution of (9.17)–(9.19), where F^{obst} is chosen as free energy. Especially $|\varphi^*| \leq 1$ holds. Details are given in Garcke et al. (2016).

9.3.3 *A posteriori Error Estimation*

For an efficient solution of (9.20)–(9.22), we next describe an a posteriori error estimator-based mesh refinement scheme that is reliable and efficient up to terms of higher order and errors introduced by the projection. We also describe how Assumption 1 on the evolution of the free energy, given in (9.23), under projection is fulfilled in the discrete setting.

Let us briefly comment on available adaptive concepts for the spatial discretization of Cahn–Hilliard/Navier–Stokes systems. Heuristic approaches exploiting knowledge of the location of the diffuse interface can be found in Kay et al. (2008), Aland and Voigt (2012), Grün and Klingbeil (2014). In Hintermüller et al. (2013), a fully adaptive, reliable and efficient, residual-based error estimator for the Cahn–Hilliard part in the Cahn–Hilliard/Navier–Stokes system is proposed, which extends the results of Hintermüller et al. (2011) for Cahn–Hilliard to Cahn–Hilliard/Navier–Stokes systems with Moreau–Yosida relaxation of the double-obstacle free energy. A residual-based error estimator for Cahn–Hilliard systems with double-obstacle free energy is proposed in Bañas and Nürnberg (2009).

In the present section, we propose a fully integrated adaptive concept for the fully coupled Cahn–Hilliard/Navier–Stokes system, where we exploit the energy inequality of (9.13).

For the numerical realization, we switch to the primitive setting for the flow part of our equation system. The corresponding fully discrete system now reads:

For $k \geq 1$, given $\varphi^{k-1} \in H^1(\Omega)$, $\varphi^k \in H^1(\Omega)$, $\mu^k \in W^{1,q}(\Omega)$, $q > n$, $v^k \in H_0^1(\Omega)^n$ find $v_h^{k+1} \in \mathcal{V}^2(\mathcal{T}^{k+1})$, $p_h^{k+1} \in \mathcal{V}^1(\mathcal{T}^{k+1})$, $\int_{\Omega} p_h^{k+1} dx = 0$, $\varphi_h^{k+1} \in \mathcal{V}^1(\mathcal{T}^{k+1})$, $\mu_h^{k+1} \in \mathcal{V}^1(\mathcal{T}^{k+1})$ such that for all $w \in \mathcal{V}^2(\mathcal{T}^{k+1})$, $q \in \mathcal{V}^1(\mathcal{T}^{k+1})$, $\Phi \in \mathcal{V}^1(\mathcal{T}^{k+1})$, $\Psi \in \mathcal{V}^1(\mathcal{T}^{k+1})$ there holds:

$$\begin{aligned} \frac{1}{2\tau}(\rho^k v_h^{k+1} - \rho^{k-1} v^k + \rho^{k-1}(v_h^{k+1} - v^k), w) + a(\rho^k v^k + J^k, v_h^{k+1}, w) \\ + (2\eta^k Dv_h^{k+1}, \nabla w) - (\mu_h^{k+1} \nabla \varphi^k + \rho^k g, w) - (p_h^{k+1}, \operatorname{div} w) = 0, \end{aligned} \quad (9.25)$$

$$-(\operatorname{div} v_h^{k+1}, q) = 0, \quad (9.26)$$

$$\frac{1}{\tau}(\varphi_h^{k+1} - \mathcal{P}^{k+1} \varphi^k, \Phi) + (m(\varphi^k) \nabla \mu_h^{k+1}, \nabla \Phi) - (v_h^{k+1} \varphi^k, \nabla \Phi) = 0, \quad (9.27)$$

$$\sigma \varepsilon(\nabla \varphi_h^{k+1}, \nabla \Psi) + \frac{\sigma}{\varepsilon}(F'_+(\varphi_h^{k+1}) + F'_-(\mathcal{P}^{k+1} \varphi^k), \Psi) - (\mu_h^{k+1}, \Psi) = 0. \quad (9.28)$$

Thus, we use the famous Taylor–Hood LBB-stable $P2 - P1$ finite element for the discretization of the velocity–pressure field and piecewise linear and continuous finite elements for the discretization of the phase field and the chemical potential.

For other kinds of possible discretizations of the velocity–pressure field, we refer to, e.g., Verfürth (2010).

Note that we perform integration by parts in (9.27) in the transport term, using the no-slip boundary condition for v_h^{k+1} . As soon as \mathcal{P}^{k+1} is a mass-conserving projection, we by testing equation (9.27) with $\Phi = 1$ obtain the conservation of mass in the fully discrete scheme.

The link between equations (9.25)–(9.28) and (9.20)–(9.22) is established by the fact that for $v_h^{k+1}, \varphi_h^{k+1}, \mu_h^{k+1}$ denoting the unique solution to (9.20)–(9.22), there exists a unique pressure $p_h^{k+1} \in \mathcal{V}^1(\mathcal{T}^{k+1})$ satisfying $\int_{\Omega} p_h^{k+1} dx = 0$, such that $(v_h^{k+1}, p_h^{k+1}, \varphi_h^{k+1}, \mu_h^{k+1})$ is a solution to (9.25)–(9.28). The opposite direction is obvious.

Next, we describe the error estimator which we use in our computations. We follow Hintermüller et al. (2011) and restrict the presentation of its construction to the main steps.

We define the following error terms:

$$\begin{aligned} e_v &:= v_h^{k+1} - v^{k+1}, & e_p &:= p_h^{k+1} - p^{k+1}, \\ e_\varphi &:= \varphi_h^{k+1} - \varphi^{k+1}, & e_\mu &:= \mu_h^{k+1} - \mu^{k+1}, \end{aligned}$$

as well as the discrete element residuals

$$\begin{aligned} r_h^{(1)} &:= \frac{\rho^k + \rho^{k-1}}{2} v_h^{k+1} - \rho^{k-1} v^k + \tau (b^k \nabla) v_h^{k+1} + \frac{1}{2} \tau \operatorname{div}(b^k) v_h^{k+1} \\ &\quad - 2\tau \operatorname{div}(\eta^k Dv_h^{k+1}) + \tau \nabla p_h^{k+1} - \tau \mu_h^{k+1} \nabla \varphi^k - \rho^k g, \\ r_h^{(2)} &:= \varphi_h^{k+1} - \mathcal{P}^{k+1} \varphi^k + \tau v_h^{k+1} \nabla \varphi^k - \tau \operatorname{div}(m^k \nabla \mu_h^{k+1}), \\ r_h^{(3)} &:= \frac{\sigma}{\varepsilon} F'_+(\varphi_h^{k+1}) + \frac{\sigma}{\varepsilon} F'_-(\mathcal{P}^{k+1} \varphi^k) - \mu_h^{k+1}, \end{aligned}$$

where $b^k := \rho^k v^k + J^k$. Furthermore, we define the error indicators

$$\begin{aligned} \eta_T^{(1)} &:= h_T \|r_h^{(1)}\|_T, & \eta_E^{(1)} &:= h_E^{1/2} \|2\eta^k [Dv_h^{k+1}]_v\|_E, \\ \eta_T^{(2)} &:= h_T \|r_h^{(2)}\|_T, & \eta_E^{(2)} &:= h_E^{1/2} \|m^k [\nabla \mu_h^{k+1}]_v\|_E, \\ \eta_T^{(3)} &:= h_T \|r_h^{(3)}\|_T, & \eta_E^{(3)} &:= h_E^{1/2} \|[\nabla \varphi_h^{k+1}]_v\|_E. \end{aligned} \tag{9.29}$$

Here, $[\cdot]_v$ denotes the jump of a discontinuous function in normal direction v pointing from the triangle with lower global number to the triangle with higher global number. Thus, $\eta_E^{(j)}$, $j = 1, 2, 3$ measures the jump of the corresponding variable across the edge E , while $\eta_T^{(j)}$, $j = 1, 2, 3$ measures the trianglewise residuals.

In (Garcke et al. 2016, Thm. 99), the following theorem is proven.

Theorem 9.2 (Garcke et al. 2016, Thm. 9) *There exists a constant $C > 0$ only depending on the domain Ω and the regularity of the mesh \mathcal{T}^{k+1} such that*

$$\begin{aligned} \underline{\rho} \|e_v\|^2 + \tau \underline{\eta} \|\nabla e_v\|^2 + \tau \underline{m} \|\nabla e_\mu\|^2 + \sigma \varepsilon \|\nabla e_\varphi\|^2 + \frac{\sigma}{\varepsilon} (F'_+(\varphi_h^{k+1}) - F'_+(\varphi^{k+1}), e_\varphi) \\ \leq C (\eta_\Omega^2 + \eta_{h.o.t} + \eta_C), \end{aligned}$$

holds with

$$\begin{aligned} \eta_\Omega^2 &= \frac{1}{\tau \underline{\eta}} \sum_{T \in \mathcal{T}^{k+1}} \left(\eta_T^{(1)} \right)^2 + \frac{\tau}{\underline{\eta}} \sum_{E \in \mathcal{E}^{k+1}} \left(\eta_E^{(1)} \right)^2 \\ &\quad + \frac{1}{\tau \underline{m}} \sum_{T \in \mathcal{T}^{k+1}} \left(\eta_T^{(2)} \right)^2 + \frac{\tau}{\underline{m}} \sum_{E \in \mathcal{E}^{k+1}} \left(\eta_E^{(2)} \right)^2 \\ &\quad + \frac{1}{\sigma \varepsilon} \sum_{T \in \mathcal{T}^{k+1}} \left(\eta_T^{(3)} \right)^2 + \sigma \varepsilon \sum_{E \in \mathcal{E}^{k+1}} \left(\eta_E^{(3)} \right)^2, \\ \eta_{h.o.t.} &= \tau (\operatorname{div}(e_v), e_p), \\ \text{and } \eta_C &= (\mathcal{P}^{k+1} \varphi^k - \varphi^k, e_\mu) - \frac{\sigma}{\varepsilon} (F'_-(\mathcal{P}^{k+1} \varphi^k) - F'_-(\varphi^k), e_\varphi). \end{aligned}$$

Here, $\underline{\rho}$, $\underline{\eta}$, \underline{m} denote lower bounds for $\rho(\varphi)$, $\eta(\varphi)$, and $m(\varphi)$.

In the numerical part, this error estimator is used together with the mesh adaptation cycle described in Hintermüller et al. (2011). The overall adaptation cycle

SOLVE \rightarrow ESTIMATE \rightarrow MARK \rightarrow ADAPT

is performed once per time step. For convenience of the reader, we state the marking strategy here.

Algorithm 2 (Marking strategy)

- Fix $a_{\min} > 0$ and $a_{\max} > 0$, and set $\mathcal{A} = \{T \in \mathcal{T}^{k+1} \mid a_{\min} \leq |T| \leq a_{\max}\}$.
- Define indicators:

1. $\eta_T = \frac{1}{\tau \underline{\eta}} \left(\eta_T^{(1)} \right)^2 + \frac{1}{\tau \underline{m}} \left(\eta_T^{(2)} \right)^2 + \frac{1}{\sigma \varepsilon} \left(\eta_T^{(3)} \right)^2$,
2. $\eta_{TE} = \sum_{E \subset T} \left[\frac{\tau}{\underline{\eta}} \left(\eta_{TE}^{(1)} \right)^2 + \frac{\tau}{\underline{m}} \left(\eta_{TE}^{(2)} \right)^2 + \sigma \varepsilon \left(\eta_{TE}^{(3)} \right)^2 \right]$.

- Refinement: Choose $\theta^r \in (0, 1)$,

1. Find a set $R^T \subset \mathcal{T}^{k+1}$ with $\theta^r \sum_{T \in \mathcal{T}^{k+1}} \eta_T \leq \sum_{T \in R^T} \eta_T$,
2. Find a set $R^{TE} \subset \mathcal{T}^{k+1}$ with $\theta^r \sum_{T \in \mathcal{T}^{k+1}} \eta_{TE} \leq \sum_{T \in R^{TE}} \eta_{TE}$.

- Coarsening: Choose $\theta^c \in (0, 1)$,

1. Find the set $C^T \subset \mathcal{T}^{k+1}$ with $\eta_T \leq \frac{\theta^c}{N} \sum_{T \in \mathcal{T}^{k+1}} \eta_T \forall T \in C^T$,
2. Find the set $C^{TE} \subset \mathcal{T}^{k+1}$ with $\eta_{TE} \leq \frac{\theta^c}{N} \sum_{T \in \mathcal{T}^{k+1}} \eta_{TE} \forall T \in C^{TE}$.

- Mark all triangles of $\mathcal{A} \cap (R^T \cup R^{TE})$ for refining.
- Mark all triangles of $\mathcal{A} \cap (C^T \cup C^{TE})$ for coarsening.

Ensuring the validity of the energy estimate

To ensure the validity of the energy estimate during the numerical computations, we ensure that Assumption 1 holds trianglewise. For the following considerations, we restrict to bisection as refinement strategy combined with the *i*FEM coarsening strategy proposed in Chen (2008). This strategy only coarsens patches consisting of four triangles by replacing them by two triangles if the central node of the patch is an inner node of \mathcal{T}^{k+1} and patches consisting of two triangles by replacing them by one triangle if the central node of the patch lies on the boundary of Ω . A patch fulfilling one of these two conditions we call a nodeStar. By using this strategy, we do not harm the Assumption 1 on triangles that are refined. We note that this assumption can only be violated on patches of triangles where coarsening appears.

After marking triangles for refinement and coarsening and before applying refinement and coarsening to \mathcal{T}^{k+1} , we make a postprocessing of all triangles that are marked for coarsening.

Let M^C denote the set of triangles marked for coarsening obtained by the marking strategy described in Algorithm 2. To ensure the validity of the energy estimate (9.23), we perform the following postprocessing steps:

Algorithm 3 (Postprocessing)

1. For each triangle $T \in M^C$:
if T is not part of a nodeStar
then set $M^C := M^C \setminus T$.
2. For each nodeStar $S \in M^C$:
if Assumption (1) is not fulfilled on S
then set $M^C := M^C \setminus S$.

The resulting set M^C does only contain triangles yielding nodeStars on which the Assumption 1 is fulfilled.

9.4 Numerics

Let us finally give a numerical example to show the applicability of the provided method to the simulation of the complex interaction at the air-water interface. We use the implementation from Garcke et al. (2016) that was developed for the validation of results for the rising bubble benchmark from (Aland and Voigt 2012; Hysing et al. 2009). For this reason, it is not adapted to the present situation and we will comment on the restrictions and the future work to tackle the given problem after showing the numerical results.

We use $\Omega = (0.0, 3.0) \times (0.0, 1.0)$ and a time horizon of $I = (0, 10.0)$ that we subdivide into equidistant time steps of length $\tau = 5e - 4$. To mimic the wind forc-

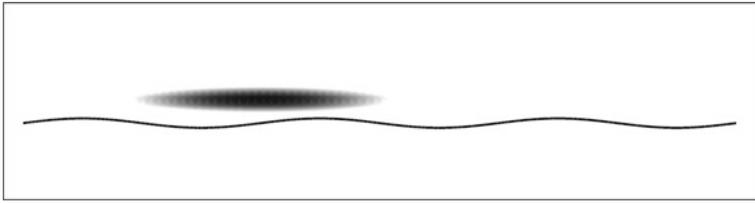


Fig. 9.1 Volume force f that generates the ‘wind’ together with the zero level line of φ_0

ing, we introduce a volume force $f = (f_1, f_2)^\top = (f_1, 0)^\top$ as shown in Figure 9.1 on the right-hand side of the Navier–Stokes equation that is defined as

$$(f_1(x)) = \begin{cases} 0 & \text{if } \|(x - m)/\sigma\| \geq 1, \\ \cos(\pi \|(x - m)/\sigma\|)^2 & \text{else,} \end{cases}$$

where $m = (1.0, 1.2)^\top$, $\sigma = (1.0, 0.1)^\top$, and the division $(x - m)/\sigma$ has to be understood componentwise. This is an approximation of a Gaussian bell with compact support. In Figure 9.1, we further show the zero-level line of the initial phase field φ_0 that is given by

$$\begin{aligned} z(x_1, x_2) &:= (x_2 - 0.02 \sin(2\pi x_1) + 0.2)/\varepsilon, \\ z_0 &= \arctan \sqrt{s - 1}, \\ \varphi_0(z) &:= \begin{cases} \sqrt{\frac{s}{s-1}} \sin(z) & \text{if } |z| \leq z_0, \\ \frac{1}{s-1} (s - \exp(\sqrt{s-1}(z_0 - z))) & \text{if } z > z_0, \\ -\frac{1}{s-1} (s - \exp(\sqrt{s-1}(z_0 + z))) & \text{if } z < -z_0. \end{cases} \end{aligned}$$

Note that z measures the distance in x_2 direction to the wave $0.2 - 0.02 \sin(2\pi x_1)$ scaled by ε^{-1} and $\varphi_0(z)$ is the first-order approximation to a phase field with relaxed double-obstacle free energy (9.12); see (Kahle 2014, Section 10). The initial velocity is $v_0 \equiv 0$, and we have no-slip boundary data at $\partial\Omega$.

As parameters we choose $\rho_{water} = 1$, $\eta_{water} = 0.01$, $\sigma_{water} = 0.0005$. Using unit velocity $V = 1$ and unit length $d = 1$, this results in a Weber number of $We = 2000$, and after a required scaling due to the chosen free energy, see Abels et al. (2012), we have $\sigma = 0.00032$. The gravity is $g = (0.0, -9.81)^\top$, and the mobility is $b = \frac{\varepsilon}{500\sigma}$. We note that especially the chosen density ρ_{water} does not correspond to the real-world parameter. We use the air–water ratio $\rho_{air} = 0.01\rho_{water}$ and $\eta_{air} = 0.01\eta_{water}$, which is ten times larger than the real-world ratio. To overcome the limitations of the current implementation with respect to the density ratio is subject of future research.

For the adaptation process from Algorithm 2, we choose $\theta^r = 0.5$, $\theta^c = 0.01$, $V_{\min} = 8e - 6$, $V_{\max} = 3e - 4$. This results in about 44000 cells for the spatial discretization, where the interface is resolved by 16 cells in orthogonal direction.

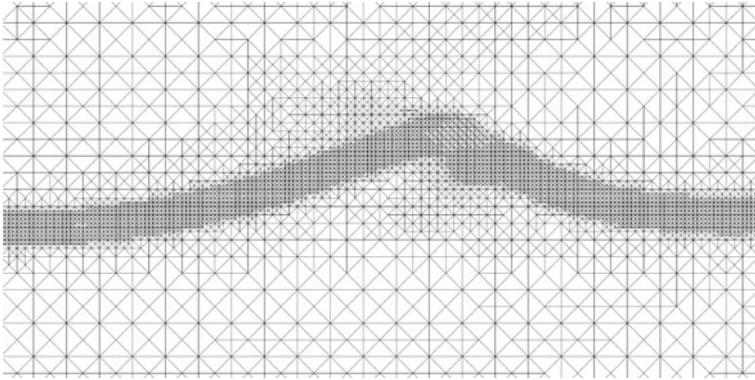


Fig. 9.2 Distribution of the elements at final time. Note the high resolution close to the surface layer, while the elements are larger in the pure phases. We only show a part of the computational domain, to be able to sufficiently resolve the elements in the picture

This corresponds to cells with an area of $|T| \approx 4e - 6$ inside the interface, while in the ocean and air we obtain $|T| \approx 3e - 4$. For the numerical results presented, we switched off the postprocessing proposed in Algorithm 3. In Garcke et al. (2016), the influence of this postprocessing on the numerical simulation of the rising bubble benchmark is investigated in detail. The resulting equations are solved by Newton’s method using a preconditioned GMRES iteration; see Garcke et al. (2016).

In Figure 9.2, we show the grid at final time, and in Figure 9.3, we show snapshots of the evolution of the interface between water and air, given by the zero-level line of φ , and the velocity field presented by streamlines of v , colored by $|v|$. We observe that, despite the unphysical parameters and boundary data, the method is able to deal with the complex two-phase interaction at the air-water interface.

9.5 Outlook on the Direction of Research

The numerical results proposed in Section 9.4 are only preliminary and should be regarded as a proof of concept for the proposed diffuse interface approach. The method is able to cope with the complex phenomena at the air–water interface. Further research is necessary to further develop our approach and to make it applicable for real-world scenarios. This includes:

Boundary data

As a first step, the application of periodic boundary data for v parallel to the water surface will be incorporated together with an open boundary on the top and the bottom of the domain. For φ and μ , periodic boundary conditions are sufficient in the water parallel directions only.

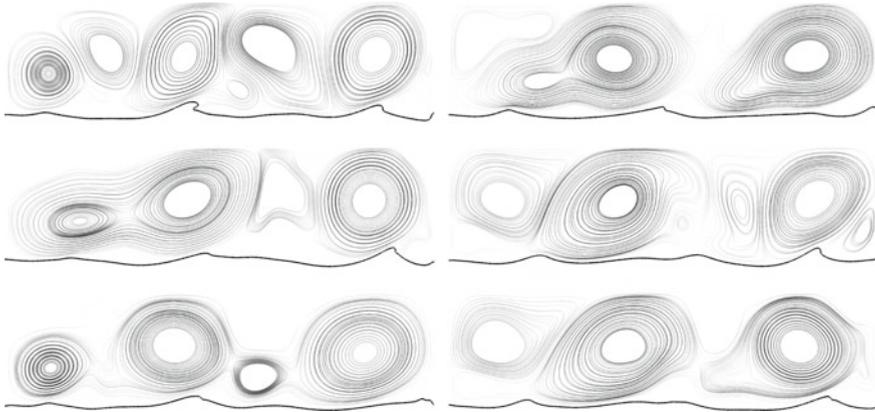


Fig. 9.3 Snapshots of the evolution of φ and v . For $t \in \{1.7, 3.3, 5.0, 6.7, 8.3, 10.0\}$ (left top to right bottom), we present streamlines of v in grayscale together with the zero-level line of φ in black. Darker streamlines means higher velocity. Due to the unphysical boundary data and the given forcing, we observe large vortices that generate several waves at different locations. We stress that especially breaking waves are captured by our approach as it is able to capture topological changes

3D computations

For 3D computations, an efficient solution of the linear systems arising throughout the simulation is essential. Here, results on preconditioning of the Cahn–Hilliard system from Bosch et al. (2014) or a multigrid approach as proposed in Kay and Welford (2006) might be used. For the solution of the Navier–Stokes equation, well-developed preconditioners exist and we refer to (Benzi et al. 2005; Kay et al. 2002).

Real-world parameter

Incorporating real-world parameters will require several changes on the architecture of the solver. Especially, we note that this will lead to large Reynolds number which require stabilization techniques like grad-div stabilization, which have to be incorporated into the finite element code. We note that the drawback of grad-div stabilization, namely a stronger coupling of the unknowns, does not appear here, as in (9.1) all variables are coupled anyway due to the term $2Dv = \nabla v + (\nabla v)^T$.

Incorporation into Earth System Models

If our concept proves applicable for the numerical simulation of the air–water region in atmosphere and ocean on the meter scale, it has to be incorporated into simulations on the next coarser (kilometer) scale. In this context, homogenization concepts might be an option. We refer to, e.g., Eck (2004), where homogenization of phase field models has been done in a different context.

Acknowledgements The authors are grateful for many discussions with Jeff Carpenter from the Helmholtz Center in Geesthacht on practical issues related to the wind-wave coupling at the interface of atmosphere and ocean. The second author acknowledges support of the TRR 181 funded by the German Research Foundation (DFG). The third author acknowledges support by the DFG through the International Research Training Group IGDK 1754 ‘Optimization and Numerical Analysis for Partial Differential Equations with Nonsmooth Structure.’

References

- Abels, H., Depner, D., Garcke, H.: Existence of weak solutions for a diffuse interface model for two-phase flows of incompressible fluids with different densities. *J. Math. Fluid Mech.* **15**(3), 453–480 (2013a)
- Abels, H., Depner, D., Garcke, H.: On an incompressible Navier-Stokes/Cahn-Hilliard system with degenerate mobility. *Annales de l’Institut Henri Poincaré (C) Non Linear. Analysis* **30**(6), 1175–1190 (2013b)
- Abels, H., Garcke, H., Grün, G.: Thermodynamically consistent, frame indifferent diffuse interface models for incompressible two-phase flows with different densities. *Math. Model. Methods Appl. Sci.* **22**(3), 1150013(40) (March 2012)
- Adams, R.A., Fournier, J.H.F.: Sobolev spaces. In: *Pure and Applied Mathematics*, 2nd edn, vol. 140. Elsevier (2003)
- Aki, G.L., Dreyer, W., Giesselmann, J., Kraus, C.: A quasi-incompressible diffuse interface model with phase transition. *Math. Model. Methods Appl. Sci.* **24**(5), 827–861 (2014)
- Aland, S., Voigt, A.: Benchmark computations of diffuse interface models for two-dimensional bubble dynamics. *Int. J. Numer. Methods Fluids* **69**, 747–761 (2012)
- Aland, S., Lowengrub, J., Voigt, A.: Two-phase flow in complex geometries: a diffuse domain approach. *CMES* **57**, **57**(1), 77–108 (2010)
- Anderson, D.M., McFadden, G.B., Wheeler, A.A.: Diffuse-interface methods in fluid mechanics. *Annu. Rev. Fluid Mech.* **30**, 139–165 (1998)
- Bañas, L., Nürnberg, R.: A posteriori estimates for the Cahn-Hilliard equation. *Math. Model. Numer. Anal.* **3**(5), 1003–1026 (2009)
- Barrett, J.W., Garcke, H., Nürnberg, R.: A stable parametric finite element discretization of two-phase Navier-Stokes flow. *J. Sci. Comput.* **63**(1), 78–117 (2014)
- Benzi, M., Golub, G.H., Liesen, J.: Numerical solution of saddle point problems. *Acta Numerica* **14**, 1–137 (2005)
- Blowey, J.F., Elliott, C.M.: The Cahn-Hilliard gradient theory for phase separation with non-smooth free energy. Part I: mathematical analysis. *Eur. J. Appl. Math.* **2**, 233–280 (1991)
- Bosch, J., Stoll, M., Benner, P.: Fast solution of Cahn-Hilliard variational inequalities using implicit time discretization and finite elements. *J. Comput. Phys.* **262**, 38–57 (2014)
- Boyer, F.: Mathematical study of multiphase flow under shear through order parameter formulation. *Asymptot. Anal.* **20**(2), 175–212 (1999)
- Boyer, F.: A theoretical and numerical model for the study of incompressible mixture flows. *Comput. Fluids* **31**(1), 41–68 (2002)
- Boyer, F., Chupin, L., Fabrie, P.: Numerical study of viscoelastic mixtures through a Cahn-Hilliard flow model. *Eur. J. Mech.-B/Fluids* **23**(5), 759–780 (2004)
- Boyer, F., Lapuerta, C., Minjeaud, S., Piar, B., Quintard, M.: Cahn-Hilliard/Navier-Stokes model for the simulation of three-phase flows. *Transp. Porous Media* **82**(3), 463–483 (2010)
- Chen, L.: iFEM: an innovative finite element method package in Matlab. <https://ifem.wordpress.com> (2008)
- Clément, P.: Approximation by finite element functions using local regularization. *RAIRO Analyse numérique* **9**(2), 77–84 (1975)

- Constantin, P., Foias, C.: Navier-Stokes-Equations. University of Chicago Press (1988)
- Ding, H., Spelt, P.D.M., Shu, C.: Diffuse interface model for incompressible two-phase flows with large density ratios. *J. Comput. Phys.* **226**(2), 2078–2095 (2007)
- Drennan, W., Donelan, M., Terray, E., Katsaros, K.: Oceanic turbulence dissipation measurements in SWADE. *J. Phys. Oceanogr.* **26**, 808–815 (1996)
- Druzhinin, O.A., Elghobashi, S.: Direct numerical simulations of bubble-laden turbulent flows using the two-fluid formulation. *Phys. Fluids* **10**, 685–697 (1998)
- Eck, C.: Homogenization of phase field models for binary mixtures. *SIAM Multiscale Model. Simul.* **3**(1), 1–27 (2004)
- Feng, X.: Fully discrete finite element approximations of the Navier-Stokes-Cahn-Hilliard diffuse interface model for two-phase fluid flows. *SIAM J. Numer. Anal.* **44**(3), 1049–1072 (2006)
- Ganesan, S., Tobiska, L.: A coupled arbitrary lagrangian-eulerian and lagrangian method for computation of free surface flows with insoluble surfactants. *J. Comput. Phys.* **228**, 2859–2873 (2009)
- Garcke, H., Hinze, M., Kahle, C.: A stable and linear time discretization for a thermodynamically consistent model for two-phase incompressible flow. *Appl. Numer. Math.* **99**, 151–171 (2016)
- Gross, S., Reusken, A.: Numerical methods for two-phase incompressible flows. In: Springer Series in Computational Mathematics, vol. 40. Springer (2011)
- Grün, G.: On convergent schemes for diffuse interface models for two-phase flow of incompressible fluids with general mass densities. *SIAM J. Numer. Anal.* **51**(6), 3036–3061 (2013)
- Grün, G., Guillén-González, F., Metzger, S.: On fully decoupled convergent schemes for diffuse interface models for two-phase flow with general mass densities. *Commun. Comput. Phys.* **19**(5), 1473–1502 (2016)
- Grün, G., Klingbeil, F.: Two-phase flow with mass density contrast: stable schemes for a thermodynamic consistent and frame indifferent diffuse interface model. *J. Comput. Phys.* **257**(A), 708–725 (2014)
- Guo, Z., Lin, P., Lowengrub, J.S.: A numerical method for the quasi-incompressible Cahn-Hilliard-Navier-Stokes equations for variable density flows with a discrete energy law. *J. Comput. Phys.* **276**, 486–507 (2014)
- Hintermüller, M., Hinze, M., Kahle, C.: An adaptive finite element Moreau-Yosida-based solver for a coupled Cahn-Hilliard/Navier-Stokes system. *J. Comput. Phys.* **235**, 810–827 (2013)
- Hintermüller, M., Hinze, M., Tber, M.H.: An adaptive finite element Moreau-Yosida-based solver for a non-smooth Cahn-Hilliard problem. *Optim. Methods Softw.* **25**(4–5), 777–811 (2011)
- Hysing, S., Turek, S., Kuzmin, D., Parolini, N., Burman, E., Ganesan, S., Tobiska, L.: Quantitative benchmark computations of two-dimensional bubble dynamics. *Int. J. Numer. Methods Fluids* **60**(11), 1259–1288 (2009)
- James, A.J., Lowengrub, J.: A surfactant-conserving volume-of-fluid method for interfacial flows with insoluble surfactant. *J. Comput. Phys.* **201**(2), 685–722 (2004)
- Kahle, C.: Simulation and control of two-phase flow using diffuse-interface models. PhD thesis, University of Hamburg (2014)
- Kay, D., Lohin, D., Wathen, A.: A preconditioner for the steady state Navier-Stokes equations. *SIAM J. Sci. Comput.* **24**(1), 237–256 (2002)
- Kay, D., Styles, V., Welford, R.: Finite element approximation of a Cahn-Hilliard-Navier-Stokes system. *Interfaces Free. Boundar.* **10**(1), 15–43 (2008)
- Kay, D., Welford, R.: A multigrid finite element solver for the Cahn-Hilliard equation. *J. Comput. Phys.* **212**, 288–304 (2006)
- Kay, D., Welford, R.: Efficient numerical solution of Cahn-Hilliard-Navier-Stokes fluids in 2d. *SIAM J. Sci. Comput.* **29**(6), 2241–2257 (2007)
- Kihara, N., Hanazaki, H., Mizuya, T., Ueda, H.: Relationship between airflow at the critical height and momentum transfer to the traveling waves. *Phys. Fluids* **19**(015102) (2007)
- Kim, J.: Phase-field models for multi-component fluid flows. *Commun. Comput. Physics* **12**, 613–661 (2012)
- Lowengrub, J., Truskinovsky, L.: Quasi-incompressible Cahn-Hilliard fluids and topological transitions. *Proc. R. Soc. A* **454**(1978), 2617–2654 (1998)

- Lubin, P., Glockner, S.: Numerical simulations of three-dimensional plunging breaking waves: generation and evolution of aerated vortex filaments. *J. Fluid Mech.* **767**, 364–393 (2015)
- McWilliams, J., Sullivan, P., Moeng, C.: Langmuir turbulence in the ocean. *J. Fluid Mech.* **334**, 1–30 (1997)
- Mellado, J.P., Stevens, B., Schmidt, H., Peters, N.: Two-fluid formulation of the cloud-top mixing layer for direct numerical simulation. *Theor. Comput. Fluid Dyn.* **24**, 511–536 (2010)
- Polton, J., Smith, J., MacKinnon, J., Tejada-Martinez, A.: Rapid generation of high frequency internal waves beneath a wind wave forced oceanic surface mixed layer. *Geophys. Res. Lett.* (L13602) (2008)
- Shen, L., Zhang, X., Yue, D., Triantafyllou, M.: Turbulent flow over a flexible wall undergoing a streamwise travelling wave motion. *J. Fluid Mech.* **484**, 197–221 (2003)
- Sullivan, P., McWilliams, J.: Turbulent flow over water waves in the presence of stratification. *Phys. Fluids* **14**, 1182–95 (2002)
- Sullivan, P., McWilliams, J.: Dynamics of winds and currents coupled to surface waves. *Ann. Rev. Fluid Mech.* **42**, 19–42 (2010)
- Sullivan, P., McWilliams, J., Melville, W.: Surface gravity wave effects in the oceanic boundary layer: large-eddy simulation with vortex force and stochastic breakers. *J. Fluid Mech.* **593**, 405–452 (2007)
- Sullivan, P., McWilliams, J., Moeng, C.: Simulation of turbulent flow over idealized water waves. *J. Fluid Mech.* **404**, 47–85 (2000)
- Sutherland, P., Melville, W.: Field measurements of surface and near-surface turbulence in the presence of breaking waves. *J. Phys. Oceanogr.* **45**, 943–965 (2015)
- Tsai, W., Chen, S., Lu, G.: Numerical evidence of turbulence generated by nonbreaking surface waves. *J. Phys. Oceanogr.* **45**, 174–180 (2015)
- Tsai, W., Hung, L.: Three-dimensional modeling of small-scale processes in the upper boundary layer bounded by a dynamic ocean surface. *J. Geophys. Res.* **112**, (C02019) (2007)
- Wan, D., Turek, S.: An efficient multigrid-FEM method for the simulation of solid-liquid two phase flows. *J. Comp. Appl. Math.* **203**(2), 561–580 (2007)
- Verfürth, R.: A posteriori error analysis of space-time finite element discretizations of the time-dependent Stokes equations. *Calcolo* **47**, 149–167 (2010)
- Wunsch, C., Ferrari, R.: Vertical mixing, energy, and the general circulation of the oceans. *Ann. Rev. Fluid Mech.* **36**, 281–314 (2004)

Index

A

Absolute vorticity, 18
Adaptive meshing, 298, 300, 303
Additive noise, 37
ADER flux evaluation, 264
Ageostrophic vorticity, 9
ALE vertical coordinate, 257
Altimetry, 133
Anisotropic tidal forcing, 111
Approximation
 beta-plane, 5
 Boussinesq, 4
 f-plane, 5
 hydrostatic, 12
 traditional, 8, 12
Arbitrary Lagrangian Eulerian, 257
Argo-CTD profiles, 114
Aspect ratio, 11
Available potential energy, 146, 159, 161, 167
Axial flow, 61, 79

B

Backscatter, 146, 148, 169, 170, 173, 174
Baroclinic, 131
 instability, 197, 205, 216
 Rossby radius, 15, 146, 151, 168
 tide, 101
Baroclinic turbulence, 162
Barotropic
 tide, 101
Barotropic vorticity equation, 152
Bernoulli function, 19
Bottom friction, 105
Boussinesq
 approximation, 4

 equations, 4, 62, 150

Brunt–Väisälä frequency, 8, 88
Buoyancy flux, 162
Burger number, 17, 19

C

Cahn–Hilliard, 291
Cascade
 baroclinic energy, 159
 barotropic energy, 159
 buoyancy variance, 167
 energy, 146, 155
 enstrophy, 155
 rotational component, 168
Casimir, 29
Conservation laws, 7, 153
Critical layer, 109
 absorption, 102
Cross-frontal shallow water equations, 66

D

Desaubies spectrum, 99
Diapycnal mixing, 245
Diffuse interface, 288
Direct averaging, 43
Dispersion relation, 60
Dissipation, 100, 128, 132

E

Eddy
 diffusivity
 diagnostics, 198
 estimates, 206
 parameterization, 214

- kinetic energy, 215
 - viscosity, 33
 - Eddy kinetic energy flux, 163
 - Eddy potential energy flux, 163
 - Eddy potential vorticity flux, 176, 178
 - Energy
 - available potential, 146, 159, 161, 167
 - backscatter, 146, 148
 - inequality, 293, 296, 297
 - kinetic, 146, 159, 161, 167
 - subgrid, 171
 - total, 7, 18, 26, 56, 153
 - Energy cascade, 146, 155
 - Energy flux, 156, 157, 162, 164
 - Energy range, 156
 - Energy spectrum, 154
 - Energy wavenumber, 154
 - Enstrophy, 153
 - Enstrophy cascade, 155
 - Enstrophy flux, 157
 - Enstrophy range, 156
 - Entropy budget equation, 227
 - Equation
 - barotropic vorticity, 152
 - Boussinesq, 4, 62, 150
 - Euler- α , 184
 - Euler–Poincaré, 25
 - Fokker–Planck, 40
 - free surface, 6
 - Kolmogorov backward, 40, 41
 - Lighthill, 57
 - primitive, 12
 - quasi-geostrophic, 12
 - quasi-geostrophic (barotropic), 16, 152
 - quasi-geostrophic (stratified), 14, 15, 166
 - quasi-geostrophic (two-layer), 16, 159
 - ray, 59
 - rotating shallow water, 17, 55
 - semi-geostrophic, 20, 21
 - shallow water (cross-front), 66
 - surface quasigeostrophic, 167, 168
 - Equatorial
 - Burger number, 23
 - Rossby number, 23
 - waves, 22
 - Euler- α equations, 184
 - Euler–Poincaré equation, 25
 - External Rossby radius, 19, 152, 178
- F**
- Finestructure method, 114
 - Finite volume particle method, 276
- Flow map, 24, 75
 - Fluctuation, 225
 - theorem, 238, 240
 - Flux
 - buoyancy, 162
 - eddy kinetic energy, 163
 - eddy potential energy, 163
 - eddy potential vorticity, 176, 178
 - energy, 156, 157, 162, 164
 - enstrophy, 157
 - momentum, 235
 - precipitation, 230
 - subgrid energy, 173
 - thermal, 231
 - thickness, 165
 - Fokker–Planck equation, 40
 - Forced-dissipative system, 225
 - f -plane approximation, 5
 - Free surface equation, 6
 - Frontal geostrophic adjustment, 65
 - Froude number, 2, 11
- G**
- Generalized Riemann problem, 265
 - Gent–McWilliams parameterization, 165, 178
 - GM spectrum, 92
 - Gravity
 - waves, 57, 59, 68
 - wave schemes, 117
 - Group velocity, 90
 - effective, 60
 - intrinsic, 59
- H**
- Hamiltonian, 7, 18, 26, 56
 - Hamilton principle, 23
 - HYCOM, 131
 - Hydrostatic
 - approximation, 12
 - balance, 13
 - pressure, 7
- I**
- IDEMIX
 - concept, 87
 - model, 94
 - Imbalance variables, 7
 - Internal
 - gravity waves, 88, 128
 - Rossby radius, 16

tides, 128
 Isonutral mixing, 261
 Itô process, 40

J

Jacobi identity, 28

K

Kernel-based approximation, 275
 Kinetic energy, 146, 159, 161, 167
 KLB theory, 156
 Kolmogorov backward equation, 40, 41
 Kuroshio front, 59

L

Lagrangian, 24, 184
 dispersion statistics, 207
 Lagrangian averaging, 184
 Lee wave generation, 105
 Leith parameterization, 146
 Lighthill
 equation, 57
 radiation, 106
 Linear stability analysis, 201
 Loss of balance, 104
 Low mode waves, 128

M

Mana–Zanna parameterization, 180
 Mixed layer turbulence, 103
 Mixing, 128, 131, 133, 136
 Modulation interaction, 113
 Mori–Zwanzig formalism, 37, 43
 M_2 tide, 103
 Multiplicative noise, 37, 44

N

Nambu
 bracket, 30
 formulation, 30
 Navier–Stokes, 291
 Near-inertial waves, 101, 128
 Non-acceleration theorem, 116
 Number
 Burger, 17, 19
 Burger (equatorial), 23
 Froude, 2
 Richardson, 62, 65
 Rossby, 2, 62, 70

Rossby (cross-frontal), 67
 Rossby (equatorial), 23
 Rossby (local), 71

O

Ocean general circulation models, 79
 Optimal balance, 74
 Ornstein–Uhlenbeck process, 39

P

Parameterization, 225
 backscatter, 170, 174
 Gent–McWilliams, 165, 178
 Leith, 146
 Mana–Zanna, 180
 Smagorinsky, 146, 187
 stochastic super-, 175
 Phase
 field variable, 288, 291
 velocity, 57
 Poisson
 bracket, 28
 formulation, 28
 Polyharmonic splines, 279
 Potential
 enstrophy, 56
 vorticity, 7, 18, 27, 56, 75
 Preconditioning, 303, 304
 Pressure
 hydrostatic, 7
 Primitive equations, 12
 Process
 Itô, 40
 Ornstein–Uhlenbeck, 39
 stochastic, 37
 Stratonovich, 40
 Wiener, 39, 177

Q

Quasi-geostrophic
 equation, 12
 barotropic, 16
 stratified, 14, 15
 two-layer, 16
 Quasi-geostrophic equation
 barotropic, 152
 stratified, 166
 two-layer, 159
 Quasi-geostrophic turbulence, 146, 151

R

- Radiative transfer equation, 91
- Ray equation, 59
- Refraction, 90
- Relative vorticity, 18, 56
- Resonant triad interactions, 113
- Richardson number, 62, 65
- Rosby
 - number, 2, 11, 62, 70
 - cross-frontal, 67
 - local, 71
 - radius
 - baroclinic, 15
 - external, 19
 - internal, 16
 - waves, 21, 35
- Rosby radius
 - baroclinic, 146, 151, 168
 - external, 152, 178
- Rotating shallow water equations, 17, 55
- Rotational eddy flux, 196

S

- Saturation concept, 116
- Semi-geostrophic equation, 20, 21
- Single-column approximation, 117, 118
- Slab model, 129
- Slow manifold, 74
- Smagorinsky model, 146, 187
- Spectrum
 - barotropic energy, 164
 - energy, 154
 - internal waves, 168
 - surface kinetic energy, 167
- Spontaneous creation, 113
- Spurious diapycnal mixing, 245
 - diagnosis, 250
- Stable scheme, 292, 294, 296
- Stochastic
 - differential equation, 37
 - mode reduction, 37–39, 43
 - process, 37
- Stochastic superparameterizations, 175
- STORMTIDE, 131
- Stratification, 8
- Stratonovich process, 40
- Subgrid energy, 171, 173

Subgrid energy flux, 173

Subgrid-scale processes, 225

Surface

internal wave interactions, 104

Surface quasigeostrophic equation, 167, 168

T

- Thickness flux, 165
- Time filter, 68
- TKE balance, 95
- Total energy, 7, 18, 26, 56, 153
- Traditional approximation, 8, 12
- Transport equation, 291
- Triad, 154
- Turbulence
 - baroclinic, 162
 - homogeneous, 185
 - quasi-geostrophic, 146, 151
 - three-dimensional, 173

V

- Variational principle, 24, 25
- Vertical diffusivity, 96
- Viscosity, 32, 33
- Vorticity
 - absolute, 18
 - ageostrophic, 9
 - potential, 7, 18, 27, 56, 75
 - relative, 18, 56

W

Wave

- action, 91
- breaking, 101, 110
- capture, 59, 62, 102, 109
- drag, 96
- equatorial, 22
- gravity, 57, 59, 68
- mean flow interaction, 97, 107
- nonlinear, 35
- Rosby, 21, 35
- WENO reconstruction, 264
- Wiener process, 39, 177
- Wind power, 129
- WKB methods, 90