Spontaneous Imbalance in Rotating Stratified Turbulence

Hossein Amini Kafiabad

Doctor of Philosophy

Department of Atmospheric and Oceanic Sciences

McGill University

Montreal,Quebec

August 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATION

To all who enjoy turbulence in their balanced life

ABSTRACT

Whereas high-frequency waves are valid solutions to the Boussinesq equations in certain limits, their amplitudes are generally observed to be small in large-scale atmospheric and oceanic data. Traditionally, this has led to the development of balance models, reducing the dynamics to only the slow subset. Their solutions, however, can spontaneously generate imbalance in the context of the full equations. To study this phenomenon, we perform high-resolution turbulence simulations of the non-hydrostatic Boussinesq equations, starting with a high-order nonlinearly balanced initial condition. To produce this initial condition, we employ the nonlinear normal mode initialization scheme that was introduced by Baer & Tribbia (1977) up to second order in the Rossby number as the next-order improvement to quasi-geostrophic theory. This initialisation is devised to limit the growth of high-frequency modes initially. To keep the flow predominantly balanced, we set the initial Rossby, Ro, and Froude, Fr, numbers to small values. By increasing these dimensionless parameters slightly, we investigate at what scales balance breaks down and after how long.

Our results show that the balanced initial condition displays a steep geostrophic energy spectrum and an even steeper ageostrophic energy spectrum. However, a shallow tail later develops in the ageostrophic energy spectrum at small scales if Ro and the Reynolds number, Re, are large enough. This shallow range moves toward smaller wavenumbers if Ro is increased. Increasing N/f (while f is held nearly constant) leads to a shift of the shallow tail to smaller scales due to lower amplitudes of

the unbalanced modes. By implementing a Baer-Tribbia-based balance/imbalance decomposition at the end of the integration time, we show that the unbalanced component of the dynamics accounts for the shallow range in the energy spectrum. This breakdown of balance, which leads to a change of slope in the energy spectrum, is consistent with observed atmospheric and oceanic spectra. The disparity between the fast and slow timescales, which is a requirement of balance, is shown to break down at scales where the energy spectrum is still steep and the scale-dependent Ro and Fr are small.

Spontaneous imbalance, which is the energy transfer from the balanced to the unbalanced modes, is also quantified in this study. More particularly, the scale at which it occurs and its dependence on the flow parameters are investigated. In our simulations, this quantity is shown to scale approximately as $Ro^{4.4}$ for our decomposition, accurate to second order in Ro. However, the wavenumber at which the maximum energy transfer from balance to imbalance occurs is invariant to Ro and Fr but dependent on the large-scale flow characterised by the Rossby radius of deformation. This wavenumber is found to be smaller than the wavenumber characterising the transition between the large scales and the shallow tail but larger than the peak of the energy spectrum. By comparing the results at different orders of balance, it is determined that all qualitative conclusions of this study are robust to the precise definition of balance.

RÉSUMÉ

Bien que les ondes à haute fréquence soient des solutions valides aux équations de Boussinesq dans certaines limites, leurs amplitudes sont généralement petites dans les données atmosphériques et océaniques à grande échelle. Traditionnellement, cela a mené au développement de modèles équilibrés, réduisant la dynamique à sa partie lente. Leurs solutions, cependant, peuvent générer spontanément un déséquilibre dans le contexte des équations complètes, ce qui consistue le sujet central de cette étude. Nous effectuons des simulations à haute résolution des équations non hydrostatiques de Boussinesq en débutant avec une condition initiale non linéairement équilibrée à ordre élevé. Pour produire cette condition initiale, nous utilisons le schéma d'initialisation non linéaire de Baer-Tribbia jusqu'au deuxième ordre du nombre de Rossby en tant qu'incrément à la théorie QG. Cette initialisation est conçue pour limiter la croissance initiale des modes haute fréquence. Pour maintenir l'équilibre de l'écoulement, nous assignons de petites valeurs initiales aux nombres de Rossby et de Froude. En augmentant légèrement ces paramètres sans dimension, nous étudions à quelle échelle se produit le bris d'équilibre et après combien de temps. Nos résultats montrent que la condition initiale équilibrée présente un spectre d'énergie géostrophique abrupte et un spectre d'énergie agéostrophique encore plus abrupte. Cependant, une queue plus douce se développe dans le spectre d'énergie agéostrophique aux petites échelles après un certain temps si Ro et Re sont assez grands. Cette gamme d'échelles moins escarpée se déplace vers les petits nombres d'onde si Roest augmenté. Augmenter N/f (alors que f est maintenu presque constant) mène à

un décalage de la queue vers les petites échelles. En utilisant la décomposition équilibre-déséquilibre de Baer-Tribbia à la fin du temps d'intégration, nous montrons que la composante déséquilibrée de la dynamique domine la portion douce du spectre énergétique. Ce bris d'équilibre, qui entraîne un changement de pente du spectre énergétique, est en accord avec les spectres atmosphériques et océaniques observés. L'écart entre les échelles de temps rapide et lente, une condition d'équilibre, disparaît à des échelles où le spectre d'énergie est encore abrupte et Ro et Fr sont petits.

La génération spontanée du déséquilibre, qui représente le transfert d'énergie des modes équilibrés vers les modes déséquilibrés, est également quantifiée dans cette étude. Plus particulièrement, l'échelle à laquelle elle se produit ainsi que sa dépendance aux paramètres de l'écoulement sont étudiés. On montre que cette quantité augmente selon $Ro^{4.4}$ pour la décomposition équilibre-déséquilibre exacte au deuxième ordre en Ro. Cependant, le nombre d'onde auquel survient le transfert d'énergie maximum entre modes équilibrés et déséquilibrés est indépendant de Ro et Fr, mais dépendant de l'écoulement à grande échelle caractérisé par le rayon de déformation de Rossby. Ce nombre d'onde est plus petit que celui de la transition spectrale, mais plus grand que celui du pic du spectre énergétique. En comparant les résultats à différents ordres d'équilibre, on confirme que toutes les conclusions de cette étude sont robustes à la définition précise d'équilibre.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my advisor Professor Peter Bartello. He was indeed the guiding light during different stages of my PhD. I genuinely want to extend this gratitude beyond a formal acknowledgement, as I greatly enjoyed 5 years of my life with him and learned from him in many different levels. This project would have not reached the present status without his continuous support and guidance. In addition to finding him a remarkable source of knowledge in turbulence and geophysical fluid dynamics, I learned from his academic manners and personality traits. Among them, patience and mutual respect were the most teaching ones to me.

Along with my adviser, I appreciate Professor David Straub for our fruitful discussions throughout my PhD. Apart from being my teacher, I significantly benefited from his insightful comments and feedback on my research. I also appreciate the financial support from Natural Sciences and Engineering Research Council of Canada (NSERC). The numerical calculations of this thesis were performed using the computational resources provided by Calcul Quebec (a Compute Canada Regional Partner).

My fellow students Oliver Asselin, Sisi Chen, Ali Asaadi and Yulia Akutina are thanked for the scientific interactions and worthwhile communications we have had during our graduate studies. Olivier Asselin kindly translated my abstract into French. Last but not least, the emotional support from my gracious family should

be highlighted. I am perpetually indebted to my mother, Maliheh, my father, Ahmad, and my brother Mohsen, for their patience, encouragement and love.

PREFACE AND CONTRIBUTION OF AUTHORS

This thesis aims to connect two realms of geophysical fluid dynamics: balance dynamics and unbalanced rotating stratified turbulence. The former refers to the large-scale atmospheric or oceanic flows that evolve with slow time scales. The slowness of these dynamics enables us to predict the weather for a few days or analyse the climate. On the other hand, unbalanced turbulence that is associated with nonlinear interactions and fast time scales makes these flows unpredictable, even on short time scales. Similar to human lives, the reality of the atmosphere and ocean is the concurrence of balance and imbalance. The dominance of one over the other depends on the spatial and temporal scales as well as other characteristics of a geophysical flow. This research is an effort to investigate the transition between these two contrasting dynamics.

As will be explained in detail, my research is based on an initial condition that is balanced to high order. This is used as an input for high-resolution turbulence simulations, where its breakdown to turbulence is studied. These simulations are performed using a pseudo-spectral code that solves the non-hydrostatic Boussinesq equations for periodic boundary conditions. The core of this solver has been developed by Peter Bartello (my supervisor) and modified by Micheal Waite to run on cluster grids using MPI. I develop a set of codes that produces the high-order-balanced initial conditions. They can be run on multiple processors using MPI.

In the study of balance breakdown, I also investigate the spontaneous generation of imbalance in time. In so doing, I propose a perturbation method that derives the

transfer between balance and imbalance at any time for any order of balance. I code this method and integrate it with the non-hydrostatic Boussinesq solver.

The chapters 2-4 of this thesis are the manuscripts of the following, closely-related articles

1. **Kafiabad, H. A.**, & Bartello, P. (2016). Balance dynamics in rotating stratified turbulence. *Journal of Fluid Mechanics*, **795**, 914-949.
2. **Kafiabad, H. A.**, & Bartello, P. (2017). Rotating stratified turbulence and the slow manifold. *Computers & Fluids*, **151**, 23-34.
3. **Kafiabad, H. A.**, & Bartello, P. Nonlinear transfers between balanced and unbalanced modes of rotating stratied flows. *Submitted to the Journal of Fluid Mechanics*.

The first article has been published in the *Journal of Fluid Mechanics*, and the second one in the special edition of the journal *Computers & Fluids* dedicated to C. E. Leith. The third paper is uder revision for the *Journal of Fluid Mechanics*. Each manuscript is co-authored by my supervisor, Peter Bartello. He played a normal supervisory role in the research and performed some editing of the manuscript text. Considering that these papers are authored only by myself and the supervisor of this thesis, no other major contributions need to be declared in this section. We most certainly benefited from discussions with our colleagues and collaborators who are acknowledged either in the published papers or the opening of this thesis, but they were not closely involved with this work.

Each of the above papers (chapters 2-4) entails a rather extensive review of the

related literature in their respective introductions. Therefore, I devoted chapter 1 of this thesis to explaining relevant concepts and phenomena. I hope this improves the readability of my thesis.

The contributions of this thesis to original knowledge are:

- A high-order nonlinear normal mode initialization (NNMI) scheme, namely that of Baer and Tribbia, is implemented numerically for turbulence simulations. In this implementation we propose a pseudo-spectral scheme to derive a set of balanced ageostrophic modes associated with any distribution of geostrophic modes. The geostrophic modes together with the balanced ageostrophic modes can serve as an initial condition that is initially devoid of fast time scales. Without our suggested numerical scheme, implementation of the Baer-Tribbia scheme for the resolutions that are commonly used in turbulence studies would be very difficult.
- High-order initial conditions enabled us to derive the slope of the balanced ageostrophic energy spectrum. This slope is shown to be steeper than the (already steep) geostrophic spectrum. It is then concluded that balance occurs in conjunction with steep spectra. Therefore, balance dynamics cannot explain the observed slope of -5/3 in the atmospheric mesoscale or the oceanic submesoscale range of the energy spectrum.
- Starting from the nonlinearly-balanced initial condition, the energy spectrum is shown to develop a shallow tail if the Rossby number (Ro) is large enough. By increasing Ro, this starts at lower wavenumbers and at earlier times. By decomposing the flow to its balanced and unbalanced parts at the end of the

simulations, it is shown that the unbalanced component accounts for the shallow range.

- The shallow spectrum in our simulation, which is shown to be the result of imbalance generation, is consistent with the observed atmospheric (see Gage & Nastrom, 1986) and oceanic (see Klymak & Moum, 2007*a*,*b*) data. Without downplaying the role of boundaries in the breakdown of balance, the internal dynamics that lead to the shallowing of the energy spectrum are demonstrated.
- A new method based on the smallness of the unbalanced motion is proposed to derive the transfer between the balanced and unbalanced parts of rotating stratified flows. Unlike linear decompositions, this method can be applied to any order of balance in which the balanced and unbalanced modes are not necessarily orthogonal. The numerical implementation of this method based on pseudo-spectral techniques was performed.
- Employing the newly proposed method, spontaneous imbalance is studied. It is shown that the balanced energy is transferred to imbalance at wavenumbers smaller than that of transition from the steep to the shallow energy spectrum. Over the range of Ro we explored, the wavenumber at which maximum transfer occurs is shown to be invariant to the strength of rotation. However, by increasing Ro the magnitude of transfer from balance to imbalance increases.
- The frequency spectra of the balanced and unbalanced modes at different scales are calculated. It is shown that at large scales both the geostrophic and balanced ageostrophic modes peak at low frequencies, whereas the unbalanced modes display a sharp peak at the linear frequency. This distinct separation

of time scales confirms that the decompositions based on frequency are maintained at these scales. These peaks in the balanced and unbalanced frequency spectra are shown to merge and form similar flat curves when the Fourier amplitudes of higher wavenumbers are considered. The scale at which they merge is shown to be larger than that of the kink of the energy spectrum, signalling that the time scale separation breaks down at scales larger than the appearance of the sub-deformation scale -5/3 range in the energy spectrum.

CHAPTER 1

Introduction

1.1 Motivation

Examples of fluid dynamics are so ubiquitously present in our daily lives that we have a hard time noticing them. The mixing of milk in our coffee, the ventilation of our offices, the circulation of blood in our veins and the weather forecast are among the many examples that we encounter every day. The applications of fluid dynamics can be extended to complicated technological advances such as combustion in car engines or the design of aircraft, as well as natural phenomena such as the formation of hurricanes. All of these problems share similar governing equations. However, the specificity of each problem makes scientists from different fields focus on particular features. In so doing, they employ different approximations, different frames of reference and different numerical or analytical approaches to solve them. Coming from an engineering background, I found myself fascinated by three rather unique features of geophysical fluid dynamics that I had not encountered in engineering problems.

The first particular feature is the effect of earth's rotation, which adds the Coriolis "force" to the system. Stratification should be counted as the second interesting

aspect of geophysical flows. This term indicates that the atmosphere and ocean consist of layers of different density. The third intriguing feature is the range of spatial scales that play significant roles in the physics of these flows. The atmospheric spectrum ranges from planetary scales (order of 10^4 kilometres) to microphysics scales (order of 10^{-6} metres). This extraordinary range makes the atmosphere a mega fluid dynamics laboratory that no human-made device can describe completely. Due to nonlinearities in the equations of motion, different scales of fluid flow interact with each other. This interaction turns these flows into turbulence. Section 1.4 elaborates upon the concept of turbulence, but in a few words it can be described as the chaotic motion of fluid particles in which different scales interact with each other. As turbulent flow contains a wider range of scales, it can display more complicated behaviour. Therefore, atmospheric and oceanic flows are some of the most interesting media in which to study turbulence. To support this statement, it is befitting to mention a historical study of turbulence in the ocean. One of the most celebrated theories of turbulence was proposed by Kolmogorov (1941), which suggests the power-law scaling of $k^{-5/3}$ (k being wavenumber) for the energy spectrum of 3D isotropic turbulence.[1] This theory was tested by Grant *et al.* (1962) for the first time with oceanic data measured in Discovery Passage, a tidal channel on the west

[1] Kolmogorov's theory predicts the slope of the inertial subrange when the Reynolds number is large enough. These terms, as well as other pertinent concepts, will be introduced below.

coast of Canada. The Reynolds number of this flow was reported to be 10^8, a value that cannot be reached in any lab even with today's modern facilities.

In short, these three particular characteristics of geophysical fluid dynamics can be put together in the term "rotating stratified turbulence", which is not only a part of the thesis title but also my motivation for my research in the subject matter.

Another concept that is essential to this study is balance dynamics, which is the reduction of dynamics to a balance between dominant forces (such as the pressure gradient force and the Coriolis force). As will be shown later, these dynamics have slow time scales since the high-frequency components are filtered out. This thesis aims to answer some fundamental questions about the mechanisms that break down balance dynamics in geophysical flows. It also sheds light on the energetics of the atmosphere and ocean through the investigation of this breakdown from a turbulence perspective. To understand these questions better, the introductory sections of this chapter familiarise the reader with the required concepts.

The rest of this chapter is organised as follows. The next section introduces the governing equations of motion. A reader who does not want to involve him/herself with too much technicality can skim through this section. This is followed by a section on balance dynamics. Afterwards, different theories of turbulence are discussed. The manifestion of these theories appears in the subsequent section on the atmospheric and oceanic energy spectra. This chaper is concluded by expressing the objectives of this thesis.

1.2 Governing equations

As mentioned in the previous section, different fluid flows are principally governed by similar equations. The motion of Newtonian fluids are described by the fully compressible Navier-Stokes equations (see e.g. Batchelor, 2000). These equations can be written as

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{\nabla P}{\rho} - g\hat{\boldsymbol{z}} + \boldsymbol{D}_{\mathbf{u}}, \tag{1.1a}$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{1.1b}$$

where $\mathbf{u} = (u, v, w)$ is the velocity, P is the pressure, $\hat{\boldsymbol{z}}$ is the vertical unit vector and ρ is the density. $\boldsymbol{D}_{\mathbf{u}}$ represents the dissipation term.[2] (1.1a) is derived from the conservation of momentum for fluid parcels, and (1.1b) from the conservation of mass. These equations need to be completed with the first law of thermodynamics as well as an appropriate equation of state since the unknown variables outnumber the equations. The atmosphere and ocean consist of air and water, which are simple Newtonian fluids. Therefore, they can be fully described by compressible Navier-Stokes equations. However, these equations are extremely complicated and difficult to solve numerically. As discussed in section 1.1, depending on the problem, different approximations are used to simplify them. In this study, we take the following steps and assumptions:

[2] In the case of incompressible Newtonian fluids, $\boldsymbol{D}_{\mathbf{u}}$ can be defined as $\nu \nabla^2 \mathbf{u}$, where ν is the molecular viscosity.

1. We observe (and measure) all the kinematic variables such as velocity from the earth's point of view, which is a rotating frame of reference. Therefore, we need to rewrite the governing equations for this rotating coordinate. Assuming earth's angular velocity to be $\mathbf{\Omega}$, this conversion of coordinate introduces two new terms to the equations: the centrifugal force $\mathbf{\Omega} \times (\mathbf{\Omega} \times \boldsymbol{r})$ and the Coriolis force $2\mathbf{\Omega} \times \mathbf{u}$. The centrifugal force, $\mathbf{\Omega} \times (\mathbf{\Omega} \times \boldsymbol{r})$, can be written as a gradient of a potential function. Hence, it can be included in the pressure term.
2. We simplify the Coriolis term to $f\hat{\boldsymbol{z}} \times \mathbf{u}$, in which $f = 2|\mathbf{\Omega}| \sin \phi_o$ is the Coriolis parameter at the latitude ϕ_o and is assumed to be constant. This assumption is commonly referred to as an "f-plane". For high- and mid-latitudes and the spatial scales considered in this study, this assumption is justified.
3. Considering that the density fluctuations are relatively small in the atmosphere and ocean, we employ the Boussinesq approximation. More specifically, we assume that density is the sum of a base profile and fluctuations

$$\rho(\mathbf{r}, t) = \rho_o(z) + \rho'(\mathbf{r}, t) \tag{1.2}$$

   in which $\mathbf{r}$ denotes the position vector, t is time, and the base profile is a linear function, $\rho_o(z) = \rho_s - \alpha z$. The Boussinesq approximation can then be expressed as $|\rho - \rho_s|/\rho_s \ll 1$.
4. Likewise, for the pressure we assume

$$P(\mathbf{r}, t) = P_o(z) + p(\mathbf{r}, t) \tag{1.3}$$

in which P_o is the hydrostatic (basic-state) pressure and p is the dynamic pressure. For the static pressure, we can derive $P_o(z) = P_s - \rho_s g z$.

5. We employ the ideal gas law and the first law of thermodynamics (in the adiabatic form) to relate ρ, P and θ (θ being the potential temperature). In so doing, θ takes a similar form to (1.2) and (1.3)

$$\theta(\mathbf{r}, t) = \theta_0(z) + \theta'(\mathbf{r}, t). \tag{1.4}$$

where, like the pressure and density, $\theta_0(z)$ is a linear function and $\theta_0(0) = \theta_s$.

After applying the above, (1.1) transforms into the following set of equations (see e.g. Vallis (2006) for details)

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + f\hat{\boldsymbol{z}} \times \mathbf{u} = -\nabla p + b\hat{\boldsymbol{z}} + D_{\mathbf{u}}, \tag{1.5a}$$

$$\nabla \cdot \mathbf{u} = 0, \tag{1.5b}$$

$$\frac{\partial b}{\partial t} + \mathbf{u} \cdot \nabla b = -N^2 w + D_b, \tag{1.5c}$$

They are written in terms of the buoyancy perturbation, b, instead of θ' or ρ'. For atmospheric flows this quantity is defined as $b = g\theta'/\theta_s$, in which g is the gravitational acceleration, and for oceanic flows it is $b = -g\rho'/\rho_s$. The Brunt-Väisälä frequency in (1.5c) is defined as

$$N = \sqrt{\frac{g}{\theta_s}\frac{d\theta_o(z)}{dz}} = \sqrt{-\frac{g}{\rho_s}\frac{d\rho_o(z)}{dz}}, \tag{1.6}$$

and is assumed to be constant.

Just like any set of partial differential equations (PDEs), (1.5) requires a set of boundary conditions as well as initial conditions to be well-posed mathematically. We discuss our choice of initial conditions in detail in chapter 2. The generation of this initial condition is one of the contributions of this thesis as it is balanced to high orders. As for the boundary conditions, we choose periodicity in all dimensions for the variables in (1.5).[3] This geometry maintains the statistical homogeneity of the flow, which is suitable for turbulence studies. While we expect that balance will break down first near boundaries, we perfer to start with this simpler problem. Periodicity also enables us to use spectral numerical schemes in all three dimensions to solve the equations more efficiently. If the flow is periodic, it can be described in terms of its discrete Fourier modes. In so doing, $\mathbf{u}$, b and p are written as

$$\mathbf{u}(\mathbf{r},t) = \sum_{\mathbf{k}} \mathbf{u}_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.7a}$$

$$b(\mathbf{r},t) = \sum_{\mathbf{k}} b_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.7b}$$

$$p(\mathbf{r},t) = \sum_{\mathbf{k}} p_{\mathbf{k}}(t) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1.7c}$$

in which $\mathbf{u}_{\mathbf{k}}(t) = (u_{\mathbf{k}}(t), v_{\mathbf{k}}(t), w_{\mathbf{k}}(t))$ is the velocity field transformed to Fourier space (likewise for $b_{\mathbf{k}}$ and $p_{\mathbf{k}}$), and the wavevector $\mathbf{k}$ has the components of (k_x, k_y, k_z). The

[3] Note that the buoyancy fluctuation, which is related to the density fluctuation, appears in (1.5) and is assumed to be periodic. The total density field, on the other hand, is not periodic.

wavevectors are discrete for periodic flows. Therefore, we write them as indices of flow variables. After taking the (discrete) Fourier transform of (1.5), the following equations are derived

$$\frac{\partial \mathbf{u_k}}{\partial t} - b_{\mathbf{k}}\hat{\boldsymbol{z}} + f\hat{\boldsymbol{z}} \times \ \mathbf{u_k} + i\mathbf{k} \cdot p_{\mathbf{k}} = \sum_{\mathbf{p+q=k}} (i\mathbf{p} \cdot \mathbf{u_p})\ \mathbf{u_q} + D_{\mathbf{u}}, \tag{1.8a}$$

$$i\mathbf{k} \cdot \mathbf{u_k} = 0, \tag{1.8b}$$

$$\frac{\partial b_{\mathbf{k}}}{\partial t} + N^2 w_{\mathbf{k}} = \sum_{\mathbf{p+q=k}} i\mathbf{p} \cdot \mathbf{u_p}\ b_{\mathbf{q}} + D_b. \tag{1.8c}$$

If the periodic domain has length L_x, L_y and L_z, the discrete wavenumbers can be written in terms of integer numbers n_x, n_y and n_z as below

$$k_x = \frac{2\pi n_x}{L_x}, \quad k_y = \frac{2\pi n_y}{L_y}, \quad k_z = \frac{2\pi n_z}{L_z}. \tag{1.9}$$

From the above equations, it is clear that the distance between two subsequent wavenumbers along the k_x, k_y and k_z axes are

$$\triangle k_x = \frac{2\pi}{L_x}, \quad \triangle k_y = \frac{2\pi}{L_y}, \quad \triangle k_z = \frac{2\pi}{L_z}. \tag{1.10}$$

Multiplying and dividing the RHS of (1.7a) by $L_xL_yL_z$ yields

$$\mathbf{u}(\mathbf{r}, t) = \sum_{\mathbf{k}=-\infty}^{+\infty} L_xL_yL_z\mathbf{u_k}(t)e^{i\mathbf{k}\cdot\mathbf{r}}\frac{1}{L_x}\frac{1}{L_y}\frac{1}{L_z}. \tag{1.11}$$

We can write $\hat{\boldsymbol{u}}(\mathbf{k}, t) = L_xL_yL_z\mathbf{u_k}(t)$, and substitute $1/L_x$ with $1/2\pi\triangle k_x$ from (1.10) (likewise for $\triangle k_y$ and $\triangle k_z$)

$$\mathbf{u}(\mathbf{r},t) = \frac{1}{(2\pi)^3} \sum_{\mathbf{k}=-\infty}^{+\infty} \hat{\boldsymbol{u}}(\mathbf{k},t) e^{i\mathbf{k}\cdot\mathbf{r}} \triangle k_x \triangle k_y \triangle k_z. \tag{1.12}$$

In the limit of $L_x, L_y, L_z \to \infty$, $\triangle k_x, \triangle k_y, \triangle k_z \to 0$ and $\hat{\boldsymbol{u}}(\mathbf{k},t)$ becomes a continuous function of wavevector. Therefore, the above equation takes an integral form

$$\mathbf{u}(\mathbf{r},t) = \frac{1}{(2\pi)^3} \int_{\mathbf{k}=-\infty}^{+\infty} \hat{\boldsymbol{u}}(\mathbf{k},t) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k}, \tag{1.13}$$

where $d\mathbf{k} = dk_x\ dk_y\ dk_z$.

A common practice in fluid dynamics is non-dimensionalising the governing equations with characteristic quantities. This procedure introduces dimensionless parameters, which are the charateristics of the problem. Moreover, it determines the order of the terms in the equations with respect to each other based on the values of these parameters. We non-dimensionalise the equations by using L for scaling of x and y, H for z, U for the horizontal velocity and L/U for time. In some numerical simulations, L and H are set equal to the size of the domain or the forcing scale (if the simulations are forced). All the simulations in this study are of decaying turbulence, and we define L and H to be the peaks of the horizontal and vertical energy spectra [4] , respectively. In other words, L is the length scale that contains more energy than any other horizontal scale (likewise for H and vertical scales). The continuity equation (1.5b) leads to the scaling of UH/L for the vertical velocity. Furthermore, we use geostrophic and hydrostatic balance to scale pressure and buoyancy terms

[4] The energy spectrum will be defined later in this chapter.

with fUL and fUL/H, respectively. The scaling of dissipation terms shows that they are proportional to the inverse of the Reynolds Number, which is defined as

$$Re = \frac{UL}{\nu}. \tag{1.14}$$

It can be shown that Re is directly related to the ratio of the largest to smallest scale in the flow. Therefore, it indicates the range of spatial scales that a fluid flow entails. As mentioned, atmospheric and oceanic flows display very large Reynolds numbers, which indicates that the dissipation terms are very small in these flows; therefore, we neglect them in the equations. After scaling, (1.5) takes the following dimensionless form

$$\frac{\partial \boldsymbol{u}_h}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u}_h + \frac{\hat{\boldsymbol{z}} \times \boldsymbol{u}_h}{Ro} = -\frac{\nabla_h\, p}{Ro}, \tag{1.15a}$$

$$\alpha^2 \left(\frac{\partial w}{\partial t} + \boldsymbol{u} \cdot \nabla w \right) = -\frac{1}{Ro}\frac{\partial p}{\partial z} + \frac{b}{Ro}, \tag{1.15b}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{1.15c}$$

$$\frac{\partial b}{\partial t} + \boldsymbol{u} \cdot \nabla b = -\frac{w}{Fr^2}. \tag{1.15d}$$

The other dimensionless parameters that appear in the equations are the Rossby number, the Froude number and the aspect ratio.

The Rossby Number is defined as

$$Ro = \frac{U}{fL}. \tag{1.16}$$

L/U, which was used as the characteristic time scale to non-dimensionalise the equations, is also known as the eddy turn-over time. This name implies that L/U measures the time it takes for the characteristic (or the most energetic) eddies of the flow to make a full turn. $1/f$, on the other hand, is proportional to the period of earth's rotation. Therefore, Ro shows how fast the flow is compared to earth's rotation. At large scales, the atmosphere and ocean have small Rossby numbers, but this number increases when the dynamics of smaller scales are considered. Weather, which is associated with synoptic scales, changes in the course of several days. Large-scale oceanic eddies are even slower. Therefore, Ro is less than unity for these flows. In (1.15a), the order of the Coriolis term, $\hat{\boldsymbol{z}} \times \boldsymbol{u}_h/Ro$, is inversely proportional to Ro. Therefore, Ro^{-1} indicates the strength of rotation in a flow.

The Froude Number is defined as

$$Fr = \frac{U}{NH}. \tag{1.17}$$

H/U is the time scale associated with the vertical shear of horizontal wind. Similar to Ro, Fr^{-1} indicates the strength of stratification in a flow. In stratified turbulence without rotation a horizontal Froude number is also defined, $Fr = U/(NL)$, as it has been found that the vertical Froude number goes to unity (e.g. Billant & Chomaz, 2001). As a consequence, numerical simulations require very high vertical resolution. In large-scale atmospheric and oceanic flows, as well as the simulations in this study, QG dynamics increases the vertical scale compared to that of stratified turbulence (Waite & Bartello, 2006a) and the vertical Froude number is the relevant quantity.

Among the dimensionless parameters, there is also the aspect ratio of the flow

$$\alpha = H/L, \tag{1.18}$$

which appears in (1.15). This ratio is usually very small in large-scale geophysical flows.

1.3 Balance dynamics

The large-scale atmospheric and oceanic flows evolve over long periods of time. On the other hand, general governing equations, such as the Boussinesq or primitive equations, can contain very fast motions. Therefore, there has been a lot of effort to simplify these such that only the slow dynamics are allowed. These simplifications have led to the development of a subfield in geophysical fluid dynamics known as balance dynamics, the goal of which is to reduce the dynamics to a slow subset that can accurately describe large-scale atmospheric and oceanic flows.

Despite its frequent use in meteorology, oceanography and climatology, the term balance dynamics is often not precisely defined or its definition varies in different contexts. Sometimes this term refers to the reduction of dynamics to two (or more) dominant forces in a flow that balance each other. Sometimes this reduction is done by perturbation techniques based on small Ro that reveal the order of each term in the equations. After deriving the order of different terms, leading orders can be retained and higher ones neglected. In general, balance dynamics are seen as the reduction of dimensionality of the flow. In this sense, the concept of balance is entwined with the concept of the slow manifold. This manifold is defined as a proper subset of the phase space on which the dynamics are slow. In this section we briefly review different topics related to balance dynamics.

1.3.1 Balance between dominant forces

Different terms in equation (1.15) represent forces acting on fluid parcels. Some of these may be stronger in some limits than in others. For example, in the regimes of high-Re, the dissipation term in (1.15) is much smaller than the other terms, since Re

appears in the denominator. If, in addition to large Re, a flow has a small Ro, only the last term in the LHS and the first term in the RHS of equation (1.15a) remain to leading order. These terms correspond to the Coriolis and horizontal pressure gradient forces, respectively. In dimensionless form, the balance between them is written as

$$\hat{\boldsymbol{z}} \times \boldsymbol{u}_h = \nabla_h \, p \tag{1.19}$$

and named geostrophic balance. These equations are diagnostic without any time derivatives, whereas the atmosphere and ocean change in time. Therefore, (1.19) is too simple to describe such flows. To derive higher orders of balance, the variables in (1.5) can be expanded in terms of Ro. Balance models of different orders can then be obtained by keeping terms up to an appropriate order, which depends on the level of desired accuracy. If the approximation is taken to an order higher than (1.19), qausi-geostrophy (QG) is derived[5] . To introduce its governing equations, we first define the stream function, ψ, as

$$u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x}, \tag{1.20}$$

and then the potential vorticity (PV) in terms of ψ as[6]

[5] see e.g. Salmon (1998) for the derivation of the QG equations.

[6] Potential vorticity may be defined in a more general form that includes QG as well as non-QG flows.

$$q = \nabla^2\psi + f + \frac{f^2}{N^2}\frac{\partial}{\partial z}\left(\frac{\partial \psi}{\partial z}\right). \tag{1.21}$$

The QG equation that describes the evolution of q is

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + u\frac{\partial q}{\partial x} + v\frac{\partial q}{\partial y} = 0. \tag{1.22}$$

which is accurate up to order Ro (Charney, 1949). After using (1.20), (1.22) can be solved for ψ as the only unknown variable. QG is of great importance in geophysical fluid dynamics for several reasons. First, it can reliably describe the large-scale atmospheric and oceanic flows more effectively than the Boussinesq approximation since it has no high-frequency modes. Second, it can be stated in terms of one scalar variable, stream function, as opposed to several vector fields. Third, as can be inferred from (1.22), potential vorticity (and any function of this quantity) is conserved on each inviscid trajectory of the flow. This introduces several important features to the dynamics. One of them is the similarity to 2D turbulence, which will be discussed in section 1.4.3. In this study, we use a balance model that is an order higher than QG, whose details are described in the following chapters. For more comprehensive descriptions of balance models and their hierarchies, one can refer to Warn *et al.* (1995); McIntyre & Norton (2000); Mohebalhojeh & Dritschel (2001).

1.3.2 The slow manifold and low-order dynamical systems

The concept of balance is closely related to the separation of timescales, which in turn depends on the smallness of the Rossby number. At the large scales of the atmosphere, Ro is small. For example, if we consider the characteristic values of $U \approx 10\, m.s^{-1}$, $L \approx 10^6 m$ and $f \approx 10^{-4}\, s^{-1}$ for synoptic-scale flows, we find $Ro \approx 0.1$.

Large-scale oceanic flows are characterised by even smaller Ro. We can show for the regimes of small Rossby number there is a distinct separation of timescales. For instance, if a flow is periodic, this separation can be seen by decomposing the flow into Normal Modes and comparing their corresponding characteristic frequencies. The normal mode decomposition is presented in appendix A. It is demonstrated there that the non-hydrostatic Boussinesq system consists of three sets of orthogonal modes with the following linear dispersion relation

$$\omega_s = 0, \quad \omega_f = \pm\left(\frac{f^2 {k_z}^2 + N^2 {k_h}^2}{{k_z}^2 + {k_h}^2}\right)^{1/2}, \tag{1.23}$$

in which ω_s and ω_f are the frequencies of the slow and fast modes of motion, repectively. These modes are termed geostrophic and ageostrophic in the appendix as well as the following chapters. However, in this section we present a more general description of the slow manifold that does not necessarily depend on normal mode decomposition. Therefore, we use the terms fast and slow variables instead of geostrophic and ageostrophic modes. The inverse of $|\omega_f|$ marks the timescale of fast modes, T_f. However, the timescale of slow modes cannot be infinity based on (1.23). Instead, the timescale of these modes, T_s, is calculated from the nonlinear part of their equation (which is the advective timescale equal to L/U). The ratio between these two timescales is then derived as

$$\begin{aligned} \frac{T_f}{T_s} &= \frac{U/L}{\left((f^2 {k_z}^2 + N^2 {k_h}^2)/({k_z}^2 + {k_h}^2)\right)^{1/2}} \\ &= \frac{U/fL}{\left(1 + (N^2/f^2 - 1){k_h}^2/({k_z}^2 + {k_h}^2)\right)^{1/2}} \leq \left(\frac{U}{fL} = Ro\right). \end{aligned} \tag{1.24}$$

The last inequality is deduced by assuming $N/f > 1$, which is a valid assumption considering that this ratio is around 100 in the atmosphere and 30 in the ocean. If Ro is very small, T_f has to be much smaller than T_s. This separation between the two timescales is central to the definition of the slow manifold and the particular balance model (Baer & Tribbia, 1977) that is used in this thesis.

If we denote the variables (modes) that can have fast time scales by f and the variables (modes) that have only slow time scales by s, we can write the following non-dimensionalised evolution equations to describe them

$$\frac{\partial s}{\partial t} = N_s(s, f), \tag{1.25a}$$

$$\frac{\partial f}{\partial t} + \frac{1}{\epsilon}\mathcal{L}\ f = N_f(s, f), \tag{1.25b}$$

in which N_f and N_s are the nonlinear terms and $\epsilon = Ro$. Note that s and f are loosely referred to as slow and fast modes, while the high-frequency part of f can be filtered out. As argued in Warn *et al.* (1995), the flow can be decomposed into s and f in different ways. For example, Mohebalhojeh & Dritschel (2001) and Vanneste (2013) considered potential vorticity as the slow variable and horizontal divergence and ageostrophic vorticity as fast variables. Leith (1980) and Bartello (1995) employed the normal (geostrophic and ageostrophic) modes to define their fast and slow modes (see Appendix A). This choice reduces the matrix $\mathcal{L}$ in (1.25b) to ω_f defined in (1.23). Bartello (1995) showed that in the limit of small Ro and Fr only geostrophic modes contribute to the potential vorticity. Therefore, in this limit geostrophic and vortical modes can be used interchangeably.

(1.25) includes the nonlinear terms in the RHS and the time derivative of all variables. Therefore, each mode can evolve in time without any constraints while it influences and is influenced by all other modes. Considering that slow dynamics are predominant in the regimes of small Rossby numbers, balance models aim to filter out the fast motion and describe the dynamics with only a slow-time evolution. These models generally embody two main components:

1. A balance relation, which filters out the fast component of the dynamics by slaving f to s. In other words, instead of calculating fast variables independently from (1.25b), they are derived from balance relations which are time-independent functions written as

$$f = F(s, \epsilon). \tag{1.26}$$

Balance relations reduce the dimensionality of the system by adding a new constraint. Some studies such as Van Kampen (1985) and Warn *et al.* (1995) termed (1.26) the "slave relation" to highlight the fact that in the full equations (1.25) f and s are in fact independent of each other, but in the reduced dynamics the evolution of f is determined by s.

2. Balance dynamics, which describe the time evolution of the slow variables, now independent of f. Hence, they can be described with

$$\frac{\partial s}{\partial t} = \tilde{N}_s(s). \tag{1.27}$$

As mentioned, some balance models consider potential vorticity as their slow variable. For these models, (1.27) is merely the QG equation introduced in (1.22).

Warn *et al.* (1995) argued that in some cases balance relations may remain accurate while balance dynamics do not. Therefore, both aspects of balance models should be considered together.

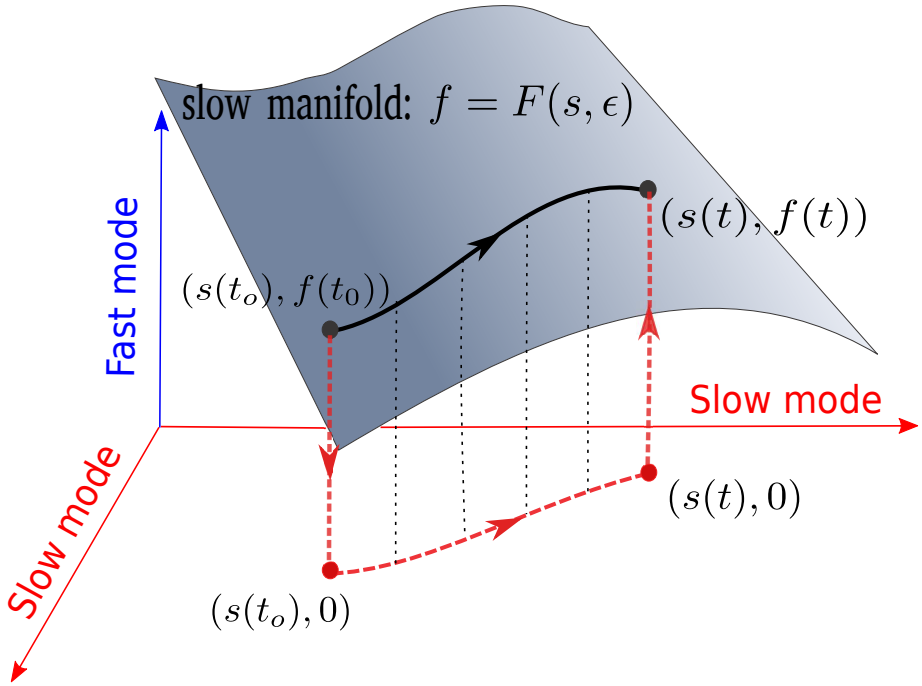


Figure 1–1: Schematic view of the slow manifold. The horizontal axes represent the slow modes and the vertical axis represents the fast modes. There are many fast and many slow modes in a geophysical flow. This allocation of axes is merely a simplified visualisation.

Leith (1980) and Lorenz (1980) pioneered the concept of the slow manifold, which can be viewed as a geometric representation of balance in phase space. A solution of the governing PDEs can be visualised as a trajectory in the infinite-dimensional phase space. Each indexed variable in (1.8) is seen as two dimensions of this space, one for each of the real and imaginary parts. Hence, a particular set of velocity, pressure and buoyancy fields is a unique point there. As the flow evolves,

these fields change in time and form a trajectory. A schematic view of phase space is portrayed in figure 1–1, where two slow variables are put on the x- and y- axes and a fast variable on the z- axis. Note that this is merely a schematic representation as the flow may have infinitely many slow and fast modes. The constraint (1.26) forms a hyper-surface or a manifold in the phase space which contains the trajectories governed by balance dynamics. Since balance models are constructed to eliminate the fast-time evolution, this hyper-surface is called the slow manifold. If balance models fully describe the dynamics, this manifold has to be invariant.

Balance relations and balance dynamics can be visualised in figure 1–1 as well. The full equations (1.25) map a balanced initial condition, $(s(t_o), f(t_o))$, directly to $(s(t), f(t))$, which is the solution at time t. Instead of following this trajectory, balance models project the initial condition onto the horizontal plane, i.e. $(s(t_o), 0)$, using the balance relation (1.26). Then balance dynamics defined by (1.27) give the slow variables at the desired point in time, i.e. $(s(t), 0)$. This point can be mapped back on the slow manifold to derive $(s(t), f(t))$. In figure 1–1, the path that the full equations and the balance model follow are marked by (solid) black and (dashed) red lines, respectively.

As will be discussed in detail in the introductions of chapters 2-4, many studies have shown there is no exactly invariant slow manifold; however, there might be many non-invariant slow manifolds of various accuracy. This means trajectories that start from the manifold do not necessarily stay on it. However, for small Ro and Fr, this departure may take place at very small scales and only at later times. Therefore,

if the Rossby and Froude numbers are small and only the large scales are the focus, the slow manifold is a useful approximation.

Early studies of the slow manifold and balance dynamics mainly employed low-order dynamical systems, which are often derived by modal truncation of the governing equations. These approaches have more historical value than real applications in atmospheric and oceanic sciences, as geophysical flows are fully turbulent and depend on a large number of modes. Among others, one can refer to Lorenz (1980), Warn (1997), Warn & Menard (1986), Lorenz (1986), Lorenz & Krishnamurthy (1987), Camassa (1995) and Bokhove & Shepherd (1996) to learn more about finite-dimensional balance models. For a more comprehensive review of the subject, interested readers are referred to a recent study by Chekroun *et al.* (2017).

1.4 Turbulence

"*I am an old man now, and when I die and go to heaven, there are two matters on which I hope for enlightenment. One is quantum electrodynamics and the other is the turbulent motion of fluids. About the former, I am really rather optimistic.*"

"*There is a physical problem that is common to many fields, that is very old, and that has not been solved. It is not the problem of finding new fundamental particles, but something left over from a long time ago -over a hundred years. Nobody in physics has really been able to analyse it mathematically satisfactorily in spite of its importance to the sister sciences. It is the analysis of circulating or turbulent fluids.*"

The latter quote is by Richard Feynman,[7] one of the most celebrated contemporary physicists, and the former by Sir Horace Lamb, a renowned mathematician and fluid dynamicist almost a century before Feynman. These, among others, are frequently seen and heard in the openings of scholarly talks, theses and books on turbulence. In addition to their beautiful witticism, they speak about the complications of this phenomenon – a complication that starts with its definition.

As classical textbooks in turbulence (e.g. Tennekes & Lumley, 1972; Lesieur, 2012) admit, giving a precise definition of turbulence would be very difficult if not impossible. Following their approach, instead of defining turbulence we try to describe it with its main characteristics:

[7] Taken from "*The Feynman Lectures on Physics Vol 1.*" See Feynman *et al.* (1963).

Randomness and chaos This might be the most sensible and intuitive property of such flows. Turbulent motion is very irregular and unpredictable. As a signature of chaotic motions, a slight change in initial conditions can lead to radically different solutions in time. Therefore, it is rather impossible to deterministically follow the trajectories of fluid particles or determine the instantaneous values of flow variables (such as velocity and pressure). Instead, the study of turbulence focuses on the statistics of flow variables. Interestingly, despite their absolute randomness, turbulent flows display a number of universal statistics. Some of these will be briefly reviewed in the following subsections.

Diffusivity The diffusivity of turbulence increases the transfer of momentum, mass and heat and enhances the mixing in the fluid. Turbulence can be pictured as an imaginary whisk that stirs the flow variables such as velocity and temperature and consequently increases the fluxes of quantities such as heat and momentum.

A wide range of spatial and temporal scales The spatial scales of a fluid flow correspond to wavenumbers in Fourier space. Dynamics with a limited number of modes cannot be considered turbulent even if they are chaotic. The Reynolds number defined in (1.14) is directly related to the ratio of the largest to smallest scales in a flow. Hence, one can say turbulence always occurs at large Reynolds numbers. A turbulent flow also contains a broad distribution of time scales. Therefore, balance dynamics may not hold over all length scales.

Nonlinear interactions The spectral (Fourier) modes of a flow evolve independently if there is no nonlinearity in the system (and coefficients in the equations

are constant). As seen in equation (1.1), there are nonlinear terms in the Navier-Stokes equations, but in some flows they are relatively small compared to other terms. If a flow has a large number of modes, it is likely that they interact nonlinearly. Therefore, this characteristic of turbulence often occurs with the preceding one. However, we counted this as a separate property to emphasise the importance of these interactions.

As mentioned in section 1.1, the largest scales in the atmosphere and ocean are many orders of magnitude larger than the dissipation scales (where viscosity is important). Most of the scales between the two bounds interact nonlinearly. The inaccuracy of our weather forecasts is often rooted in the chaos and randomness. Hence, we can easily claim that the ocean and atmosphere are turbulent. As a matter of fact, they display different forms and regimes of turbulence at different scales. We first introduce different types of turbulence in sections 1.4.2-1.4.3, and then we situate them in the atmosphere and oecan in section 1.5.

1.4.1 Theory of 3D isotropic turbulence

In introducing different theories of turbulence, we start by citing the most prominent figures of each field whose contributions led to major developments. Kolmogorov (1941) should be honoured for the most remarkable contribution to the theory of isotropic turbulence. His phenomenology only applies to the very small scales in the atmosphere and ocean that are not the focus of this study. Nevertheless, it can be viewed as the foundation of several other theories used heavily in this thesis. Hence, it is reviewed in this section.

The theory of isotropic turbulence is based on the cascade of energy from larger to smaller eddies. This notion was first introduced by Richardson (1922), who wrote:

"Big whirls have little whirls that feed on their velocity,

and little whirls have lesser whirls and so on to viscosity."

As his eloquent words describe, the larger eddies transfer their energy to smaller eddies, and then they transfer the received energy to even smaller ones. This cascade of energy continues to the very small scales where it is finally dissipated by viscosity.

Elaborating on the concept of energy cascade, Kolmogorov's phenomenology predicts how much energy is contained at a particular spatial scale. In simple words, it explains how the total energy of all eddies with radius R scale with their R. Five assumptions can be viewed as the building blocks of Kolmogorov's theory: isotropy, homogeneity, statistical stationarity, large values of the Reynolds number and locality of energy transfer between scales. An isotropic flow does not have a preferred direction statistically. In homogeneous flows the statistics of each point in space are the same as at all other points. These conditions do not hold in the vicinity of solid boundaries as there is a preferred direction on the boundary and the variables are not homogeneous as we move away from the boundary. Therefore, Kolmogorov's theory should be applied to periodic boundary conditions or far in the flow interior where the presence of solid boundaries are not felt. Considering the small thickness of geophysical boundary layers, these assumptions are not too constraining. If the flow is statistically stationary, the variables do not depend on time when they are ensemble averaged. We will explain what we mean by the last condition, locality of transfer, later.

Since Kolmogorov's theory concerns different scales of a flow, it is suitable to transform the flow variables such as velocity to Fourier space. In this space we consider the amplitude of velocity at each wavenumber as opposed to the velocity of each point in real space. As defined in (1.13), the velocity field may be written in terms of its Fourier modes, $\hat{\mathbf{u}}(\mathbf{k})$. These can be seen as the amplitude of oscillations with the wavevector $\mathbf{k}$. Superposition of all such oscillations retrieves the function $\mathbf{u}(\mathbf{r})$. After describing the variables in Fourier space, we can define the modal energy of the wavevector $\mathbf{k}$ as

$$\mathcal{E}(\mathbf{k}) = \frac{1}{2}\,\hat{\mathbf{u}}(\mathbf{k}) \cdot \hat{\mathbf{u}}(\mathbf{k})^* \tag{1.28}$$

($\hat{\mathbf{u}}^*$ denoting the complex conjugate of $\hat{\mathbf{u}}$). The energy density per unit mass at each point in real space is defined as

$$\mathtt{E}(\mathbf{r}) = \frac{1}{2}\,\mathbf{u}(\mathbf{r}) \cdot \mathbf{u}(\mathbf{r}). \tag{1.29}$$

Using Parseval's theorem, $\mathcal{E}$ and $\mathtt{E}$ can be related to each other as below

$$\tilde{E} = \frac{1}{\mathtt{V}} \int_D \mathtt{E}(\mathbf{r})\; d\mathbf{r} = \int\limits_{\mathbf{k}=-\infty}^{+\infty} \mathcal{E}(\mathbf{k})\; d\mathbf{k} \tag{1.30}$$

in which $\mathtt{V}$ is the volume of the entire domain, D. Hence, $\tilde{E}$ can be simply viewed as the energy of the system. Knowing that the flow is isotropic (one of the fundamental assumptions of this theory), variables at wavevectors of equal magnitudes should have similar statistics. For instance, if two wavevectors $\mathbf{k}$ and $\mathbf{k}'$ have the same magnitudes ($k = |\mathbf{k}| = |\mathbf{k}'|$), their corresponding modal energies should be statistically equal.

Therefore, if the ensemble average is denoted by $\langle\ \rangle$, we can write $\langle \mathcal{E}(\mathbf{k})\rangle = \langle \mathcal{E}(\mathbf{k}')\rangle$. We then define the energy spectrum as

$$E(k) = \int_{S_k} \mathcal{E}(\mathbf{k})\ d\mathbf{s} \tag{1.31}$$

in which the surface integral is carried over the sphere of radius k denoted by S_k. Therefore, the 3D integral in the last term of (1.30) can be calculated by integrating over the spherical shell of S_k and then summing these shells together as below

$$\tilde{E} = \int_{\mathbf{k}=-\infty}^{+\infty} \mathcal{E}(\mathbf{k})\ d\mathbf{k} = \int \left(\int_{S_k} \mathcal{E}(\mathbf{k}) d\mathbf{s} \right) dk = \int E(k) dk. \tag{1.32}$$

The last term in this equation suggests that $E(k)$ is the energy density contained between scales of k and $k + dk$. Hence, the average energy in (1.32) is the sum of energy at different wavenumbers. The energy spectrum is one of the fundamental quantities of turbulence. For a statistically stationary flow, it can be written as a function of the following variables

$$E = f(F, \epsilon, k, \nu), \tag{1.33}$$

in which F is the forcing that is applied to the flow to maintain the motion. The forcing usually occurs at the larger scales of the flow. ϵ is the energy flux defined as the rate of energy transfer from one scale to another per unit volume. Knowing that energy is input into the flow at forcing scales and is dissipated by visocisity, if the flow is statistically stationary, ϵ can also be viewed as the rate of energy input by the forcing per unit volume or the rate of energy dissipation per unit volume.

At this point, the last two assumptions of Kolmogorov's theory should be exercised: the flow has a large Reynolds number and energy is exchanged locally. The locality hypothesis emphasises that the energy transfer between different modes is more effective when their wavenumbers have similar magnitudes. More particularly, the eddies of bigger size transfer their energy locally to eddies with slightly smaller size and this transfer continues to the very small scales in an energy cascade. It should be noted that the locality assumption does not preclude the interaction of eddies of very different sizes. They can still interact with one another without a significant energy transfer. For example, a large eddy can move very small eddies around without exchanging energy with them, since the larger eddy appears as a mean translation of the small ones without shear, leaving the energetics unchanged. It can be shown that dissipation occurs at very small scales, whereas we assumed the forcing to be exerted only at large scales. These together with the largeness of Re indicate there is a substantial separation of length scales between forcing and dissipation. Hence, there is a range of wavenumbers that feels neither the forcing nor the dissipation. For this range, which is known as the "inertial subrange", it was hypothesised by Kolmogorov (1941) that the energy spectrum no longer depends on F and ν. Hence, (1.33) reduces to

$$E = f_i(\epsilon, k). \tag{1.34}$$

Using Buckingham's π theorem,[8] one finds that (1.34) contains only one dimensionless parameter. Hence, we can write

$$\frac{E}{\epsilon^m k^n} = C \tag{1.35}$$

in which m and n are some exponents, and C is a dimensionless constant. Denoting the unit of time by T and the unit of length by L, the quantities on the left hand side of (1.35) have the following dimensions

$$[E(k)] = L^3 T^{-2}, \quad [\epsilon] = L^2 T^{-3}, \quad [k] = L^{-1}. \tag{1.36}$$

(1.35) then implies $m = 2/3$ and $n = -5/3$ to have similar units on both sides of the equation. Therefore,

$$E(k) = C\epsilon^{2/3} k^{-5/3}, \tag{1.37}$$

which has been verified by numerical studies (e.g. Kaneda *et al.*, 2003), experiments (e.g. Mydlarski & Warhaft, 1996) and observations (e.g. Grant *et al.*, 1962). Figure 1–2 schematically shows the energy spectrum of 3D isotropic turbulence as well as the forcing and dissipation scales.

[8] See e.g. chapter 8 of Kundu & Cohen (2007) for more information on this theorem and dynamic similarity.

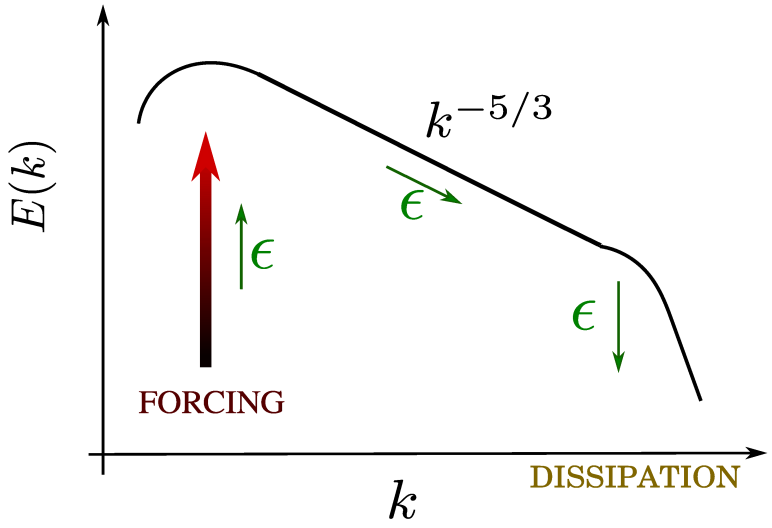


Figure 1–2: The energy spectrum of isotropic turbulence.

1.4.2 2D turbulence

The landmark studies on 2D turbulence are the contribution of Kraichnan (1967, 1971), followed by the work of Batchelor (1969) and Leith (1971). Neither atmospheric nor oceanic flows are two dimensional. Nevertheless, the theory of 2D turbulence was shown by Charney (1971) to be the cornerstone of geostrophic turbulence, which explains the statistics of large-scale geophysical flows very well. This is because quasi-geostrophic and 2D flows have similar invariants. Before explaining this theory, we investigate the differences between 2D and 3D fluid flows. Taking the curl of (1.1) and assuming the flow to be incompressible and inviscid, we derive the following equation for the budget of vorticity ($\boldsymbol{\omega} = \nabla \times \mathbf{u}$)

$$\frac{\partial \boldsymbol{\omega}}{\partial t} + (\mathbf{u} \cdot \nabla)\, \boldsymbol{\omega} = (\boldsymbol{\omega} \cdot \nabla)\, \mathbf{u}. \tag{1.38}$$

For 2D flows, only the vertical component of vorticity ($\boldsymbol{\omega} \cdot \hat{\boldsymbol{z}} = \zeta$) remains. Therefore, (1.38) reduces to

$$\left(\frac{\partial}{\partial t} + \mathbf{u}_h \cdot \nabla_h\right) \zeta = 0. \tag{1.39}$$

in which $\mathbf{u}_h = (u, v)$ and $\nabla_h = (\partial/\partial x, \partial/\partial y)$. For an incompressible 2D flow, the conservation of mass enables us to define the stream function as below [9]

$$u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x}. \tag{1.40}$$

Therefore, the equation of motion in (1.39) can be written in terms of the stream function as

$$\left(\frac{\partial}{\partial t} + \mathbf{u}_h \cdot \nabla_h\right) \nabla_h^2 \psi = 0, \tag{1.41}$$

The energy of 2D flows can then be expressed as

$$\mathtt{E} = \frac{1}{2} \iint \nabla_h \psi \cdot \nabla_h \psi \; dxdy. \tag{1.42}$$

(1.41) shows that in the absence of viscosity the vorticity is conserved on each trajectory of the flow. The conservation of vorticity is stronger than the conservation of energy in a sense that it is conserved pointwise, whereas energy is conserved once

[9] The stream function was previously defined for QG flow in (1.20) which could depend on z as well. The 2D stream function has a similar definition but depends only on x, y and time.

integrated over the entire domain. Therefore, any analytic function of vorticity, $F(\nabla_h^2\psi)$, is also conserved and can be seen as a new invariant of the dynamics. Of particular interest is the following quadratic function of vorticity named enstrophy

$$\mathtt{Z} = \frac{1}{2}\iint \left(\nabla_h^2\psi\right)^2 \, dxdy. \tag{1.43}$$

Like the energy spectrum, the enstrophy spectrum, $Z(k)$, can be defined as

$$Z(k) = \oint_{\Gamma_k} \left((k_x^2 + k_y^2)\hat{\psi}\right)^2 d\boldsymbol{s}, \tag{1.44}$$

in which Γ_k are rings of radius k, and $\hat{\psi}$ is the Fourier transform of ψ. This spectrum can be related to the energy spectrum defined in (1.43)

$$\tilde{Z} = \frac{\mathtt{Z}}{\mathtt{A}} = \int Z(k)dk = \int k^2 E(k)dk, \tag{1.45}$$

in which $\mathtt{A}$ is the surface area of the domain. The same scenario as in section 1.4.1 for energy in 3D flows can be restated for enstrophy in 2D flows. Just like energy, enstrophy is injected into the flow by the forcing and is dissipated by viscosity. Therefore, if there is a separation between the forcing and dissipation scales, enstrophy can be cascaded downscale. By anology with ϵ, η is defined as the flux of enstrophy which is transferred locally from one scale to another. It can also be viewed as the rate of enstrophy dissipation per unit volume if the flow is statistically stationary.

Let's assume a steady flux of enstrophy and energy are injected into the flow at the forcing scales. These fluxes are transferred through the inertial subrange (if such a subrange exists). With similarity analysis, it can be shown that ϵ and η cannot

coexist in an intertial subrange (see Kraichnan, 1967). If $\epsilon = 0$ and $\eta \neq 0$, the energy spectrum will be a function of the enstrophy flux and wavenumber, $E = f_\eta(\eta, k)$. Therefore, a dimensional analysis similar to that presented in (1.37) leads to

$$E(k) = \mathtt{K}_\eta \ \eta^{2/3} k^{-3}. \tag{1.46}$$

On the other hand, if $\epsilon \neq 0$ and $\eta = 0$, the result of (1.37) will be repeated

$$E(k) = \mathtt{K}_\epsilon \ \epsilon^{2/3} k^{-5/3}. \tag{1.47}$$

This prompts the question of which one of these inertial subranges occurs. So far, we did not consider the signs of η and ϵ (or the direction of energy and enstrophy cascades). Using statistical mechanics, it can be shown that enstrophy is cascaded *forward* toward the dissipation range and energy displays an inverse cascade toward smaller wavenumbers (Kraichnan, 1967). If the flow is forced at some intermediate scale k_F, the forward cascade of enstrophy occurs at $k > k_F$ and the inverse cascade of energy at $k < k_F$. This is shown schematically in figure 1–3 where the direction of cascades as well as the slopes of the spectra are shown.

1.4.3 Geostrophic turbulence

The most significant contribution to geostrophic turbulence should unequivocally be accredited to Charney (1971). His work gave birth to this subject, which was followed by a number of other studies. It was an answer to a paradox on the

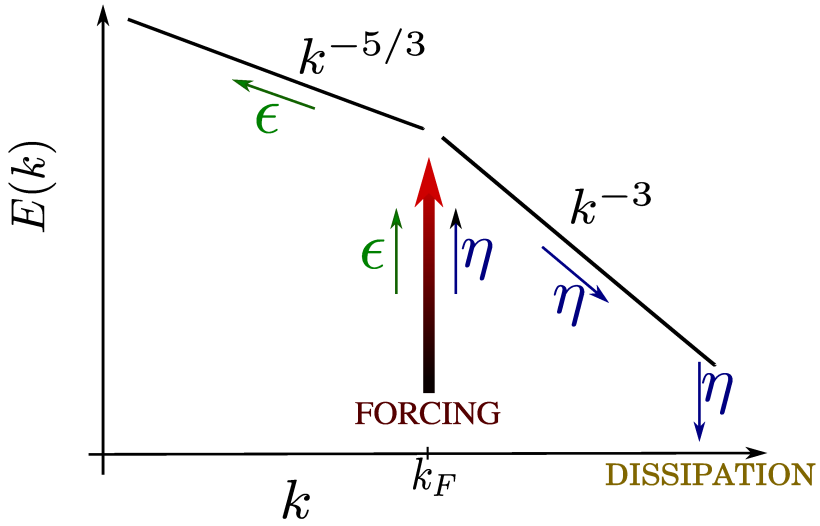


Figure 1–3: Energy spectrum of 2D turbulence with both forward cascade of enstrophy and inverse cascade of energy.

synoptic-scale atmospheric spectrum. On one hand, it had been reported in preceding observations (Horn & Bryson, 1963; Wiin-Nielsen, 1967; Julian *et al.*, 1970)[10] that the large-scale atmospheric energy spectrum displays an approximate scaling of k^{-3}, which was similar to that of the 2D enstrophy cascade in Kraichnan's (1967) theory. On the other hand, it was also clear from the observations that the large-scale atmosphere is baroclinic and hence not 2D.

Charney's theory is based on QG, which reliably describes the dynamics in strongly rotating flows in the large-scale atmosphere and ocean. Therefore, it can

[10] The observations after Charney (1971) with a wider range of data and more accuracy confirmed these scalings. For instance, see Nastrom & Gage (1985) and Callies *et al.* (2016).

explain the observations in this range of scales. To simplify the argument, we assume periodic boundaries and constant N and f.[11] However, the theory of geostrophic turbulence can be extended to some solid boundaries and variable N and f (see e.g. Charney (1971) or Salmon (1998)). The total energy

$$\mathsf{E} = \int_V \left(\nabla\psi \cdot \nabla\psi + \frac{f^2}{N^2} \frac{\partial\psi}{\partial z}^2 \right) dV = \int_V \left(u^2 + v^2 + \frac{f^2}{N^2} b^2 \right) dV \tag{1.48}$$

is conserved, in which $\int_V () dV$ is the volume integral over the entire domain. Note that QG flows can have vertical velocity, w, when the Boussinesq equations are expanded up to second order in Ro. However, the vertical velocity component does not evolve independently and can be determined by the omega equation (see e.g. Hoskins *et al.*, 1978). This is an example of the relation (1.26).

As discussed in section 1.3.1, q (indeed any function of q) remains constant on each inviscid trajectory of QG flows. Hence, q^2, known as potential enstrophy, is conserved

$$\mathsf{Z} = \int_V q^2 \, dV. \tag{1.49}$$

The flow described by (1.22) is not isotropic, but Charney (1971) suggested the following change of coordinate to make the scaled flow (nearly) isotropic

[11] Our numerical simulations in the rest of this thesis also satisfy these conditions.

$$\tilde{\mathrm{z}} = \frac{N}{f} z. \tag{1.50}$$

In the new coordinate, (1.22) and (1.21) transform to

$$\frac{\partial}{\partial t}\tilde{\boldsymbol{\nabla}}^2\psi + u\frac{\partial}{\partial x}\tilde{\boldsymbol{\nabla}}^2\psi + v\frac{\partial}{\partial y}\tilde{\boldsymbol{\nabla}}^2\psi = 0 \tag{1.51}$$

in which $\tilde{\boldsymbol{\nabla}}$ is the 3D gradient vector in the new coordinate,

$$\tilde{\boldsymbol{\nabla}} = \big(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial \tilde{\mathrm{z}}}\big), \tag{1.52}$$

which is an isotropic operator, while (1.51) is still anisotropic. Energy and potential enstrophy take the forms below in the new coordinate

$$\mathsf{E} = \int_V \big(\tilde{\boldsymbol{\nabla}}\psi \cdot \tilde{\boldsymbol{\nabla}}\psi\big)\, dV, \quad \mathsf{Z} = \int_V \big(\tilde{\boldsymbol{\nabla}}^2\psi\big)^2\, dV. \tag{1.53}$$

(1.53) resembles the invariants of 2D turbulence, defined in (1.42) and (1.43), respectively. The only differences are that ∇_h is replaced by $\tilde{\boldsymbol{\nabla}}$, and the surface integral by the volume integral. The theory of 2D turbulence, which is based on similarity analysis and statistical mechanics, depends only on the form of its invariants (energy and enstrophy). Considering their similarity, the same argument presented for 2D turbulence can be repeated for QG in the stretched coordinate, noting that there are some subtleties in the precise definition of the energy spectrum.

The major forcing of the earth's atmosphere is by the sun, which occurs at scales larger than synoptic scales. Similar to 2D turbulence, it can be argued that (potential) enstrophy is cascaded forward in the range of synoptic wavenumbers

which are larger than the forcing wavenumber. Therefore, the energy spectrum of the corresponding inertial subrange is of the form

$$E(\tilde{k}) = \mathtt{K}_Q \ \eta^{2/3}\tilde{k}^{-3} \tag{1.54}$$

in which $\tilde{k} = (k_x^2 + k_y^2 + k_{\tilde{z}})^{1/2}$ and $\mathtt{K}_Q$ is a universal constant. Asssuming isotropy, the energy should be equipartitioned among its components. We label these by E_x, E_y and $E_{\tilde{z}}$.

$$E_x(\tilde{k}) = \oint\limits_{S_{\tilde{k}}} k_x^2\hat{\psi}^2 \ d\mathbf{s}, \quad E_y(\tilde{k}) = \oint\limits_{S_{\tilde{k}}} k_y^2\hat{\psi}^2 \ d\mathbf{s}, \quad E_{\tilde{z}}(\tilde{k}) = \oint\limits_{S_{\tilde{k}}} k_{\tilde{z}}^2\hat{\psi}^2 \ d\mathbf{s} \tag{1.55}$$

in which $S_{\tilde{k}}$ is the sphere of radius $\tilde{k}$ and $\hat{\psi}$ is the Fourier transform of the stream function in the stretched coordinate. If there is equipartition of energy, E_x, E_y and $E_{\tilde{z}}$ will have the same form as (1.54). E_z is the (available) potential energy once it is transformed back to the original coordinate. Therefore, Charney's theory predicts the logarithmic slope of -3 for potential as well as kinetic energy spectra, a fact that corroborates the observations reported by Nastrom & Gage (1985).

Charney's theory derives the 3D energy spectrum, $E(\tilde{k})$, as a function of the magnitude of the 3D wavevector in the stretched coordinate, $\tilde{k}$. The three-dimensional isotropic energy spectrum may be related to the horizontal isotropic energy spectrum by (cf. Batchelor, 1953)

$$E(k_h) = k_h \int_{k_h}^{\infty} \frac{E(\tilde{k})}{\tilde{k}(\tilde{k}^2 - k_h^2)^{1/2}} d\tilde{k}. \tag{1.56}$$

If $E(\tilde{k})$ is locally equal to $\mathcal{C}\tilde{k}^{-n}$ with $n > 1$ and at very large wavenumbers has a slope of less than 1, the integral in (1.56) can be approximated by

$$E(k_h) \approx k_h \int_{k_h}^{\infty} \frac{\mathcal{C}\tilde{k}^{-n}}{\tilde{k}(\tilde{k}^2 - k_h^2)^{1/2}} d\tilde{k} = k_h^{-n} \int_1^{\infty} \frac{\mathcal{C}\kappa^{-n}}{\kappa(\kappa^2 - 1)^{1/2}} d\kappa = \mathcal{C}' \; k_h^{-n}. \qquad (1.57)$$

Therefore, under these conditions, $E(k_h)$ displays the same power-law dependence as $E(\tilde{k})$. The observations and numerical studies mostly report horizontal spectra. Considering the above argument, these can be compared to the results of Charney's theory.

The analogy between Charney's theory and 2D turbulence does not fully hold for the assumption of isotropy. It is true that the spatial derivatives of ψ in Charney's stretched coordinate are the same. Nevertheless, q is advected only by horizontal velocities in (1.51). Despite this, large-scale atmospheric and oceanic observations support the prediction of this theory. As a matter of fact, observations agree with the k^{-3} scaling better than numerical QG simulations, which often have steeper spectra.

1.5 Atmospheric energy spectrum

So far this chapter has tried to familiarise the reader with the knowledge of different theories of turbulence and balance dynamics. In this section, we demonstrate how these are employed to explain various features of the atmospheric and oceanic energy spectra. Figure 1–4, which is taken from Nastrom & Gage (1985), is one of the landmark observational studies on the atmospheric spectrum that includes both synoptic scales and mesoscales with reliable statistics. Prior to this study, there

were other reports on the synoptic (Horn & Bryson, 1963; Wiin-Nielsen, 1967; Julian *et al.*, 1970) and mesoscale (Vinnichenko & Dutton, 1969; Pinus *et al.*, 1967) spectra separately, featuring similar statistics. The data used in Nastrom & Gage (1985) was obtained from over 6000 commercial aircraft flights as a part of the Global Atmospheric Sampling Program (GASP). Since then a number of studies have been performed to explain the characteristics of figure 1–4 or to reproduce it from other observations. For instance, Lindborg (1999) used MOZAIC data with higher accuracy and expressed that "the results are, on the whole, similar" to those derived by Nastrom & Gage (1985). In the synoptic range of scales (up to wavelength 500 km), the zonal and meridional velocities as well as potential temperature display a slope of -3, which is consistent with Charney's theory of geostrophic turbulence in the subrange of forward enstrophy cascade (discussed in section 1.4.3). Note that even if we assume that the flow is 2D in this range of scales (which is an unrealistic assumption), the theory of 2D turbulence on its own cannot explain the slope of potential temperature – a fact that highlights the contribution of Charney's theory. Depending on the particular features of a flow, more accurate balance models than QG can be employed over this range of scales.

At wavelengths smaller than 500 km, there is a sharp transition to a shallower slope of -5/3 in figure 1–4. As much as there is a consensus on how to explain the range of k^{-3}, there is controversy on the explanation of this shallow range. The early explanation for the emergence of the -5/3 slope was the inverse cascade of energy similar to 2D turbulence (see section 1.4.2). In studies such as Gage (1979), Lilly (1983) and Falkovich (1992), it was hypothesised that this subrange is a form of

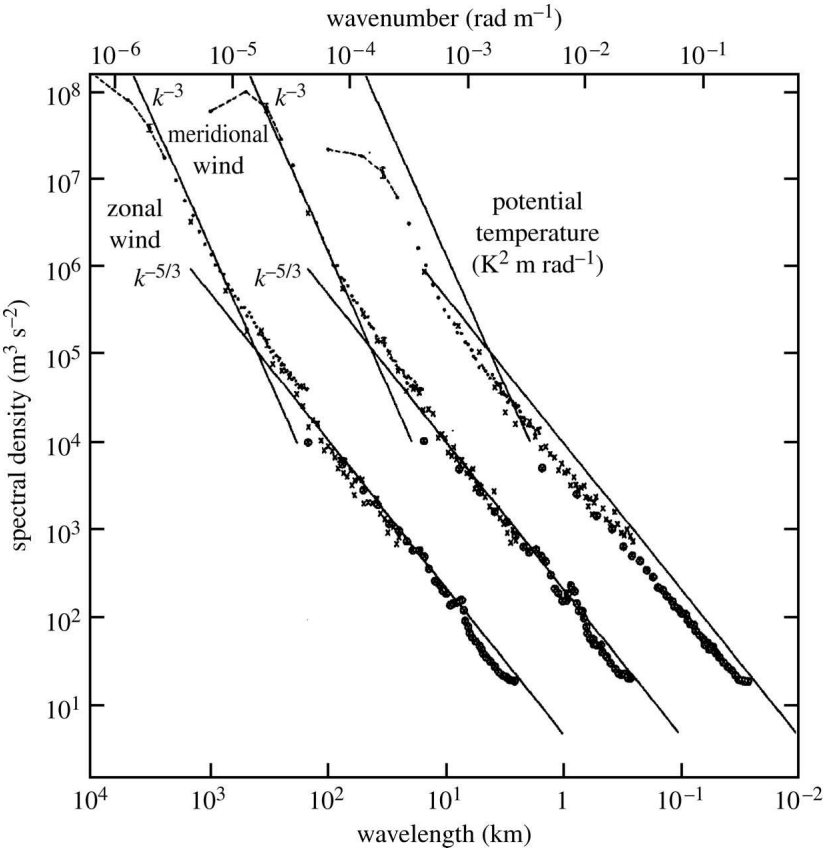


Figure 1–4: The wavenumber spectra of zonal (u) and meridional (v) velocity and potential temperature taken from Nastrom & Gage (1985). The guiding lines indicate the slopes of -3 and -5/3. For better visualisation, the meridional velocity and potential temperature are shifted to the right for one and two decades, respectively.

layerwise 2D stratified turbulence with a flux of energy toward larger scales and no significant cascade of potential enstrophy. The source of this inverse energy cascade was attributed to convective forcing at even smaller scales. There have been many arguments against this hypothesis. First, although Ro becomes larger in this regime the Coriolis term is still important in the equations. Therefore, theories based on purely stratified turbulence may not be applied. Second, numerical studies show

that strong stratification favours the forward cascade of energy (see e.g. Herring & Métais, 1989; Laval *et al.*, 2003; Riley & deBruynKops, 2003; Lindborg, 2006). Third, a number of studies using Global Climate Models (GCM) and Numerical Weather Prediction (NWP) models were able to retrieve a $-5/3$ spectrum without forcing at such small scales (Skamarock, 2004; Hamilton *et al.*, 2008; Evans *et al.*, 2013). Fourth, it is hard to explain the energetics of the atmosphere and ocean with this hypothesis. If at synoptic scales there is no significant cascade of energy and at mesoscales there is an inverse cascade of energy, the energy keeps accumulating at larger scales. Fifth, the break in spectral scaling is also observed in the ocean (e.g. Klymak & Moum, 2007*a*; Callies & Ferrari, 2013), where there is comparably less convection. Therefore, one may speculate that there is one mechanism behind it in both fluids.

The other early hypothesis, offered by Dewan (1979) and followed by VanZandt (1982), was based on the forward cascade of gravity (buoyancy) waves. Dewan (1979) decomposed the flow into three components – mean flow, waves and turbulence – and derived the budget of their energies. Then, he argued that the wave part of the flow can display a Kolmogorov-like cascade of energy. Therefore, he used the same similarity analysis to derive the $k^{-5/3}$ scaling. One may cast doubt on the validity of Kolmogorov's assumptions for the wave part of the flow, while it interacts with the mean and turbulent part. Aside from this, his theory suggests a cascade that does not depend on rotation. Many idealised numerical simulations (e.g. Bartello (2010) and Vallgren *et al.* (2011)) as well as GCM studies (e.g. Hamilton *et al.* (2008) and Evans *et al.* (2013)) have shown that the shallow part of the energy

spectrum depends on the strength of rotation. More specifically, the flux of forward energy and the wavenumber at which the spectrum starts shallowing are functions of *Ro*. We discuss in detail how the kink of the energy spectrum changes with *Ro* in chapter 2 as well. More recently, Callies *et al.* (2014, 2016) claimed that the shallowing of the mesoscale spectrum is due to weakly nonlinear inertia-gravity waves. These authors first suggested a wave-vortex decomposition method based on the linearised equations to decompose the one-dimensional spectra of the observed horizontal velocity and buoyancy (Bühler *et al.*, 2014). Then, they applied this decomposition to the atmospheric data in Callies *et al.* (2014, 2016) and reported that in the mesoscale subrange the energy spectrum of wave modes dominates that of vortical modes. Hence, they concluded that inertia-gravity waves dominate the dynamics in this range of scales. One can argue that even if the energy spectrum of wave modes is larger, they can still nonlinearly interact with each other and also with vortical modes. Lindborg (2015) argued that the flow is strongly nonlinear in this range of scales. Therefore, linear wave-like behaviour cannot dominate the dynamics. Additionally, the frequency analysis of idealised simulations, which are offered in chapter 4 as well as in Asselin *et al.* (2017), do not agree with the dominance of inertia-gravity waves. These results show that the wave modes peak around the linear frequency at scales larger than the mesoscales, but at mesoscales they do not display a wave-like frequency spectrum.

The other stream of research associates the shallowing in the mesoscale spectrum with a forward cascade of energy due to strong nonlinear interactions. A subgroup of these studies, such as Lindborg (2006) and Lindborg (2015), explain the $k^{-5/3}$

spectrum by stratified turbulence, in which rotation does not play a significant role (some of the stratified-turbulence hypotheses do not hold when rotation is included in the dynamics). Others emphasise the importance of rotation in tandem with stratification (Bartello, 2010; Vallgren *et al.*, 2011; Deusebio *et al.*, 2013; Nadiga, 2014). In other words, they describe the mesoscale dynamics as a form of rotating stratified turbulence with a forward cascade of energy. A number of these studies claim that the ageostrophic part of the flow accounts for the scaling of $k^{-5/3}$ (see e.g. Bartello (2010) and Deusebio *et al.* (2013)).

A more recent hypothesis attributes the scaling of the mesoscale spectrum to the sharp change of stability at the tropopause (Tulloch & Smith, 2006, 2009). Surface quasi-geostrophy (SQG) has been proposed to model the dynamics near the tropopause (see e.g. Blumen, 1978; Lapeyre, 2017). Using a finite-depth SQG model, Tulloch & Smith (2006) showed that the energy spectrum transitions from a -3 to a $-5/3$ slope. Asselin *et al.* (2016) questioned whether SQG dynamics can be maintained in a Boussinesq simulation when the flow is strong. In a follow-up paper, the same authors demonstrated that, in the regime of weak flows (characteristic velocity of less than 1 m/s), the existence of the tropopause leads to a shallow spectrum (Asselin *et al.*, 2017). However, for more typical velocities of atmospheric flows (larger than 10 m/s), the statistics of flows with and without the tropopause are similar. Therefore, they inferred that the ageostrophic dynamics play a more important role in the shallowing of the mesoscale spectrum than tropopause-induced effects. Hence, their arguments support the rotating-stratified-turbulence hypothesis.

Similar to figure 1–4, the break in spectral scaling is also reported for the ocean but at a different scale (see e.g. Callies & Ferrari, 2013, and references therein). Considering the values of dimensionless parameters, oceanic mesoscales may be seen as the dynamical equivalent of atmospheric synoptic scales, and likewise oceanic submesoscales as the equivalent of atmospheric mesoscales. Therefore, this research can be applied to both atmospheric and oceanic flows. To be more inclusive, we use the term sub-deformation scales to refer to both mesoscales in the atmosphere and submesoscales in the ocean.

1.6 Thesis objectives

Having the observations of atmospheric spectrum in mind, this thesis aims to elucidate the following three issues:

1. Breaking of the spectral scaling in the energy spectrum

As explained in section 1.5, the atmospheric and oceanic energy spectra display a kink where the geostrophic slope of -3 changes to the sub-deformation slope of -5/3. The different hypotheses that try to explain this transition were reviewed in the previous section. Among them, 2D-like stratified turbulence with an inverse cascade of energy is refuted with a relative consensus. The current debate is whether tropopause-induced effects, inertia-gravity waves or rotating stratified turbulence can better explain this phenomenon. This thesis is a supporting argument for the rotating-stratified-turbulence hypothesis.

This research is performed in a turbulence platform with periodic boundaries. This choice of boundary as well as the exclusion of moist processes and salinity distance our simulations from the real atmosphere and ocean. Therefore, we cautiously

avoid claiming that this study explains the characteristics of the atmospheric energy spectrum. Nevertheless, we present a mechanism that explains the transition from a balanced steep spectrum to an unbalanced shallow one, which is both robust and consistent with the observations of Nastrom & Gage (1985) and Klymak & Moum (2007*a*,*b*). We initialise the simulations with balanced conditions so that the steep part of our energy spectrum contains high-order balance dynamics. This initialisation makes our simulations more realistic as the atmosphere and ocean are balanced at large scales.

2. Quantifying the breakdown of balance dynamics

As mentioned above, balance dynamics predominate in the steep subrange in figure 1–4. Starting from a balanced initial condition with a steep spectrum, we study how a shallow tail develops in time. We show that this shallow tail is the unbalanced ageostrophic part of the flow. Therefore, by analysing its formation, we study the generation of imbalance from a balanced initial condition.

It has been observed that classical QG breaks down at smaller scales when it is introduced in the Boussinesq equations. By considering higher-order balance than QG, we investigate if this still occurs and if so, after how long and at what scales.

The balance initialisation that we use in this study can be viewed as a projection on the hypothetical slow manifold. By changing the flow parameters such as Ro and Fr, we study if the solution stays close to the slow manifold. In other words, we investigate if the slow manifold is invariant or how good that is as an approximation.

3. Internal route to dissipation

One of the important subjects in geophysical sciences is the budget of energy in the atmosphere and ocean. As discussed for figure 1–4, at synoptic scales there is no significant cascade of energy to smaller scales. On the other hand, the energy of the atmosphere, which comes from solar radiation at very large scales, is dissipated at very small scales in the dissipation range (order of centimetres and millimetres). This raises the question of how the large-scale energy is transferred to small scales.

Solid boundaries (earth's topography) are one of the ways that can make this happen. However, there have been several studies showing that even without boundaries, geophysical flows are capable of transferring the balanced energy to unbalanced motion (see e.g. Molemaker *et al.*, 2010; Bartello, 2010; Deusebio *et al.*, 2013; Nadiga, 2014)). This study tries to shed light on this mechanism, which has been termed "internal route to dissipation" (Molemaker *et al.*, 2010).

By thorough investigation of the transfer spectra, we show how energy is leaked from the balanced large-scale flow and eventually cascaded forward by the interactions between the unbalanced modes at smaller scales. This route to dissipation can explain, for example, how energy is dissipated in the ocean interior far from boundaries. Here, we study the internal route to dissipation to be able to compare it with boundary-layer dissipation.

CHAPTER 2

Balance dynamics in rotating stratified turbulence

We begin this work with an implementation of high-order balance in turbulence simulations. Previous studies such as Bartello (2010) and Deusebio *et al.* (2013) have shown that the linearly balanced initial conditions or forcing breaks down at smaller scales. In linear balance, the flow has geostrophic modes but it is void of ageostrophic dynamics (for the definition of geostrophic and ageostrophic dynamics, one can refer to Appendix A or section 2.3.2 in this chapter). With an initial condition that is nonlinearly balanced to high orders, we can investigate whether higher-order balance breaks down. If it does, we can see how it is affected by the order of balance and the strength of rotation and stratification. Moreover, we can derive the turbulence statistics of balanced ageostrophic modes since higher-order balance includes ageostrophic modes. The nonlinearly balanced initial conditions can also better simulate large-scale atmospheric and oceanic flows.

To produce a nonlinearly balanced initial condition, we implement an initialisation scheme proposed by Baer & Tribbia (1977) in a turbulence setting. The details of this implementation can be found in Appendix B. Non-hydrostatic Boussinesq simulations are then performed using this balanced initial condition. We study the

characteristics of this initially balanced flow by looking at energy spectra, time series, variables in real space and frequency spectra.

By re-implementing the initialisation scheme after the integration, the flow can be decomposed into a balanced and an unbalanced part, which demonstrates how imbalance was generated in time.

This chapter is based on the following paper:

Kafiabad, H. A., & Bartello, P. (2016). Balance dynamics in rotating stratified turbulence. *Journal of Fluid Mechanics*, 795, 914-949.

2.1 Abstract

If classical quasigeostrophic (QG) flow breaks down at smaller scales, it gives rise to questions of whether higher-order nonlinear balance can be maintained, to what scale and for how long. These are naturally followed by asking how this is affected by stratification and rotation. To address these questions, we perform nonhydrostatic Boussinesq simulations where the initial data is balanced using the Baer-Tribbia Nonlinear Normal Mode Initialization scheme (NNMI), which is accurate to 2nd order in the Rossby number, as the next-order improvement to first-order QG theory. The NNMI procedure yields an ageostrophic contribution to the energy spectrum that has a very steep slope. However, as time passes, a shallow range emerges in the ageostrophic spectrum when the Rossby number is large enough for a given Reynolds number. It is argued that this shallow range is the unbalanced part of the motion that develops spontaneously in time and eventually dominates the energy at small scales. If the initial flow is not nonlinearly balanced, the shallow range emerges at even lower Rossby number and it appears at larger scales. Through numerous simulations at different rotation and stratification this study gives a clear picture of how energy is cascaded in different initially balanced regimes of rotating stratified flow. We find that at low Rossby number the flow mainly consists of a geostrophic part and a balanced ageostrophic part with a steep spectrum. As the Rossby number increases, the unbalanced part of the ageostrophic energy increases at a rate faster than the balanced part. Hence, the total energy spectrum displays a shallow range above a transition wavenumber. This wavenumber evolves to smaller values as rotation weakens.

2.2 Introduction

In the study of rotating stratified flow two avenues of research have been explored. The first focuses on "balance dynamics". In the limit of strong rotation and stratification the linearised governing equations describe a distinct separation between fast and slow time scales. Since atmospheric and oceanic flows vary slowly, balance models try to reduce the dynamics to only the slow subset. In order to make the total nonlinear solution slowly-varying, it is necessary to introduce judiciously some variance in the high-frequency linear modes. This is what we refer to as nonlinear balance and is described in detail below. We emphasise here that, although there are high-frequency linear solutions, there is no *a priori* reason why the total nonlinear solution has to display wave-like motion. In the second avenue, attempts have been made to extend turbulence theory by studying the statistics of rotating stratified flow, mainly the energy and potential enstrophy transfer and their cascades in different limits. In this section, we briefly review both groups. Then, we discuss the focus of this study which is merging these two avenues by looking at balance dynamics from a turbulence perspective. In particular, we investigate how an initially balanced flow evolves into a more general form of turbulence in well-resolved numerical simulations. Throughout this paper, we consider the regimes of low Rossby number (Ro) and low Froude Number (Fr) defined as

$$Ro = \frac{U}{fL}, \quad Fr = \frac{U}{NH} \tag{2.1}$$

where f is the Coriolis parameter, N the Brunt-Väisälä frequency, L the horizontal length scale, H the vertical length scale and U the horizontal velocity scale.

As a result of computational advances, different aspects of high Reynolds number rotating stratified turbulence have been simulated recently. Bartello (2010) used quasi-geostrophic forcing in the non-hydrostatic Boussinesq equations with different Rossby numbers to present large-scale quasigeostrophic flow with a small-scale transition to a more general form of turbulence. Marino *et al.* (2013) changed the relative linear frequency of gravity to inertial waves, N/f, and looked at its effect on the inverse cascade of kinetic energy. Deusebio *et al.* (2013) studied the route to dissipation in strongly rotating stratified regimes by looking at the energy cascades at different rotations. Whitehead & Wingate (2014) numerically analysed the effect of fast ageostrophic modes on slow ones in three different asymptotic limits of stratification and rotation. Pouquet & Marino (2013) showed the existence of a dual cascade in one simulation. Starting from an unbalanced initial condition and randomly forcing a mid-range wavenumber, their simulation showed an inverse cascade of energy at large scales with the slope of -3 invariant to Ro. However, at wavenumbers larger than the forcing a forward cascade of energy was established as Ro increased and the slope of the energy spectrum changed from -3 to $-5/3$.

In all these studies a wide variety of initial conditions or forcings were used but none were nonlinearly balanced. However, a number of studies used geostrophic initial conditions or geostropihc forcing, which can be referred to as linearly balanced. Starting from an unbalanced initial condition, Bartello (1995) showed the ageostrophic energy is cascaded to the dissipation range via a catalytic interaction with geostrophic modes at low Fr and Ro. This can be seen as a form of nonlinear geostrophic adjustment that takes place by transferring ageostrophic energy to

smaller scales and dissipating it. This transfer of ageostrophic energy was seen in the simulations of Deusebio *et al.* (2013) as well. It provides a mechanism to explain the observed reality of the large-scale atmosphere and ocean, which is a slowly-varying balance (or something rather close to it). These scales can be thought of as synoptic scales in the atmosphere and mesoscales in the ocean, where there is little energy with frequencies much higher than planetary motion. Therefore, the fundamental question is whether balanced or unbalanced initial conditions and/or forcing change the turbulence in a fundamental way. We address this question and show that at low Rossby number, initially balanced flow shows considerably different evolution, at least over several eddy turnover times. However, Ro can be increased to the point that initially balanced and unbalanced flows act rather alike. Although this is not surprising, we explore the limits of these regimes.

In studies of balance dynamics the approach is different. Unlike most studies of rotating stratified turbulence, many of these start with a nonlinear balanced state and investigate if it is maintained. Of course, this depends on one's ability to define such a state and to project initial data onto it. If so, and it is indeed found that balance is maintained by the subsequent evolution, then it has been speculated there exists a slow manifold (Leith, 1980; Lorenz, 1980), or at least something close to it. It is strictly defined as an invariant manifold in phase space on which the dynamics are devoid of any high-frequency variability (Leith, 1980; Warn *et al.*, 1995; Ford *et al.*, 2000), implying the maintenance or breakdown of balance can be equated to the degree of invariance of the slow manifold. At the other extreme, in studies where

fast motion is initially present, the issue is whether the state evolves toward the slow manifold.

Numerous studies have discussed the existence of the slow manifold (Warn & Menard, 1986; Warn, 1997; Lorenz, 1986; Lorenz & Krishnamurthy, 1987; Lorenz, 1992; Ford *et al.*, 2000). They show that the solution does not necessarily stay on it but remains close to it in some sense. Equivalently, some high-frequency motion is observed, but its amplitude remains sufficiently small. Following this, Warn & Menard (1986) suggested the concept of a fuzzy manifold. They demonstrated that as Rossby number increases, less information on the fast modes can be deduced from the slowly-varying modes alone, implying they act as independent degrees of freedom. Hence, it is not a manifold in the strict sense, but rather a "thin" region of the full dimension of phase space.

A more recent line of research exploring the non-invariance of the slow manifold is spontaneous wave generation. Ford *et al.* (2000) showed that unsteady moderately-rotating stratified flows "must emit inertia-gravity waves". Vanneste & Yavneh (2004) considered the regime of strongly rotating stratified flows and showed the amplitude of inertia-gravity waves generated from an initially-balanced solution scales with $e^{-a/Ro}$, where a is a constant. By using multi-scale time expansion and assuming small Ro, Zeitlin *et al.* (2003) showed that time splitting is valid for $t \leqslant (fRo)^{-1}$ (f being the Coriolis parameter). Beyond this they derived the next order approximation to QG scaling and showed in this limit the time splitting is incomplete. Therefore, the vortical component and inertial oscillations evolve with similar time scales.

As an example of spontaneous wave generation, vortex dipoles have also been studied. Snyder *et al.* (2007) showed that quasigeostrophic vortex dipoles generate inertia-gravity waves that are close to stationary relative to them. They persisted after a long integration time. Therefore, it was concluded that these were inherent features of the dipoles. Viúdez (2007, 2008) investigated the characteristics of these waves further in his numerical simulations. For a comprehensive review on balance and spontaneous generation, see Vanneste (2013).

Many studies of balance dynamics and the slow manifold are restricted to low-order dynamical systems. Therefore, their results are difficult to generalize to the infinite-dimensional system governed by the Boussinesq equations. For instance, Kreiss & Lorenz (1994) showed the spatially discrete version of their system generated an exactly invariant slow manifold, whereas its spatially continuous counterpart could not. To overcome this shortcoming, at least numerically, we study the breakdown of balance from a turbulence perspective. Unlike low-order dynamical systems, the turbulence platform of this study considers a broad distribution of wavenumbers and their interactions. In our view a very limited number of modes can only present a limited number of triadic Rossby and Froude numbers, whereas in reality these parameters can be considered as a function of scale or wavenumber, k, ranging over many orders of magnitude in geophysical and astrophysical flows.

Even though most studies of rotating stratified flow examined either turbulence or balance aspects, there have been some exploring both. A series of papers by Dritschel and coworkers are highlights of this group. Dritschel & Viudez (2003), first proposed a numerical approach based on integration of a balance and two imbalance

variables. They chose potential vorticity as their balance and ageostrophic horizontal vorticity components as imbalance variables based on geostrophic and hydrostatic balances. In their subsequent papers, they introduced optimal potential vorticity to render their balance more precise than geostrophic and hydrostatic balance (Viudez & Dritschel, 2004). Their definition of balance was constructed to minimize inertia-gravity waves. Using nearly balanced initial conditions in this sense, McKiver & Dritschel (2008) studied properties of rotating stratified turbulence over a range of Rossby numbers. They also extracted the balanced part of the flow using optimal potential vorticity balance to access the degree of imbalance generated in time. In another recent study, Dritschel & McKiver (2015) thoroughly investigated the generation of imbalance in initially balanced flows with different frequency ratios N/f. In this series of studies, the authors found balance to be very robust in the regime of low Rossby numbers and large frequency ratios, yielding steep spectra characteristic of quasigeostrophic flow.

Another recent study that looked at balance dynamics in a turbulence context was carried out by Nadiga (2014). In this paper the author constructed an initial condition that was in linear balance. He then investigated its subsequent evolution in two parallel simulations of the non-hydrostatic Boussinesq and quasigeostrophic equations. Comparing the two simulations using diagnostics of balance, he examined how the unbalanced part of the flow developed.

Motivated by previous work we investigate the effect of the following parameters in limiting the accuracy of balance dynamics (or the degree of approximate invariance of the slow manifold);

1. The order of balance
2. Time
3. The length scale
4. The strength of rotation and/or stratification.

In so doing, we first generate a set of geostrophic data using a QG model and then use a high-order initialization scheme, namely that of Baer & Tribbia (1977) to produce nonlinearly balanced ageostrophic modes. The Baer-Tribbia scheme yields initial data that are slowly-varying at second order in Ro, at least at $t = 0$. These sets of ageostrophic and geostrophic modes provide us with a representation of balance dynamics suitable as initial conditions in a more general model. Starting with these initial conditions we run a set of well-resolved nonhydrostatic Boussinesq simulations to study how and when balance breaks down. The compromise that is made concerns the flow boundaries. Our simulations are carried out in a triply-periodic configuration that is the only possibility in numerical studies of homogeneous turbulence. We explain the limitations and advantages of this choice.

Our platform provides several advantages that suit the detailed study of influential parameters in the breakdown of balance. First, we can see how different orders of balance affect its breakdown. The Baer-Tribbia initialization scheme that we employ can be carried out to an order higher than the QG approximation. In a similar study Bartello (2010) used fully QG initial conditions, which are linearly balanced, and studied their evolution in a nonhydrostatic Boussinesq model with forcing only in the geostrophic modes. Linear balance cannot form a complete representation of the slow manifold as it sets all ageostrophic modes to zero, thereby rapidly generating

high-frequency oscillations. Higher-order nonlinear balance schemes, on the other hand, yield sets of ageostrophic modes producing more slow evolution.

The other advantage of a turbulence theory-based analysis is to look at the breakdown of balance through the cascade and transfer of energy. Below, we rely on the energy spectrum of normal modes. Their evolution in time draws a sufficiently clear picture of the interplay between linear fast and slow modes in a nonlinear context whose temporal variability can take on any time scale in between.

In light of analysing the breakdown of balance we also make reference to the shallow range of the Gage-Nastrom spectrum. Gage & Nastrom (1986) used the Global Atmospheric Sampling Program (GARP) data to calculate the atmospheric kinetic and potential energy spectra. These consisted of two distinct ranges; a steep range of k^{-3}, followed by a more shallow range of $k^{-5/3}$ at smaller scales with a relatively sharp transition between the two. The steep part of the spectrum is explained by quasigeostrophic turbulence resulting from injection of eddy energy at the deformation scale (Charney, 1971). However, QG theory does not predict the existence of the shallow range without invoking other sources or boundaries. To explain this shallow range two separate, but not mutually exclusive, paths have been followed;

1. The effects of external boundaries such as topography (Vanneste, 2013) or internal ones such as the tropopause (Tulloch & Smith, 2006) that break down the balance and create a shallow range
2. The generation of unbalanced motion and more general forms of turbulence that project on other degrees of freedom (Bartello, 1995, 2010; Vallgren *et al.*, 2011; Nadiga, 2014)

Being aware of the important role of boundaries in the breakdown of balance dynamics, we choose to focus on the second issue. In addition to being more tractable statistically, it can conceivably describe the energy cascade in the ocean/atmosphere interior far from boundaries. Moreover, it shows that even without boundaries there is a mechanism that generates unbalanced ageostrophic flow in homogeneous turbulence, even when starting from balanced large-scale dynamics at low Ro and Fr.

Aside from the idealised simulations cited above, the shallow part of the Gage-Nastrom spectrum has been studied with Numerical Weather Prediction (NWP) and Global Climate Models (GCM) recently (Skamarock, 2004; Hamilton *et al.*, 2008; Evans *et al.*, 2013). The models used in these studies can simulate the atmosphere with more realistic geometry, inhomogeneity and physics. Of this group Waite & Snyder (2009) is one of the most relevant to the current paper, since it employed a relatively high vertical resolution with a simple geometry and started from a linearly balanced initial condition describing a baroclinically unstable zonal jet. It was then perturbed with the "fastest-growing gravest normal mode". This unbalanced perturbation was added to make the unstable jet transition to turbulence whose characteristics were then studied. In their simulations the Rossby number increased with time, which led to the shallowing of the energy spectrum at certain vertical levels. Using the Advanced Research core of the Weather Research and Forecast (WRF) model, they examined the energy spectrum at different levels in the troposphere and lower stratosphere.

It should be noted that the dynamics of NWP/GCM models are generally not nearly as well resolved as in turbulence simulations in simplified geometry, since

much of the computational effort is devoted to realistic features such as surface fluxes, convection, radiation, topography and vertical inhomogeneity. Clearly, one of the weak points of such simulations, when compared with turbulence simulations, is the relatively coarse vertical grid. In addition, many global models still employ the hydrostatic approximation. The degree to which it reproduces the true dynamics at scales where balance breaks down is still an open question.

To have reliable statistics of stratified turbulence it was shown that the buoyancy scale, U/N, which is around 1 km near the tropopause, should be resolved (Waite & Bartello 2004, Lindborg (2006),Waite (2011)). It has even been argued that it is necessary to resolve down to the small-scale regime of 3D isotropic turbulence, i.e. the Ozmidov scale, which is on the order of tens of meters (Brethouwer et al, 2007; Bartello & Tobias 2013). Knowing that these resolution requirements may not be as severe with the addition of weak rotation, the authors' view is that it is best to explore them using idealised simulations of only the dynamics in the first instance, without adding all the other complications of meteorology and oceanography.

These scales are not resolved by GCM/NWP models with today's computers. For example, Hamilton *et al.* (2008) used 24 vertical levels from the ground to about 1 hPa, which led to about 1.5 km grid spacing in the upper troposphere. Evans *et al.* (2013) employed 26 vertical levels with a model top at 2.2 hPa, resulting in an even coarser grid. Given that vertical momentum and thermal diffusion restrict the fully nonlinear scales to larger than 1.5 km, it is clear these studies come short of resolving the vertical outer scale, U/N, of stratified turbulence. However, with their sub-grid dissipation schemes these simulations still offer realistic large-scale

dynamics. For this reason, the authors feel their mesoscale transitions must at this point be interpreted with caution.

In addition to providing higher vertical resolution, our idealised configuration enables us to study a wider range of parameters compared to NWP/GCM models. These models are difficult to tune in order to explore the effect of resolution, rotation, stratification and dissipation. Since we model only the dynamics in as simple a geometry as possible, we were able to explore objectively their sensitivity to Rossby, Froude and Reynolds numbers (at much higher effective Reynolds numbers) in order to advance our understanding of these dynamics. For this reason we feel that studies such as this are complementary to studies using more realistic models.

The organisation of this paper is as follows. Section 2.3 lays out the governing equations of motion and briefly describes the normal-mode decomposition used in this paper. The initialization scheme that balances the ageostrophic modes given the geostrophic modes is explained in Section 2.4. Next, we propose a three-step procedure to study the breakdown of balance dynamics. In the results Section 2.6, we show how the initially balanced and unbalanced flow differ in various Rossby number regimes. Our results demonstrate that the balanced part of the ageostrophic energy has a steep spectrum. However, by increasing the Rossby number a shallow unbalanced range emerges. In fact, it seems reasonable to speculate that such a shallow tail in the ageostrophic spectrum will emerge at any Ro after a sufficient time if the Reynolds number is sufficiently large. These ideas are developed in our concluding remarks.

2.3 Mathematical formalism and normal modes

2.3.1 Governing equations

The three-dimensional equations governing the motion of rotating stratified flow under the Boussinesq approximation are

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u} \cdot \nabla \boldsymbol{u} + f\hat{\boldsymbol{z}} \times \boldsymbol{u} = -\nabla p + b\hat{\boldsymbol{z}} + \boldsymbol{D_u}, \tag{2.2a}$$

$$\nabla \cdot \boldsymbol{u} = 0, \tag{2.2b}$$

$$\frac{\partial b}{\partial t} + \boldsymbol{u} \cdot \nabla b = -N^2 w + D_b, \tag{2.2c}$$

where $\boldsymbol{u} = (\boldsymbol{u}_h, w) = (u, v, w)$ is the velocity field; b is the buoyancy perturbation (which can be defined based on potential temperature as $-g\theta'/\theta_0$, or based on density as $-g\rho'/\rho_0$); p is the pressure perturbation divided by a constant reference density, ρ_0. The operator D_q represents the dissipation of quantity q. We also assume Coriolis parameter, f, and Brunt-Väisälä frequency, N, to be constants.

2.3.2 Normal mode decomposition

To close the problem mathematically one needs a set of boundary and initial conditions. We will discuss the choice of the initial conditions in the following sections. As for boundaries, we assume periodicity. This configuration maintains statistical homogeneity and provides for efficient direct numerical simulation (DNS) using pseudo-spectral methods.

Following Leith (1980) and Bartello (1995), we use the eigenvectors of the linearised equation as our orthonormal basis. To derive these we first take the Fourier transform of (1.5), then linearise the system around a state of rest. For each

wavenumber $\boldsymbol{k} = (k_x, k_y, k_z)$, one finds the following eigenfrequencies (in dimensional form)

$$\lambda_{\boldsymbol{k}}^{(0)} = 0, \quad \lambda_{\boldsymbol{k}}^{\pm} = \pm\sigma_{\boldsymbol{k}} = \frac{(f^2 {k_z}^2 + N^2 {k_h}^2)^{1/2}}{k}, \tag{2.3}$$

where $k_h = (k_x^2 + k_y^2)^{1/2}$ and $k = (k_h^2 + k_z^2)^{1/2}$. We denote the associated orthonormal eigenvectors by $\boldsymbol{\xi}_{\boldsymbol{k}}^{(0)}$ and $\boldsymbol{\xi}_{\boldsymbol{k}}^{(\pm)}$. We can group all physical variables (e.g., velocity components and buoyancy) in a dynamical state vector $X_{\boldsymbol{k}}$. Then, the normal modes can be derived by projecting $X_{\boldsymbol{k}}$ onto the set of eigenvectors

$$G_{\boldsymbol{k}} = X_{\boldsymbol{k}} \cdot \overline{\xi_{\boldsymbol{k}}^{(0)}}, \quad A_{\boldsymbol{k}} = X_{\boldsymbol{k}} \cdot \overline{\xi_{\boldsymbol{k}}^{(\pm)}}. \tag{2.4}$$

where over-bar denotes the complex conjugate. We refer to the slow modes $G_{\boldsymbol{k}}$ as *geostrophic* (also known as rotational or vortical modes), and the fast ones $A_{\boldsymbol{k}}$ as *ageostrophic* (also known as gravitational or internal wave modes).

By applying the projections in (2.4) to the Fourier-space form of the linear terms of (1.5) and then non-dimensionalising it, the evolution equations of the normal-mode amplitudes follow as

$$\frac{\partial G_{\boldsymbol{k}}}{\partial t} = Ro\, \Phi_{\boldsymbol{k}}(G, A) + D_G \tag{2.5a}$$

$$\frac{\partial A_{\boldsymbol{k}}}{\partial t} + i\sigma_{\boldsymbol{k}} A_{\boldsymbol{k}} = Ro\, \Psi_{\boldsymbol{k}}(G, A) + D_A. \tag{2.5b}$$

In (2.5), we also grouped all quadratic nonlinearities in terms of $\Phi_{\boldsymbol{k}}(G,A)$ and $\Psi_{\boldsymbol{k}}(G,A)$ for notational economy. They can be expanded as convolution sums. For instance

$$\Phi_{\boldsymbol{k}}(G,A) = \sum_{\boldsymbol{k}=\boldsymbol{p}+\boldsymbol{q}} \left[\phi_{GG} G_{\boldsymbol{p}} G_{\boldsymbol{q}} + \phi_{AG} A_{\boldsymbol{p}} G_{\boldsymbol{q}} + \phi_{AA} A_{\boldsymbol{p}} A_{\boldsymbol{q}}\right], \tag{2.6}$$

where the coefficients, ϕ, are functions of $\boldsymbol{k}$, $\boldsymbol{p}$, $\boldsymbol{q}$ and constant parameters defined in (2.1). A similar description holds for $\Psi_{\boldsymbol{k}}(G,A)$.

Multiplying (2.5) by $\overline{G_{\boldsymbol{k}}}$ and $\overline{A_{\boldsymbol{k}}}$, the evolution of geostrophic and ageostrophic modal energy follows

$$\frac{\partial G_{\boldsymbol{k}}\overline{G_{\boldsymbol{k}}}}{\partial t} = Ro\,\Phi_{\boldsymbol{k}}\overline{G_{\boldsymbol{k}}} + D_G\overline{G_{\boldsymbol{k}}} + c.c. \tag{2.7a}$$

$$\frac{\partial A_{\boldsymbol{k}}\overline{A_{\boldsymbol{k}}}}{\partial t} = Ro\,\Psi_{\boldsymbol{k}}\overline{A_{\boldsymbol{k}}} + D_A\overline{A_{\boldsymbol{k}}} + c.c. \tag{2.7b}$$

where c.c. denotes the complex conjugate.

The nonlinear transfers in (2.7) can be summed as below to derive the horizontal energy fluxes

$$\Pi_G(k_h) = -\sum_{|\boldsymbol{k'}-\boldsymbol{k'}\cdot\hat{\boldsymbol{z}}|<k_h} Ro\,\Phi_{\boldsymbol{k'}}.\overline{G_{\boldsymbol{k'}}} + c.c. \tag{2.8a}$$

$$\Pi_A(k_h) = -\sum_{|\boldsymbol{k'}-\boldsymbol{k'}\cdot\hat{\boldsymbol{z}}|<k_h} Ro\,\Psi_{\boldsymbol{k'}}.\overline{A_{\boldsymbol{k'}}} + c.c.. \tag{2.8b}$$

2.4 Balance dynamics and nonlinear normal mode initialization

2.4.1 Time scales

From (2.3) there are two time scales in the flow when $Ro \to 0$. The time scale of $G_{\boldsymbol{k}}$ is L/U, which is determined by the nonlinear terms since $\lambda_{\boldsymbol{k}}^{(0)} = 0$. The time scale of $A_{\boldsymbol{k}}$, on the other hand, is derived as the inverse of the linear ageostrophic mode frequencies, $1/\sigma_{\boldsymbol{k}}$. Therefore the ratio of the two is

$$\frac{t^*}{T} = \epsilon = \frac{1/\sigma_{\boldsymbol{k}}}{L/U} \leq Ro, \tag{2.9}$$

where t^* and T denote the time scale of $A_{\boldsymbol{k}}$ and $G_{\boldsymbol{k}}$, respectively.

This separation enables us to split the time derivative into two parts

$$\frac{\partial()}{\partial t} = \frac{\partial()}{\partial t^*} + \epsilon \frac{\partial()}{\partial T} \tag{2.10}$$

This two-time-scale method is a foundation of a Nonlinear Normal Mode Initialization (NNMI) scheme proposed by Baer & Tribbia (1977). More detail on it can be found in classic texts on perturbation techniques such as Kevorkian & Cole (2013).

Rewriting equation (2.5) in terms of fast and slow time derivatives leads to

$$\epsilon \frac{\partial G_{\boldsymbol{k}}}{\partial T} = \epsilon \Phi_{\boldsymbol{k}}(G, A) + D_G \tag{2.11a}$$

$$\frac{\partial A_{\boldsymbol{k}}}{\partial t^*} + \epsilon \frac{\partial A_{\boldsymbol{k}}}{\partial T} + i\sigma_{\boldsymbol{k}} A_{\boldsymbol{k}} = \epsilon \Psi_{\boldsymbol{k}}(G, A) + D_A. \tag{2.11b}$$

The fast time derivative of the geostrophic modes are zero by construction, as they are slowly-varying.

A more advanced time-scale separation method was used by Reznik *et al.* (2001) and Zeitlin *et al.* (2003) to develop the theory of nonlinear geostrophic adjustment.

In these studies, instead of two time scales, the method of multiple time scales was used. More specifically, in addition to a fast time scale t^*, they used several slow time scales such that each one is $O(Ro)$ slower than the next. As will be seen in the following sections, their analytical results are compatible with the numerical simulations of the current study.

2.4.2 The Baer-Tribbia scheme

The essence of the Baer-Tribbia scheme is to ensure evolution on the slow time scale by forcing $\partial A_{\boldsymbol{k}}/\partial t^*|_{t=0} = 0$ in (2.11b). This is done by expanding (2.11) in terms of ϵ, and then applying perturbation techniques up to the desired order. It is useful to introduce a new notation for variable expansion in powers of both T and ϵ as

$$X = \left(X^{0,0} + X^{0,1}T + X^{0,2}T^2\right) + \left(X^{1,0} + X^{1,1}T + X^{1,2}T^2\right)\epsilon + O(\epsilon^2, T^3), \qquad (2.12)$$

where X can be any variable in (2.11). In other words, $X^{i,j}$, refers to terms proportional to ϵ^i and T^j.

After setting $\partial A_{\boldsymbol{k}}/\partial t^*|_{t=0} = 0$, expanding both sides of (2.11) and equating each order in ϵ, one obtains

$$A_{\boldsymbol{k}}^{0,0} = O(T) \qquad (2.13a)$$

$$i\sigma_{\boldsymbol{k}} A_{\boldsymbol{k}}^{1,0} + i\sigma_{\boldsymbol{k}} A_{\boldsymbol{k}}^{1,1} T = \Psi_{\boldsymbol{k}}^{0,0} + \Psi_{\boldsymbol{k}}^{0,1} T + O(T^2) \qquad (2.13b)$$

$$A_{\boldsymbol{k}}^{1,1} + i\sigma_{\boldsymbol{k}} A_{\boldsymbol{k}}^{2,0} = \Psi_{\boldsymbol{k}}^{1,0} + O(T) \qquad (2.13c)$$

The time derivative of (2.13b) yields one more equation

$$i\sigma_{\boldsymbol{k}} A_{\boldsymbol{k}}^{1,1} = \Psi_{\boldsymbol{k}}^{0,1} + O(T) \tag{2.14}$$

By setting $T = 0$ in (2.13) and (2.14), the initial ageostrophic modes are derived up to $O(\epsilon^2)$

$$A_{\boldsymbol{k}}^{0,0} = 0 \tag{2.15a}$$

$$A_{\boldsymbol{k}}^{1,0} = \frac{\Psi_{\boldsymbol{k}}^{0,0}}{i\sigma_{\boldsymbol{k}}} \tag{2.15b}$$

$$A_{\boldsymbol{k}}^{2,0} = \frac{\Psi_{\boldsymbol{k}}^{1,0}}{i\sigma_{\boldsymbol{k}}} - \frac{A_{\boldsymbol{k}}^{1,1}}{i\sigma_{\boldsymbol{k}}} = \frac{\Psi_{\boldsymbol{k}}^{1,0}}{i\sigma_{\boldsymbol{k}}} + \frac{\Psi_{\boldsymbol{k}}^{0,1}}{{\sigma_{\boldsymbol{k}}}^2}. \tag{2.15c}$$

Since our focus is on higher-order balance, it is necessary to keep terms up to $O(\epsilon^2)$ as a minimum requirement since QG theory is only one order lower. Equation (2.15a) shows that the ageostrophic modes are $O(Ro)$. Therefore, the scaling of ageostrophic energy goes as $O(Ro^2)$. With some algebra, the nonlinear terms in (2.15) can be expressed in terms of the geostrophic modes as

$$\begin{aligned}
\Psi_{\boldsymbol{k}}^{0,0} &= \sum_{\boldsymbol{k}=\boldsymbol{p}+\boldsymbol{q}} \psi_{GG} G_{\boldsymbol{p}} G_{\boldsymbol{q}} \\
\Psi_{\boldsymbol{k}}^{1,0} &= \sum_{\boldsymbol{k}=\boldsymbol{p}+\boldsymbol{q}} \frac{\psi_{GA}}{i\sigma_{\boldsymbol{q}}} G_{\boldsymbol{p}} \left(\sum_{\boldsymbol{q}=\boldsymbol{m}+\boldsymbol{n}} \psi_{GG} G_{\boldsymbol{m}} G_{\boldsymbol{n}} \right) \\
\Psi_{\boldsymbol{k}}^{0,1} &= 2 \sum_{\boldsymbol{k}=\boldsymbol{p}+\boldsymbol{q}} \psi_{GG} G_{\boldsymbol{p}} \left(\sum_{\boldsymbol{q}=\boldsymbol{m}+\boldsymbol{n}} \phi_{GG} G_{\boldsymbol{m}} G_{\boldsymbol{n}} \right)
\end{aligned} \tag{2.16}$$

Note that we only expanded the fast modes in terms of ϵ and kept geostrophic modes unchanged, as justified by Warn *et al.* (1995). The direct calculation of convolution

sums in (2.16), especially nested ones, is too costly. To circumvent this problem we propose a novel algorithm that uses the pseudo-spectral method to calculate them. It is briefly described in Appendix B. The dissipation terms can be kept in the procedure above. We performed initialization with and without viscosity and diffusion. The results were very similar other than in the very small scales. This is to be expected since balance is most apparent at larger scales where viscosity and diffusion are rather unimportant.

Knowing that the geostrophic spectrum admits a slope of -3, one might try to analytically derive the slope of the balanced ageostrophic energy spectrum at least to first order. Thus, $A^{1,0}_{\boldsymbol{k}}.\overline{A^{1,0}_{\boldsymbol{k}}}$ (over-bar denoting the complex conjugate) should be calculated from (3.7b) and (3.8)

$$\begin{aligned} A^{1,0}_{\boldsymbol{k}}.\overline{A^{1,0}_{\boldsymbol{k}}} &= \frac{1}{\sigma^2_{\boldsymbol{k}}}\Psi^{0,0}_{\boldsymbol{k}}.\overline{\Psi^{0,0}_{\boldsymbol{k}}} = \frac{1}{\sigma^2_{\boldsymbol{k}}}\sum_{\boldsymbol{k}=\boldsymbol{p}+\boldsymbol{q}}\psi_{GG}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q})G_{\boldsymbol{p}}G_{\boldsymbol{q}}\sum_{\boldsymbol{k}=\boldsymbol{m}+\boldsymbol{n}}\overline{\psi_{GG}(\boldsymbol{k},\boldsymbol{m},\boldsymbol{n})G_{\boldsymbol{m}}G_{\boldsymbol{n}}} \\ &= \sum_{\boldsymbol{k}=\boldsymbol{p}+\boldsymbol{q}}\sum_{\boldsymbol{k}=\boldsymbol{m}+\boldsymbol{n}}\frac{1}{\sigma^2_{\boldsymbol{k}}}\psi_{GG}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q})\overline{\psi_{GG}(\boldsymbol{k},\boldsymbol{m},\boldsymbol{n})}G_{\boldsymbol{p}}G_{\boldsymbol{q}}\overline{G_{\boldsymbol{m}}G_{\boldsymbol{n}}} \end{aligned} \tag{2.17}$$

Knowing only that the spectrum formed by geostrophic modes scales as k^{-3} does not lead to a solution for the inertial range of the ageostrophic energy spectrum unless one calculates or approximates the coefficient $\psi_{GG}(\boldsymbol{k},\boldsymbol{p},\boldsymbol{q})\overline{\psi_{GG}(\boldsymbol{k},\boldsymbol{m},\boldsymbol{n})}/\sigma^2_{\boldsymbol{k}}$, which is a function of wavenumber, and then performs the summation. We therefore explore the slope of ageostrophic energy spectrum numerically and discuss it in section (2.5.2).

2.5 Balancing procedure and numerical configuration

In this section, we develop a procedure to investigate the effect of initial Baer-Tribbia balance as well as the role of rotation and stratification in the transition from balance dynamics to more general smaller-scale turbulence. It is carried out in three steps: first, generation of the geostrophic data, second, finding balanced ageostrophic modes using NNMI and finally exploring the robustness of balanced initial conditions in the more general dynamical context of the nonhydrostatic Boussinesq simulations. Before describing each in more detail, we lay out the numerical configuration of our simulations.

Our Boussinesq model employs the de-aliased pseudo-spectral method with second-order time stepping. The state variables that are integrated in time are 3D vorticity and buoyancy. We set the size of our domain to be 2π in the horizontal and $2\pi \times (f/N)$ in the vertical. In this way the aspect ratio, $\alpha = H/L$, is equal to f/N, which makes the domain a cube of $(2\pi)^3$ using Charney (1971) scaling. Throughout this paper we match stratification to rotation by fixing $N/f = L/H$. Therefore, the change in Ro is equivalent to a change in Fr, following classical QG scaling. For the sake of brevity we usually refer only to changes in rotation (or Ro), but the reader should keep in mind this implies changes to stratification as well.

In the real atmosphere the ratio of N to f is usually reported as 100, and in the ocean as $30 - 50$. We also know that at least $O(10^2)$ grid points are required to resolve something more than just viscous and diffusional coupling of vertical layers. This vertical grid, together with high L/H $(= N/f)$, would require a large horizontal resolution, as explained in Bartello (2010). To maintain the vertical resolution high

enough while keeping horizontal resolution within our computational resources, we reduced the ratio N/f to 8. This compromise distances us from exact atmospheric and oceanic scaling somewhat, but we have varied N/f and are certain that the features described below are generic, while recognising the future need to explore this question with larger computers.

For the physical domain of $[0, 2\pi]^2 \times [0, 2\pi f/N]$, two different grids were employed:

- Grid 1) $\Delta x = \Delta y = (N/f)\Delta z$. This grid is more loyal to QG structures since it is isotropic in Charney stretched coordinates.
- Grid 2) $\Delta x = \Delta y = \Delta z$. This grid is isotropic in unstretched physical space. Hence, it is felt to be a better choice to capture the smaller scales where the dynamics, and hence the aspect ratio, are still unknown.

If N_h and N_z are the number of horizontal and vertical grid points, respectively, grid 2) requires $N_z = (f/N)N_h$ vertical points, whereas the Charney grid requires $N_z = N_h$, which is far larger. Grid 2) therefore allows for a wider range of horizontal scales, for given computational resources in addition to providing unbiased numerics for dynamics in the small-scale transition range.

We performed most of our simulations using grid 2) as we believe it is a better choice in analysing the breakdown of balance. Nevertheless, we performed several simulations using grid 1) and compared them. Table 1 shows a list of all simulation parameters used in this study.

The other parameters that need to be set in the Boussinesq model are the viscosity and diffusion. In addition to the Laplacian operator of the Boussinesq

set, an iterated Laplacian (hyperviscosity) was often used. To be consistent with our cylindrical truncation of Fourier modes, we employed a cylindrical dissipation operator expressed as

$$D_{\boldsymbol{u}} = D_b = \nu_h(-1)^{n+1}\nabla_h^{2n} + \nu_z(-1)^{n+1}\frac{\partial^{2n}}{\partial z^{2n}}, \tag{2.18}$$

which provides the same dissipation for buoyancy and velocity. In all our simulations on grid 2) we keep $\nu_h = \nu_z = \nu$. The Newtonian Laplacian operator is recovered setting $n = 1$. Hyperviscosity, here set to $n = 4$, restricts dissipation to a narrower range of scales. For further details on the comparison between the two in rotating stratified turbulence see Bartello *et al.* (1996). Although the majority of our simulations used hyperviscosity, we also performed a number of sensitivity DNS runs in Figure 2–14 below. Although there are quantitative differences, the observed trends are the same. All the times reported were normalized with the rms geostrophic vertical vorticity, τ, which is measured at the end of the preliminary QG runs and can be found in Table 2–1.

2.5.1 Generation of geostrophic data

In this step we run a decaying QG model to produce the geostrophic modes that are later used in our balancing scheme. Our QG model is obtained from the Boussinesq model by setting the ageostrophic modes to zero at each time step. Although, this is not efficient, it suffices for the generation of our initial data. As shown in Figure 2–1, our initial energy distribution peaks at $k_i = 20$ to provide a wavenumber range for the considerable upscale transfer of energy. Then the model is run until

Run	N_h	N_z	Ro_u	Ro_ω	Fr	τ	$D_{\boldsymbol{u}}$	ν_h	ν_z
QGv18	1536	192	—	—	—	4.35	∇^8	5×10^{-18}	5×10^{-18}
QGv19	1536	192	—	—	—	5.28	∇^8	5×10^{-19}	5×10^{-19}
QGv20	1536	192	—	—	—	6.45	∇^8	5×10^{-20}	5×10^{-20}
QGg1	1024	1024	—	—	—	12.73	∇^8	5×10^{-18}	1.6×10^{-23}
QGg2	1024	128	—	—	—	6.23	∇^8	1×10^{-18}	1×10^{-18}
B768r005	768	96	0.005	0.044	0.015	4.8	∇^8	1×10^{-17}	1×10^{-17}
B768r01	768	96	0.01	0.089	0.030	4.8	∇^8	1×10^{-17}	1×10^{-17}
B768r03	768	96	0.03	0.269	0.091	4.8	∇^8	1×10^{-17}	1×10^{-17}
B768r05	768	96	0.05	0.453	0.160	4.8	∇^8	1×10^{-17}	1×10^{-17}
B768r07	768	96	0.07	0.646	0.244	4.8	∇^8	1×10^{-17}	1×10^{-17}
B768r09	768	96	0.09	0.852	0.351	4.8	∇^8	1×10^{-17}	1×10^{-17}
B768r11	768	96	0.11	1.078	0.491	4.8	∇^8	1×10^{-17}	1×10^{-17}
B1536r01lv	1536	192	0.01	0.105	0.036	6.4	∇^8	5×10^{-20}	5×10^{-20}
B1536r02lv	1536	192	0.02	0.211	0.075	6.4	∇^8	5×10^{-20}	5×10^{-20}
B1536r03lv	1536	192	0.03	0.320	0.118	6.4	∇^8	5×10^{-20}	5×10^{-20}
B1536r07lv	1536	192	0.07	0.817	0.403	6.4	∇^8	5×10^{-20}	5×10^{-20}
B1536r01	1536	192	0.01	0.104	0.036	5.3	∇^8	5×10^{-19}	5×10^{-19}
B1536r03	1536	192	0.03	0.314	0.112	5.3	∇^8	5×10^{-19}	5×10^{-19}
B1536r05	1536	192	0.05	0.553	0.204	5.3	∇^8	5×10^{-19}	5×10^{-19}
B1536r07	1536	192	0.07	0.771	0.330	5.3	∇^8	5×10^{-19}	5×10^{-19}
B1536r09	1536	192	0.09	1.041	0.511	5.3	∇^8	5×10^{-19}	5×10^{-19}
B2048r01	2048	256	0.01	0.112	0.04	6.4	∇^8	5×10^{-21}	5×10^{-21}
B2048r03	2048	256	0.03	0.343	0.131	6.4	∇^8	5×10^{-21}	5×10^{-21}
B2048r05	2048	256	0.05	0.596	0.262	6.4	∇^8	5×10^{-21}	5×10^{-21}
B2048r07	2048	256	0.07	0.896	0.474	6.4	∇^8	5×10^{-21}	5×10^{-21}
B2048r09	2048	256	0.09	1.278	0.809	6.4	∇^8	5×10^{-21}	5×10^{-21}
B2048r01n	2048	256	0.01	0.084	0.024	3.0	∇^2	5×10^{-5}	5×10^{-5}
B2048r05n	2048	256	0.05	0.425	0.124	3.0	∇^2	5×10^{-5}	5×10^{-5}
B2048r09n	2048	256	0.09	0.781	0.248	3.0	∇^2	5×10^{-5}	5×10^{-5}
B2048r11n	2048	256	0.11	0.970	0.327	3.0	∇^2	5×10^{-5}	5×10^{-5}

Table 2–1: The runs starting with QG are based on the quasigeostrophic equations and the rest are based on the non-hydrostatic Boussinesq equations. The values of Ro_u, Ro_ω, Fr and τ in the Boussinesq runs are calculated at $t = 0$.

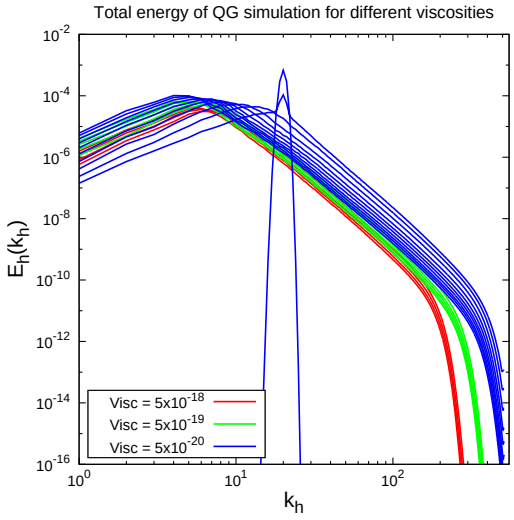


Figure 2–1: The spectra of total energy for QG simulations with different viscosities: runs listed as QGv18, QGv19 and QGv20 in table 2–1. For QGv20, the spectra is plotted from $t = 0$ to 5τ. For the other two simulations only $t \in [3\tau, 5\tau]$ is plotted.

$t = 5\tau$, based on the initial rms vorticity. As time evolves, the spectrum fills out quickly. Subsequent changes to its shape occur much more slowly.

According to Charney (1971), QG flow is mathematically analogous to 2D isotropic turbulence, where the logarithmic slope of the energy spectrum in the potential-enstrophy cascade range is predicted to be -3. Our slope, not surprisingly, is steeper than this as in previous studies at comparable resolutions, arguably due to emerging coherent structures (cf. McWilliams *et al.* (1994)). To avoid any rapid small-scale adjustment when inputting these QG modes into the non-hydrostatic Boussinesq model, we decided to perform separate preliminary QG simulations for all values of (hyper)viscosity used below.

As explained in Section 2.5, two types of grid were employed; grid 1) is isotropic in the Charney stretched coordinate and grid 2) is isotropic in unstretched real

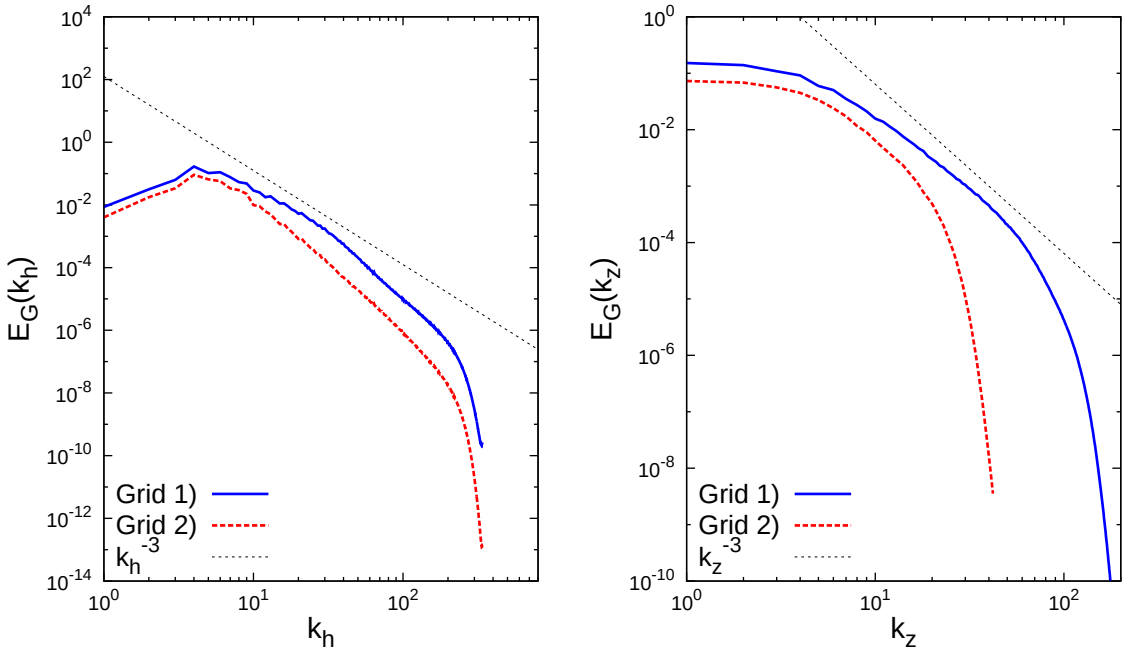


Figure 2–2: Horizontal (left) and vertical (right) spectra of total energy for QG simulations for the two different grids.

coordinates. Figure 2–2 shows the horizontal and vertical spectra of the total energy using grid 1) with 1024^3 collocation points (Run QGg1 in table 2–1) and grid 2) at $1024^2 \times 128$ (Run QGg2). The latter had the same hyperviscosity in all directions, whereas in the former the vertical viscosity was reduced by a factor $(f/N)^8$ to keep νk_{max}^8 the same in each direction, since we used a ∇^8 operator.

Figure 2–2 shows the horizontal and vertical energy spectra at $t = 4\tau$ using $1024^2 \times 128$ points on grid 2) and at 1024^3 on grid 1). The slope of the horizontal spectrum is similar at large horizontal wavenumbers. Not surprisingly the spectrum corresponding to grid 2) is lower, since the vertical dissipation is higher. The vertical spectra of grid 1) extends to much higher vertical wavenumbers. Nonetheless, they are reasonably parallel at low k_z. Acknowledging that grid 1) resolves QG structures isotropic in Charney stretched coordinates better, figure 2–2 shows that the results of grid 2) are reliable as well for the QG flow, while at the same time not biasing the small-scale dynamics emerging from the shallow spectra discussed below.

2.5.2 Generation of balanced ageostrophic modes using NNMI

Once geostrophic data are generated, they can be inserted into the balance scheme to derive the ageostrophic modes that ensure initial evolution on the slow time scale. Our high-order scheme and its numerical implementation are described in §2.4 and the Appendix, respectively. The initialized data depend on Ro (and consequently Fr), since the frequency, $\sigma_{\boldsymbol{k}}$, in equation (2.3) depends on f and N. There are different ways to calculate Ro in fully-developed turbulence. Following its definition in (2.1), it can either be derived based on mean-square vorticity or velocity.

The vorticity-based Rossby number is formulated as

$$Ro_\omega = \frac{\langle (\nabla \times \boldsymbol{u}(x,y,z) \cdot \hat{\boldsymbol{z}})^2 \rangle^{1/2}}{f}, \tag{2.19}$$

whereas the velocity-based one is given by

$$Ro_u = \frac{\langle \boldsymbol{u}.\boldsymbol{u} \rangle^{1/2}}{f \,.\, L} \tag{2.20}$$

In both definitions, $\langle\ \rangle$ represents a spatial average. In (2.20), L is the characteristic length scale, plausibly calculated by inverting the wavenumber at which the energy spectrum peaks.

The vorticity-based Rossby number is more affected by the dynamics at small scales than its velocity-based counterpart. Since we want to classify different rotating regimes based on their large scale QG-like motion, we favour the velocity-based definition and drop the subscript. As shown below, some of our simulations show a transition to a more shallow spectrum at larger wavenumbers. While this affects the root mean square vorticity and hence Ro_ω, its influence on Ro_u is much weaker. In addition, the slope of this shallow range shows some dependence on model dissipation, which further justifies our choice of velocity-based Rossby number. Note that even in the QG model, the vorticity-based Rossby number is approximately one order of magnitude larger than Ro_u.

A third definition of Rossby number is based on the extremal value of vertical vorticity in real space, $Ro_{max} = |\nabla \times \boldsymbol{u}(x,y,z) \cdot \hat{\boldsymbol{z}}|_{max}/f$. This definition has been used frequently by Dritschel and co-workers (see for instance Dritschel & Viudez (2003), Viudez & Dritschel (2004) and Dritschel & McKiver (2015)). This maximum

Ro_u	Ro_ω	Ro_{max}
0.01	0.10	0.81
0.03	0.31	2.016
0.09	1.04	16.49

Table 2–2: Different Rossby numbers at $t = 0$ on a grid of $1536^2 \times 192$

Ro is far greater than Ro_ω since the vertical vorticity at large Reynolds numbers displays a close to exponential distribution. We derived Ro_{max}, Ro_ω and Ro_u after initialization and compared them in Table 2–2. This considerable difference should be taken into account when values are compared to Ro_{max} and Ro_ω employed in other studies. It may be argued that values of Ro_{max} in excess of unity are unrealistically large. However, Hakim (2000) and Hoskins & Hodges (2002) have shown this not to be the case. In fact, the former study shows relative vorticity values in excess of f in 40% of the 500 hPa weather charts considered.

To investigate the effect of dissipation and rotation separately, we keep Ro constant for different viscosities. In doing so a slight change of the Coriolis parameter (and consequently a change of N) is necessary, as the peak of the energy spectra, L, and the total energy, U^2, in the initial data resulting from the QG run are slightly different for different viscosities.

To describe the initialization we investigate the ageostrophic modal spectrum that the Baer-Tribbia scheme produces. We begin by taking the slow modes from an output of a QG simulation. While keeping their complex Fourier phases constant, we scale their amplitudes to yield geostrophic modal spectra with various slopes. After implementing the initialization scheme on each set of these data, we produce the spectrum of the balanced ageostrophic modes. It is, of course, also interesting to see

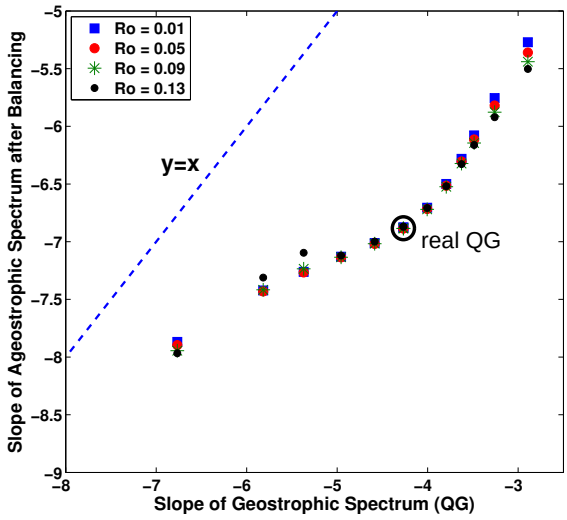


Figure 2–3: The logarithmic slope of the ageostrophic energy spectrum after the Baer-Tribbia balancing scheme as a function of the geostrophic slope. The resolution was $1536^2 \times 192$, $\nu = 5 \times 10^{-20}$

how the balanced output changes at different Rossby numbers. Figure 2–3 presents the slope of the Baer-Tribbia ageostrophic energy (E_A) spectrum versus the slope of the input geostrophic spectrum (E_G).

The real quaisgeostrophic data, derived from a decaying QG simulation, have a slope of -4.2, as seen in Figure 2–1 whose value is marked with a large circle. It is interesting that for the true QG data, all Rossby numbers have almost the same ageostrophic spectral slope, which is -6.8. As we shall see, in all our simulations at different Ro, viscosities and resolution, the slope of the balanced ageostrophic spectrum is between -6.5 and -7, that is, much steeper than the slope of the geostrophic spectrum. We see similar results at other geostrophic slopes as well since the entire curve falls below the $y = x$ line, but as we vary the geostrophic spectral slope

further from the circled value, the ageostrophic slope becomes a function of the Rossby number. Instead of using our QG simulation complex phases, we followed the same procedure with completely random phases and obtained qualitatively similar results. The conclusion is simply that balanced turbulence occurs in conjunction with rather steep energy spectra, as the total energy spectrum follows that of the less steep geostrophic modal spectrum at small scales. As such, balanced rotating stratified turbulence cannot explain the shallow -5/3 range in sub-deformation scale atmospheric and oceanic data.

2.5.3 Boussinesq simulations

We ran the non-hydrostatic Boussinesq model using the geostrophic modes from the QG model and the balanced ageostrophic data derived in the NNMI step as an initial condition in a more general dynamical setting. Our goal is to establish the robustness of this sort of balance within a nonhydrostatic Boussinesq framework. It is, of course, well known that this will depend on the Rossby number. We therefore calculated separate ageostrophic initial conditions for a variety of Ro. The following section presents the results of these simulations.

2.6 Results

Our aim is first to examine the effect of balance by comparing initially-balanced simulations with their unbalanced counterparts. The unbalanced initial data were generated by taking the output ageostrophic modes of the Baer-Tribbia data and scrambling their Fourier phases. In this way the comparison is fair since the balanced and unbalanced ageostrophic modes share the same amplitude. Hence, all simulations of §2.6.1-2.6.3 start with the same E_G, and for each Rossby number, initially balanced

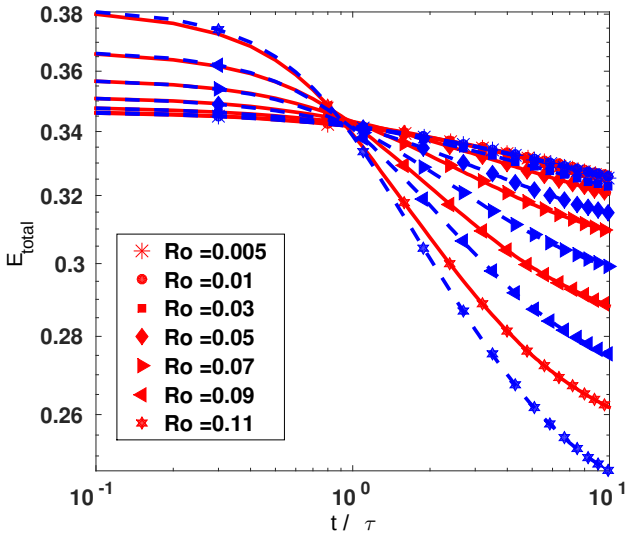


Figure 2–4: The time series of the total energy. The solid lines are the initially balanced flow, dashed lines are the initially phase-scrambled flow. The plots correspond to runs starting with B768 listed in table 2–1.

and unbalanced simulations have the same E_A and indeed the same spectrum. In fact, the energetics of the two simulations are identical. After generating a set of balanced and unbalanced data, we examine

1. the time series of their spatially-averaged quantities such as energy,
2. the evolution of their energy spectra in time,
3. the frequency spectrum of a flow variable at a grid point in physical space.

2.6.1 The time series of initially balanced and unbalanced flow

In Figure 2–4 the total energy, which is the sum of the geostrophic and ageostrophic energy, is depicted. At low Rossby number the energy stays relatively more constant in time, signalling that the dominant direction of transfer is toward larger scales. As Ro increases, more forward transfer takes place and more energy is dissipated.

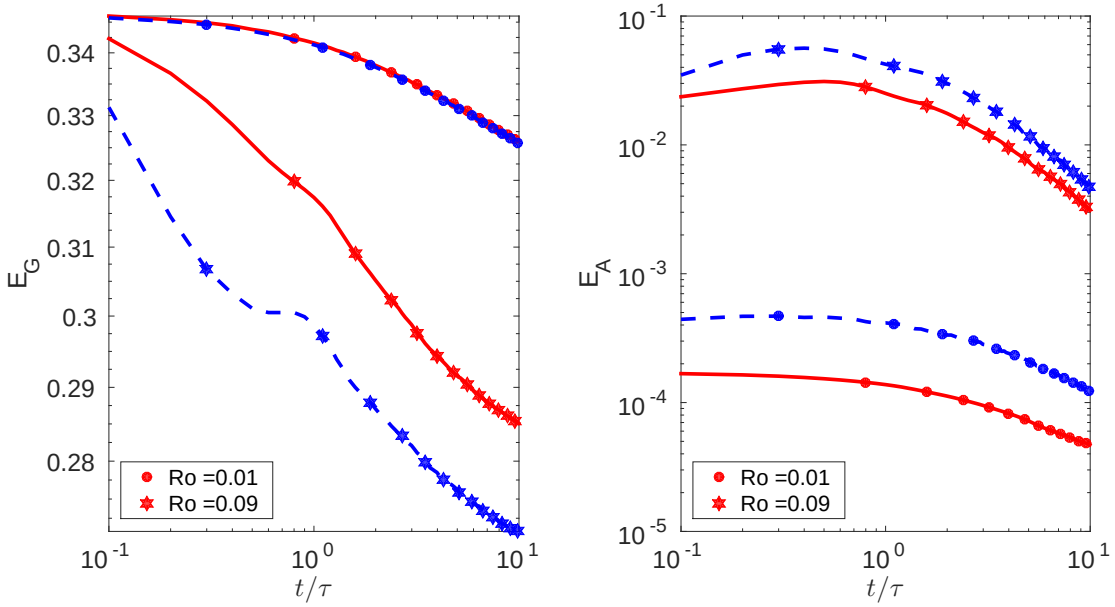


Figure 2–5: Left: Geostrophic energy vs. time. Right: Ageostrophic energy vs time. The solid lines are the time series of the initially balanced flow and the dashed lines are the initially phase-scrambled flow. The simulations correspond to B768r01 and B768r09.

Comparing the initially balanced and unbalanced flows shows that at strong rotations, both have similar levels of energy. However, as *Ro* increases, more energy is dissipated when initial conditions are unbalanced, implying that most of the forward cascade is due to the ageostrophic modes.

Figure 2–5 shows the total geostrophic and ageostrophic energy of the flow as a function of time. Noting the scale of the vertical axis of both panels, one can conclude that the decay of ageostrophic energy is much larger than that of geostrophic energy. This is consistent with an inverse cascade for the geostrophic modes and a forward

cascade for ageostrophic modes as described for initially unbalanced turbulence by Bartello (1995).

Comparing the two Rossby numbers in the left panel of Figure 2–5 one finds that E_G of the higher Rossby number decays faster than that of the lower Rossby number. Since the dominant transfer of geostrophic energy by geostrophic modes alone is toward larger scales, the faster geostrophic decay at higher Ro is presumably due to stronger interaction with ageostrophic modes. This is in accordance with the ageostrophic time series of these Rossby numbers in the right panel. The higher Rossby number has higher E_A, and it decays faster.

Both initially phase-scrambled and balanced flows have the same level of geostrophic and ageostrophic energy at $t = 0$. This is not visible on the logarithmic horizontal axis of Figure 2–5. At $Ro = 0.01$, the geostrophic energy of the initially balanced and unbalanced flow stay very close to each other during the entire integration time. The ageostrophic energy of the initially unbalanced flow, however, is larger at this Ro. This shows that the ageostrophic modes of the unbalanced flow take more energy from geostrophic modes and dissipate it at small scales. In spite of this, the total geostrophic energy is not much affected, since E_A is several orders smaller than E_G at $Ro = 0.01$. Unlike the low Rossby number, at the higher $Ro = 0.09$, geostrophic energies exhibit a clear difference. Compared to $Ro = 0.01$, E_A is several orders larger at $Ro = 0.09$, implying the interaction between geostrophic and ageostrophic modes is much stronger.

Figure 2–6 illustrates how balance dynamics breaks down with time in different rotating regimes. In this figure, the total ageostrophic energy is depicted versus the

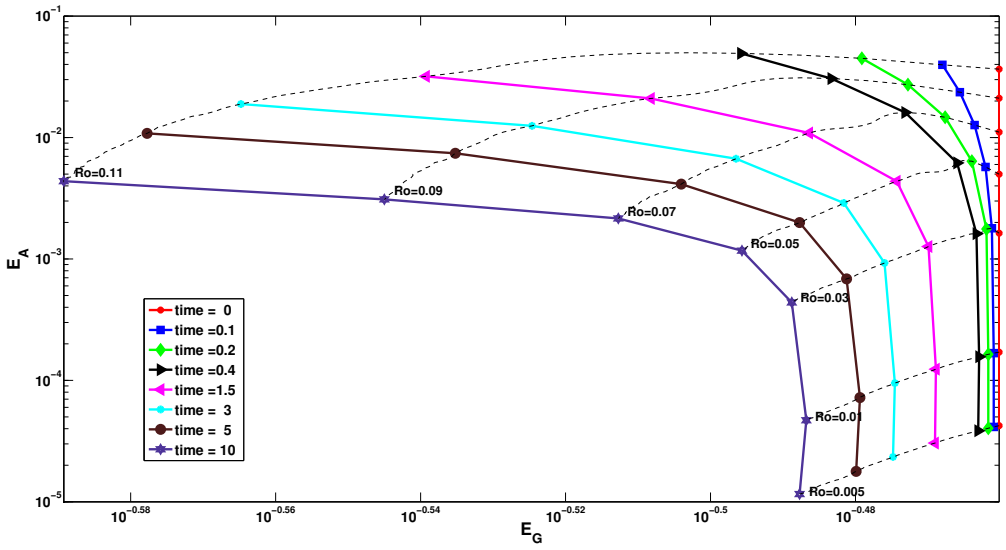


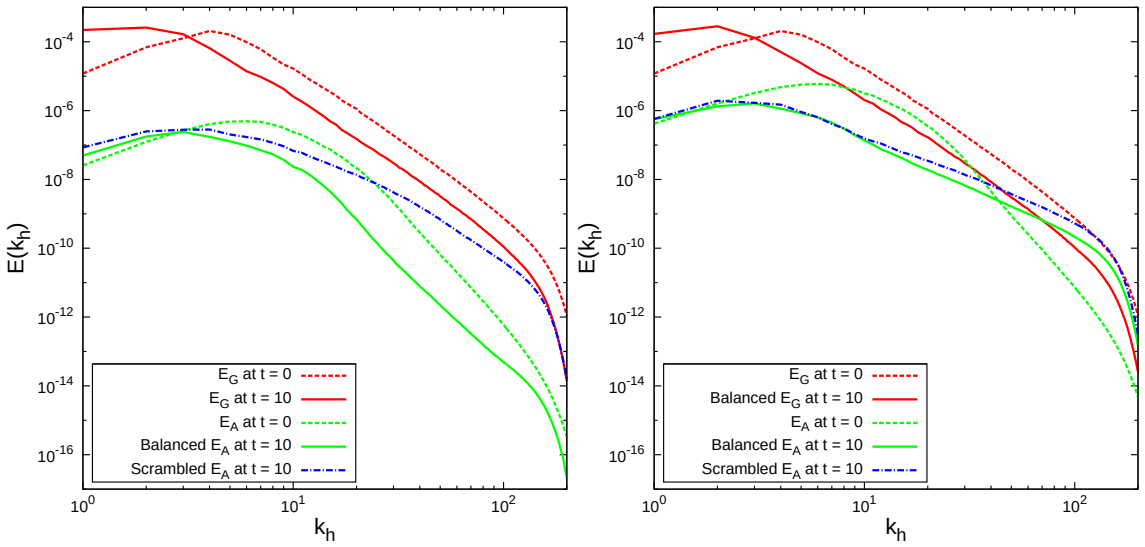
Figure 2–6: The total ageostrophic energy versus total geostrophic energy at different times for different Ro. Solid lines connect different Ro at the same time, while dashed lines do the reverse. The plots correspond to runs starting with B768 listed in table 2–1.

total geostrophic energy on logarithmic axes. Each solid curve with identical marker shapes connects different Ro at the time given in the legend. The dashed lines connect the same Rossby numbers at different times. At $t = 0$, the curve is a vertical line, as we started all our simulations with the same amount of geostrophic energy. However, for different Ro the initial ageostrophic energy produced by the Baer-Tribbia scheme is different. More specifically, equation (2.15a) shows that $A_{\boldsymbol{k}} \sim O(Ro)$, since $A_{\boldsymbol{k}}^{0,0} = 0$. Therefore, the ageostrophic energy produced by the Baer-Tribbia scheme scales as Ro^2.

As time increases, the dynamics fall into two distinct regimes. First, one which is balanced and remains as vertical lines. Since rotation is strong in this regime, the dominant transfer is an inverse cascade. Hence, the geostrophic energy stays similar for different Ro, but the balanced ageostrophic energy changes. The second regime occurs at weaker rotation and behaves as horizontal lines. The marked feature of this regime is that different Rossby numbers seem to approach the same level of ageostrophic energy, a characteristic we could not infer from other figures. The geostrophic energies at different Rossby numbers are, as expected, different as Ro is increased, since more of it is transferred to ageostrophic modes, where it is then dissipated via a forward cascade. The other notable feature of this figure is that the transition zone between these two dynamics gets sharper with time.

2.6.2 The energy cascade of initially-balanced and unbalanced flows

The energy cascades of fast and slow modes can be the simplest indicator of balance. In unforced simulations we speculate that the ageostrophic modal specrum is a power law whose amplitude decays with time (Bartello, 1995). Figure 2–7 shows



(a) $Ro = 0.03$ (b) $Ro = 0.09$

Figure 2–7: The upper dashed curve corresponds to the geostrophic spectrum at $t = 0$, and the lower dashed one corresponds to the ageostrophic spectrum at $t = 0$. The solid curves represent the spectrum of the initially balanced data at $t = 10\tau$, where the geostrophic spectrum is above the ageostrophic one. The dash-dotted line is the ageostrophic spectrum of initially phase-scrambled data. The corresponding simulations are B768r03 and B768r09.

the evolution of initially balanced and unbalanced states at $t = 10\tau$. As explained previously, the initially unbalanced data is generated by scrambling the phases of the ageostrophic modes, implying their initial spectra coincide.

The geostrophic spectrum of both balanced and unbalanced flow evolve similarly in time. However, the ageostrophic part of the energy spectrum is considerably different for initially balanced and unbalanced data. For $Ro = 0.03$ (Figure 2–7a), this difference both in terms of value and slope is drastic, whereas for $Ro = 0.09$ (Figure 2–7b) it is much less. In the lower Rossby number case, the slope of the ageostrophic balanced spectrum is much steeper than its unbalanced counterpart, showing that the initialization scheme prevented the excitation of wave modes. In the higher Ro simulation, the growth of balanced ageostrophic modes is more similar to unbalanced ones. This is expected as the Baer-Tribbia scheme uses perturbation techniques based on expansions in terms of Ro. As it increases convergence is less rapid, presumably degrading the quality of the balance.

2.6.3 The frequency spectrum of initially balanced and unbalanced flows

The role of the initial balancing is perhaps more directly observed in the frequencies of the flow variables. To derive these in the current statistically homogeneous geometry we simply take the Fourier transform with respect to time of a flow variable at a point in our physical-space collocation grid. The choice of variable does not affect the result as long as both geostrophic and ageostrophic modes contribute to it. Here, we arbitrarily chose the x component of velocity. Figure 2–8 shows its time series and frequency spectra for both initially-balanced and unbalanced conditions in the simulation denoted B768r03. High-frequency fluctuations make the

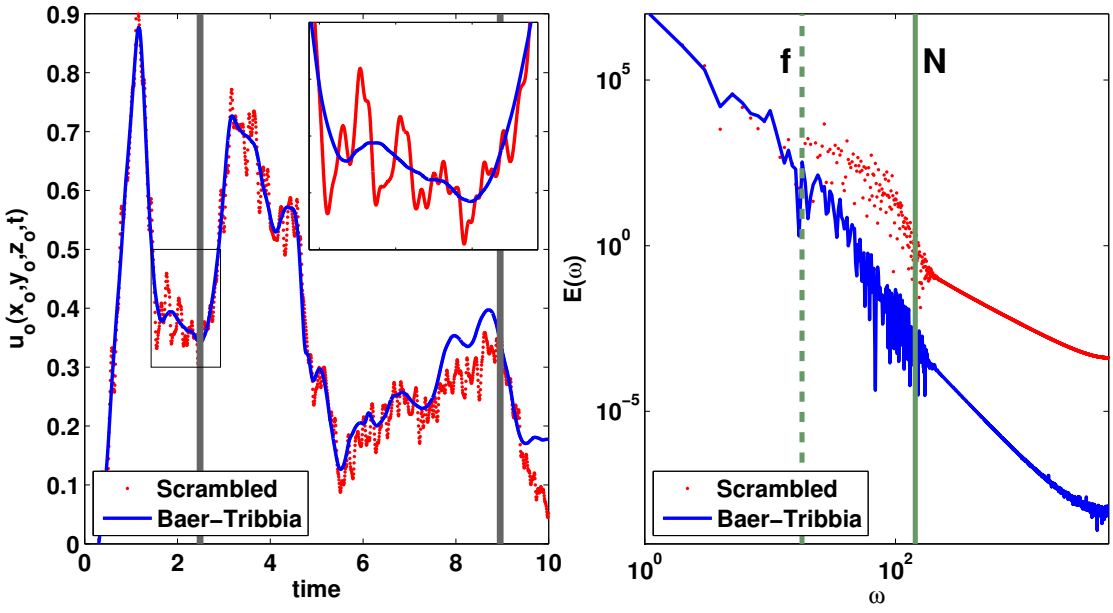


Figure 2–8: (left): The time series of $u(x_0, y_0, z_0, t) = u_0(t)$. The thicker smooth line shows the evolution of the balanced condition in time. The thinner fluctuating line shows the evolution of initially phase-scrambled data. (right): The Fourier transform of $u_0(t)$ with respect to time, for initially unbalanced and balanced conditions. The dashed vertical line represents the coriolis parameter, f, and the solid vertical line shows the Brunt-Väisälä frequency, N.

evolution of unbalanced initial data clearly different from their balanced counterpart. However, the slow part of the flow evolves similarly at early times. Later, even the slowly-varying part of the motion displays differences between initially balanced and unbalanced simulations.

To take the Fourier transform we chose a time interval that satisfies two conditions. First, the beginning is after $t = \tau$, discarding early non-stationary effects. Second, the function at its two ends had approximately the same value. The interval is marked by two thick vertical lines in the left panel of Figure 2–8. Its Fourier transform is portrayed on the right panel.

Equation (2.3) shows that the linear frequencies of the ageostrophic modes lie between Coriolis parameters, f, and Brunt-Väisälä frequency, N. These parameters are shown with vertical lines in the right panel of figure 2–8. Unsurprisingly, in this range the amplitude of balanced flow is less than that of the unbalanced one.

To understand the effect of initial balancing in different rotating regimes better, the frequency spectra at four different Ro's are shown in Figure 2–9. Similar to the those Figure 2–8, the most significant feature is the difference between the amplitudes of balanced and unbalanced flows in the band of inertia-gravity frequencies. Here we see a jump in amplitude of the unbalanced system at low Rossby numbers. The jump diminishes as the Rossby number increases. Note that the wave band moves to the left as Ro increases, since we decrease both N and f. However, the logarithmic width of the band does not change as we keep N/f constant. Not surprisingly, we see that reducing the frequency disparity between fast and slow modes reduces the difference

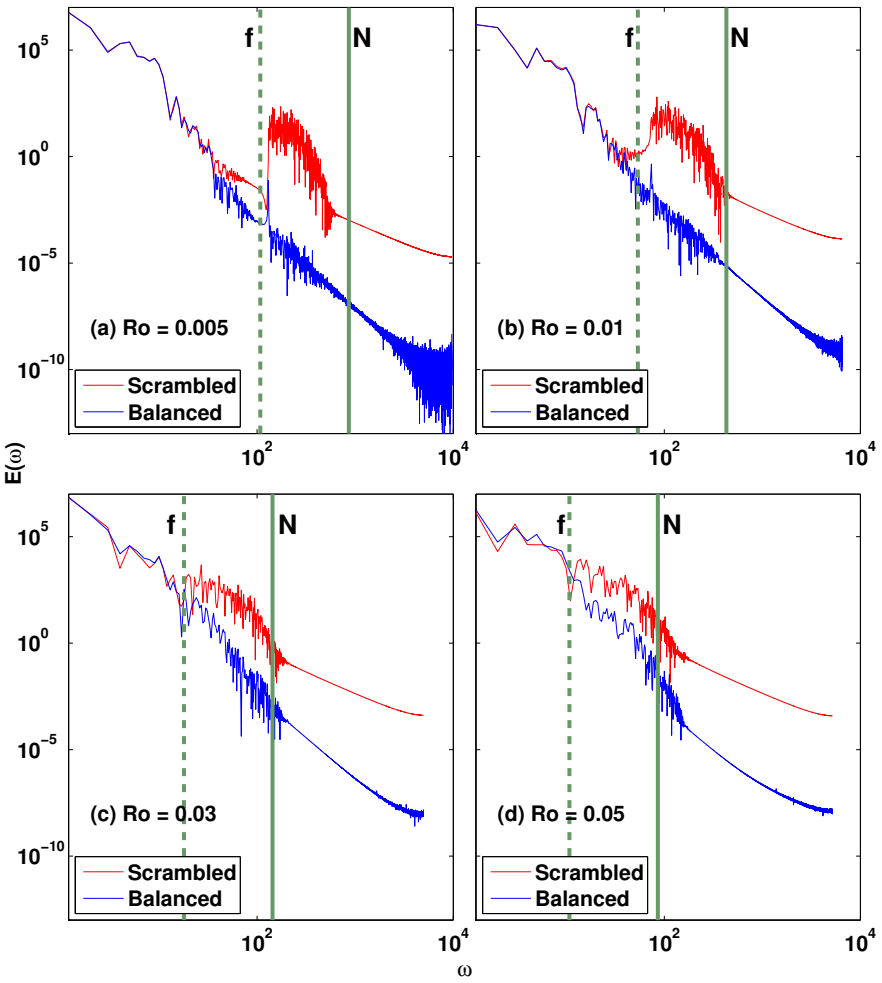


Figure 2–9: The time spectrum of $u(x_0, y_0, , z_0, t)$ for four different Rossby numbers in the runs of B768r005, B768r01, B768r03 and B768r05. The dashed vertical line represents the coriolis parameter, f, and the solid vertical line shows Brunt-Väisälä frequency, N. The time spectrum of phase-scrambled initial condition is above the balanced initial condition in Rossby numbers. All the panels have the same vertical and horizontal axes.

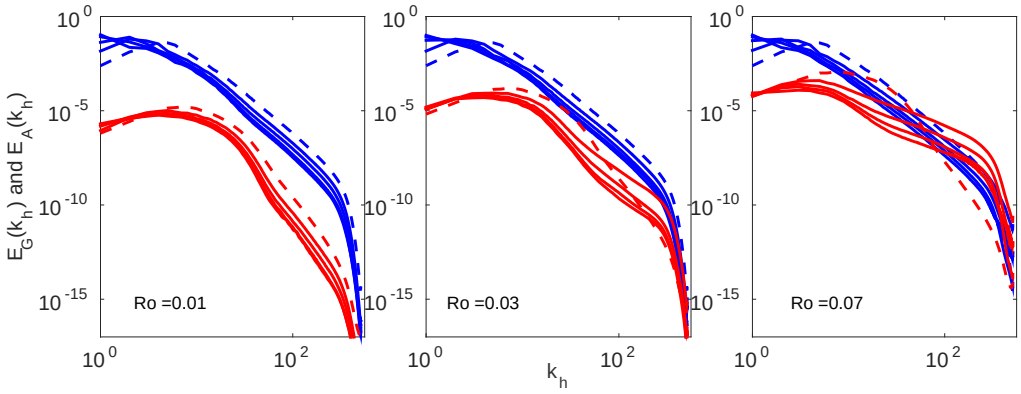


Figure 2–10: The geostrophic (top curves) and ageostrophics (lower curves) contributions to the energy spectrum for different Rossby numbers and at different times. The dashed lines correspond to $t = 0$, i.e. after Baer-Tribbia initialization and before Boussinesq simulation. The solid curves represent later times, namely $t = 5\tau$, $t = 10\tau$, $t = 15\tau$ and $t = 20\tau$. The corresponding runs are B1536r01lv, B1536r03lv and B1536r07lv.

in amplitude between initially balanced and unbalanced simulations, which makes the interaction between geostrophic and ageostrophic modes stronger.

"Subinertial" frequencies (less than f) correspond to the slow part of the dynamics. At low Rossby numbers it is almost the same for balanced and unbalanced flow, indicating they are dominated by the geostrophic modes, which are only weakly coupled to the ageostrophic modes. As *Ro* is increased even the slowly-varying part of the flow shows differences, a fact that was also seen in the time series of Figure 2–8. This, together with decreasing differences in amplitude of the wave band, makes fast and slow dynamics less distinct in the high-*Ro* regime. Although this behaviour was expected and predicted by previous studies employing two-timescale approaches, here we can quantify it.

2.6.4 The role of wave frequency in the breakdown of balance

After studying the difference between the dynamics of initially balanced and unbalanced flow, henceforth we focus on initially balanced flow and try to analyse its breakdown as a function of Ro and Fr, while also exploring sensitivity to the dissipation.

To investigate the role of rotation, in Figure 2–10 we vary Ro. In each panel the geostrophic and ageostrophic parts of the energy are portrayed together. Since the large scale dynamics are close to QG at low Rossby numbers, E_G spectra are well above E_A. However, this difference is reduced as rotation becomes weaker. Even in the initial condition (dashed lines) the difference between the two spectra tapers off with increasing Rossby number, which is consistent with the scaling of ageostrophic energy ($E_A \sim Ro^2$).

In all panels of Figure 2–10, the slope of E_A is steeper than E_G at $t = 0$ as expected from Figure 2–3. As time increases, the slope of E_G stays relatively insensitive to Ro at this resolution. By contrast, the ageostrophic slope changes substantially with Ro number at later times.

When rotation is as strong as $Ro = 0.01$ in the left panel of Figure 2–10 the slope of the ageostrophic spectrum remains steeper than the geostrophic slope. This is consistent with the maintenance of balance as unbalanced wave modes stay weaker than the higher-order corrections to geostrophic flow. As Ro increases, the ageostrophic slope quickly becomes more shallow. This can lead to a crossing of the two spectra when rotation is weak enough. This crossover may suggest strong interaction between geostrophic and ageostrophic modes. For instance, when Ro is increased to

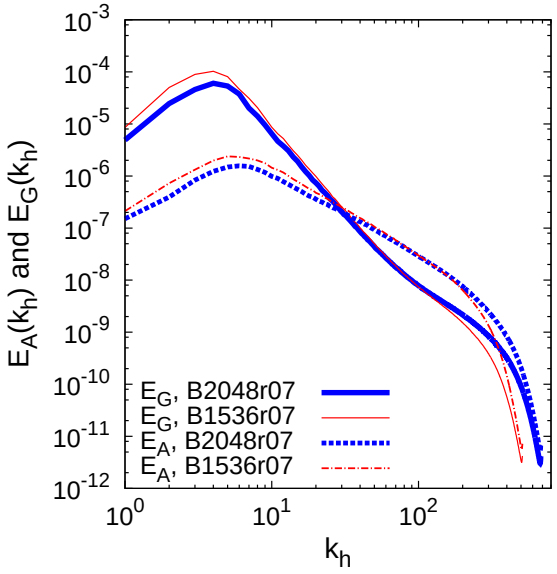


Figure 2–11: E_G and E_A spectra at $t = \tau$ for B1536r07 and B2048r07, whose initial Rossby number was 0.07.

0.07 in the right panel of Figure 2–10, E_A becomes larger than E_G above a certain wavenumber. However, these results are derived using grid 2) defined in Section 2.5; we performed similar simulations on grid 1) and found that imbalance grew even more quickly. This confirms that the generation of balance is a real feature of the flow, and is not imposed by our choice of numerical grid.

The geostrophic energy at a higher resolution (or equivalently higher Re) and at an earlier time is presented in Figure 2–11. Unlike the right panel of Figure 2–10, the geostrophic spectrum develops a shallow tail. The geostrophic shallow range, however, always occurs at wavenumbers larger than the onset of the ageostrophic shallow spectrum. If the resolved wavenumber range is sufficient, we expect both E_G and E_A to exhibit parallel slopes of $-5/3$ at very small scales as in Bartello (2010).

The E_A spectrum in this range is twice as large as E_G in amplitude since there are two ageostrophic modes corresponding to $\pm\sigma_{\boldsymbol{k}}$ (cf. (2.3) and (2.4)). At these small scales the turbulence becomes more 3D and closer to isotropic with our linear decomposition losing its meaning. Similar results were reported by Deusebio *et al.* (2013) in unbalanced turbulence. Even in the analysis of atmospheric data, Callies *et al.* (2014) observed that the tail of the geostrophic energy spectrum becomes shallow at smaller scales. They expressed that this shallowing "is likely an artifact, because at these scales the geostrophic component makes up a small fraction of the observed signal". It should also be noted that in our simulations the shallow tails decay in amplitude and are eventually suppressed as time advances, the turbulence decays, and Ro decreases.

Figure 2–12 portrays the geostrophic part of the energy spectrum for different rotations. In the left panel horizontal spectra are shown, while the right panel displays vertical spectra. It reveals that the geostrophic energy at different Ro displays very similar spectra over this range. However, it was shown in Figure 2–11 that at larger Re and at early times, the geostrophic spectrum can develop a shallow tail. The slope of the vertical geostrophic spectra is very close to -3, as predicted by the potential enstrophy cascade phenomenology, whereas the slope of the horizontal spectra is somewhat steeper. As discussed in §2.5.1, the steeper spectra are often attributed to coherent vorticity structures.

Unlike $E_G(k_h)$, $E_A(k_h)$ is highly affected by the strength of rotation and stratification. Figure 2–13 shows this dependence. The left panel presents E_A spectra for different Rossby numbers. They are scaled by the instantaneous $Ro(t)^2$, as justified

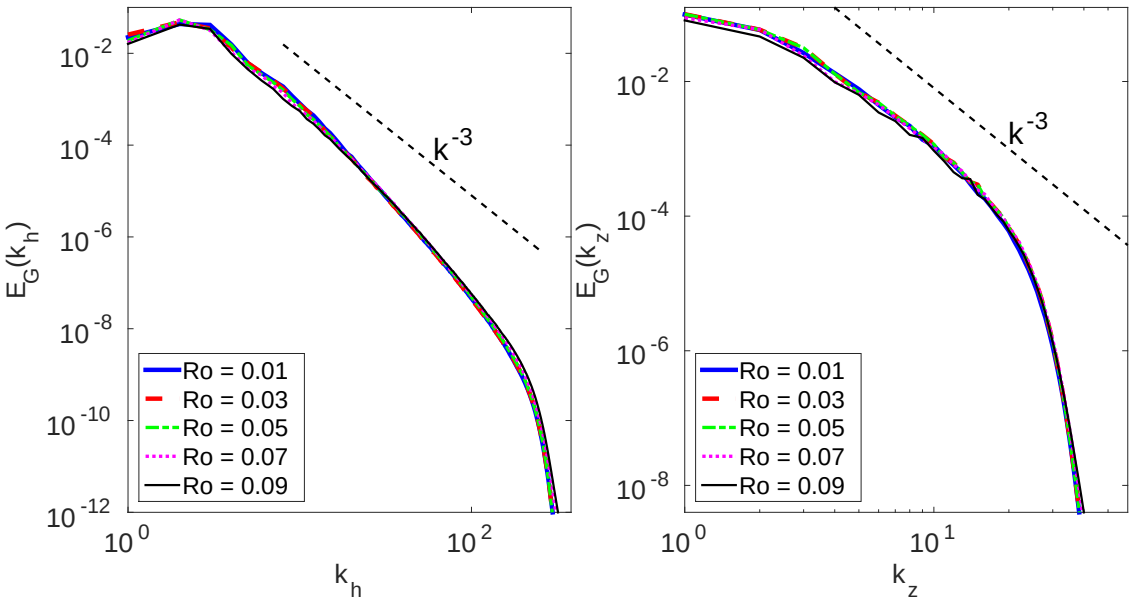


Figure 2–12: Right: E_G integrated vertically as a function of horizontal wavenumber, for different Ro. Left: horizontally integrated E_G as a function of vertical wavenumber. All the spectra are derived at $t = 10\tau$. Dashed lines in both plots have the slope of -3. The plots correspond to B1536r01, B1536r03, B1536r05, B1536r07 and B1536r09.

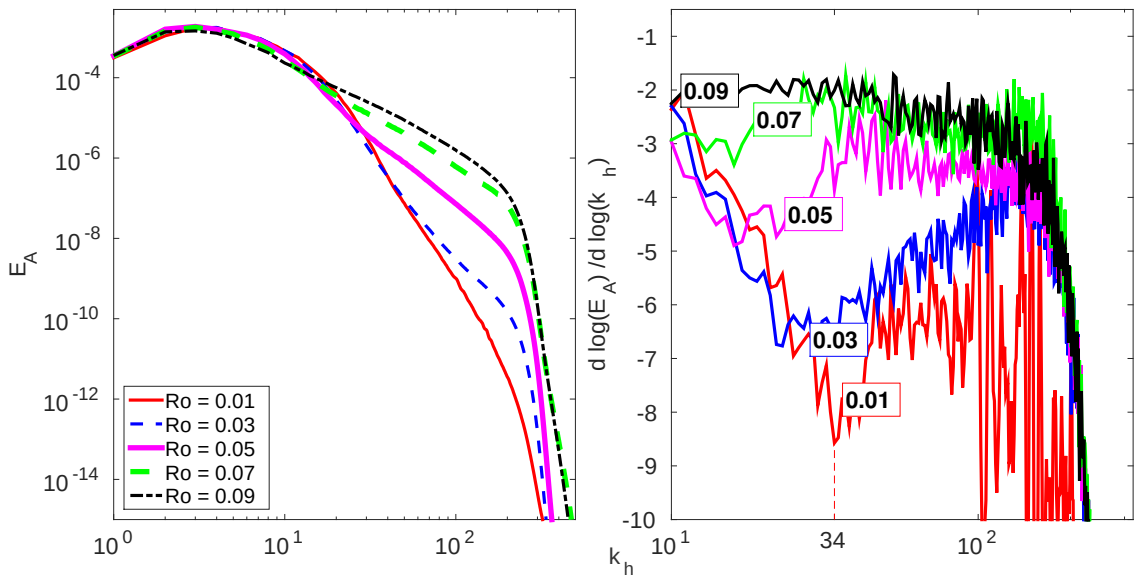


Figure 2–13: Left: E_A spectrum at different Ro. Right: the slope of E_A corresponding to the left panel. The same simulations as those of Figure 2–12 were used.

by (2.15a). The utility of this scaling appears in the ageostrophic energy spectra of large scales in Figure 2–13, where those at different Rossby numbers collapse to the same curve. Clearly, the balanced curve is the steep envelope with individual simulations departing from it at wavenumbers that decrease as Ro increases.

Similar to the observations made of Figure 2–10, in Figure 2–13 the spectral slopes at this fixed time become shallower as Ro increases. This difference can better be seen in the right panel of Figure 2–13, where the slopes of the ageostrophic spectra are depicted. The slopes are derived by calculating $\triangle(log(E_A))/\triangle(log(k_h))$ ($\triangle$ denoting forward Euler differencing). The flat range in the right panel corresponds to the power-law range of spectra in the left, and the local minima portray the kinks in the ageostrophic spectra.

At the lowest Rossby number, $Ro = 0.01$, there is a clear power-law range with a slope between $[-7, -6]$, similar to the results of Figure 2–3. At low Rossby numbers, there is a well-pronounced local minimum that indicates the kink in the spectrum. Another distinguishing feature of the $Ro = 0.01$ simulation is the fluctuation in slopes, which are greater than simulations at higher Rossby numbers. To understand their cause we should recall that in the left panel of Figure 2–13, E_A is scaled with Ro^2. The unscaled spectrum of $Ro = 0.01$ is below those of the higher Rossby numbers and its slope is steeper. Therefore, the ageostrophic energy is very small at large wavenumbers, where its value is close to machine round-off error. This is corroborated by the fact that the slope gets noisier as we go to smaller scales. As Ro is increased the kink moves to larger scales. This is best explained by the onset of spontaneous imbalance at larger scales at weaker rotations. When rotation is weak enough, the kink disappears and the slope asymptotes to a constant that is close to -2. Other simulations with lower hyperviscosity coefficients have more shallow slopes, approaching $-5/3$, but whose dissipation was not judged to be sufficiently well-resolved to be presented here.

In the limits of $Ro \to 0$ and $Ro \to \infty$, geostrophic and 3D turbulence are recovered, respectively. In these regimes, there are inertial ranges with universal slopes. The appearance of a power-law range in the spectra of Figure 2–10 gives rise to the question of whether there is a similar universal inertial range. To demonstrate this it would be necessary to show that its slope is not affected by small-scale dissipation. In fact, for the resolutions and Ro considered in this study, there is a dependence on

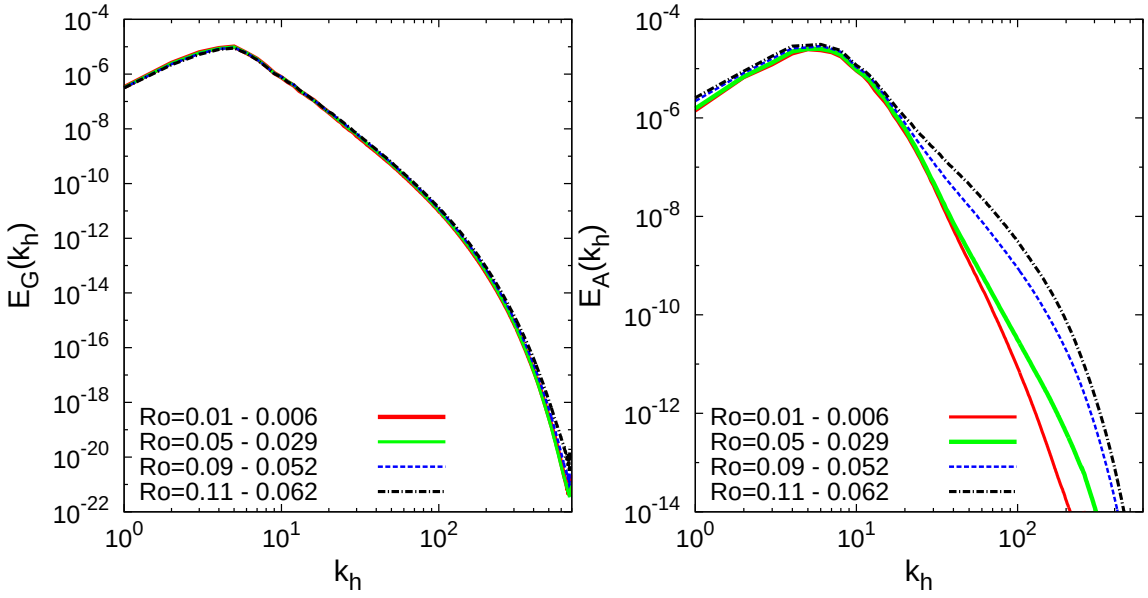


Figure 2–14: The geostrophic (left) and ageostrophic (right) part of energy spectrum for simulations with Laplacian viscosity at $t = 6\tau$. The ageostrophic energy spectra are scaled with Ro^2 as in Figure 2–13. The corresponding runs are B2028r01n, B2028r05n, B2028r09n and B2028r11n.

the viscosity and diffusion coefficients, even outside of the dissipation range and especially for mid-range Rossby numbers. Despite this fact, the general characteristics of Figure 2–13, such as the shallowing of spectra at increased Ro, were observed at four other hyperviscosity coefficients.

To verify these results using hyperviscosity we performed several simulations with Laplacian viscosity and diffusion at the higher resolution of $2048^2 \times 256$ (Figure 2–14). Similar to the geostrophic spectra of Figure 2–12, in Figure 2–14 there is not much variation in the geostrophic spectra at the different Rossby numbers. Similarities can also be seen in the ageostrophic spectra of Figures 2–13 and 2–14 as well.

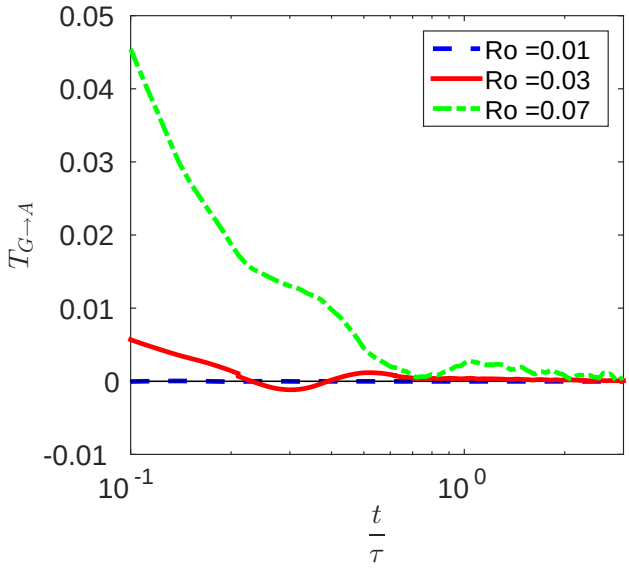


Figure 2–15: The total energy transferred from the geostrophic to ageostrophic modes via nonlinear interactions ($T_{G\to A}$) versus time for the runs: B2048r01, B2048r03 and B2048r07.

As before, we see in Figure 2–14 that the spectrum of ageostrophic energy becomes more shallow as Rossby number is increased. The slope of the ageostrophic modes in the lowest Rossby number simulation in Figure 2–14 is between -6 and -7, similar to the rapidly rotating cases using hyperviscosity.

For a better understanding of the dynamics, examining the energy fluxes and transfers in tandem with the energy spectra is very helpful. Figure 2–15 depicts the net transfer from the geostrophic to ageostrophic modes through nonlinear interactions shown in the equation (2.7). At low Rossby number this transfer is almost zero, showing that the exchange of energy between geostrophic and ageostrophic modes is very small. As rotation weakens the transfer between geostrophic and ageostrophic modes becomes larger. This exchange is higher at early times and it gets smaller with

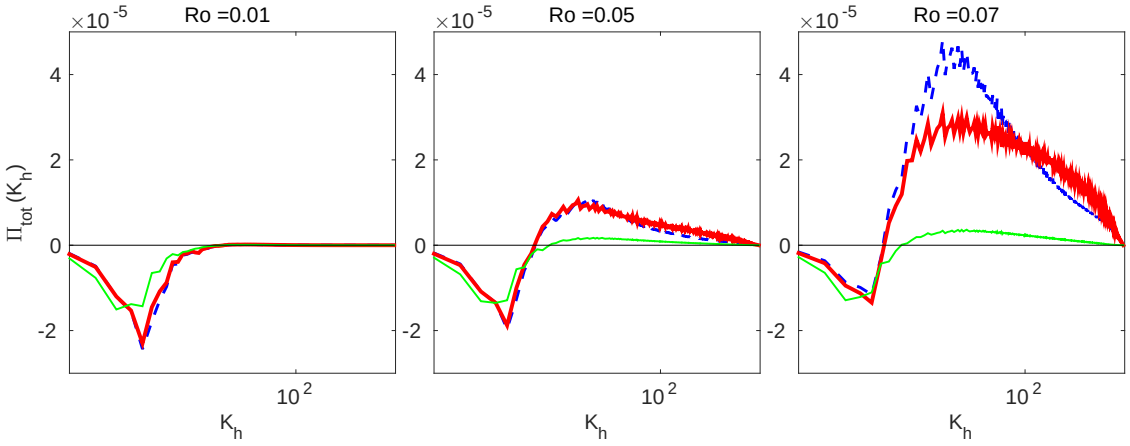


Figure 2–16: The energy flux spectra for B2048r01, B2048r05 and B2048r07. The dashed lines (blue online) are the flux spectra averaged over $[0, 0.2\tau]$, the thick solid lines (red online) averaged over $[0.2\tau, 0.4\tau]$ and the thin solid lines (green online) averaged over $[1.6\tau, 1.8\tau]$.

time. At the higher Ro, the transfer is dominantly from geostrophic to ageostrophic modes, which is consistent with the energy plots of Figure 2–5. In Figure 2–5 the level of E_A increases at the beginning. Given there is no forcing, this increase must come from the geostrophic modal energy.

The total energy flux, which is the sum of $\Pi_G(k_h)$ and $\Pi_A(k_h)$ defined in (2.8), is plotted in Figure 2–16. At $Ro = 0.01$ it follows the QG scenario. There is a negative trough in the large scales, implying an inverse cascade of energy. As the Rossby number increases a positive-flux range emerges at scales below the peak in the spectrum that signals a forward cascade of energy in these flows. It is stronger at early times and becomes weaker in time as Ro decreases.

To provide more insight on the role of rotation and stratification on the breakdown of balance, the local Rossby number is portrayed in 3D in Figure 2–17. It is derived by scaling the vertical vorticity by the Coriolis parameter, i.e. $Ro(x, y, z, t) =$

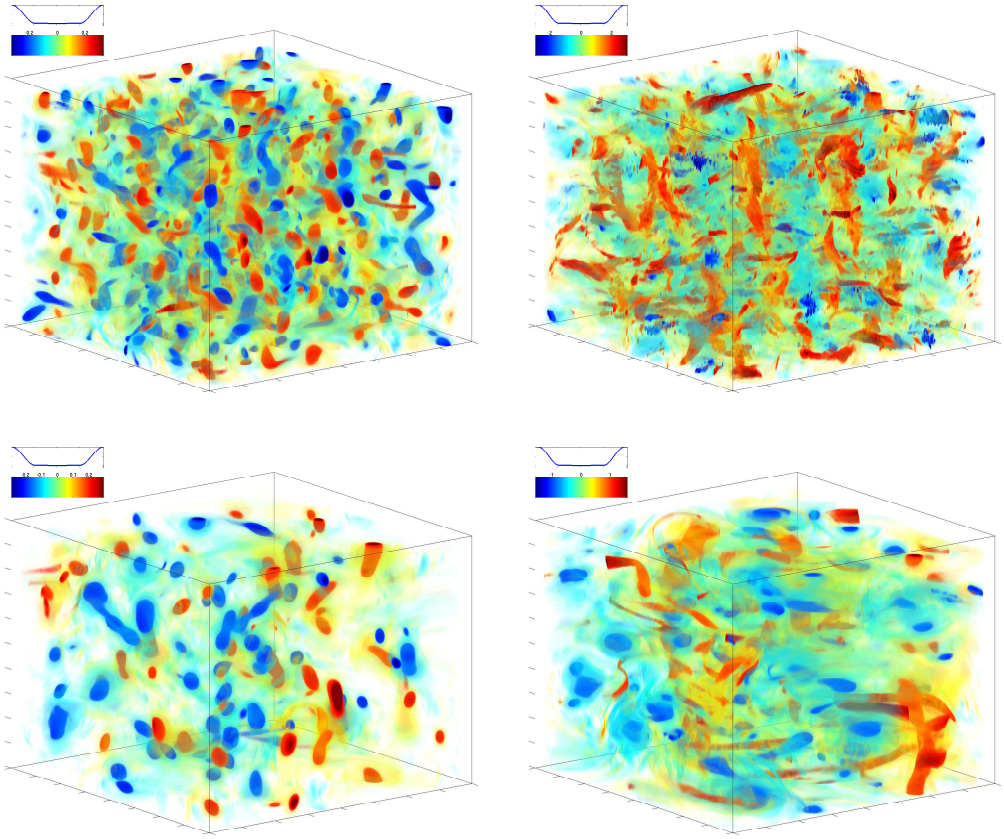


Figure 2–17: Local Rossby number, $Ro(x, y, z) = (\nabla \times \boldsymbol{u}(x, y, z) \cdot \hat{\boldsymbol{z}})/f$, in 3D space at $t = 5\tau$ (the first row) and $t = 20\tau$ (second row). The initial Ro is set to 0.01 in the first column and 0.09 in the second column. Values are displayed by assigning a color and a degree of transparency to them. There are two horizontal bars on the top left corner showing the transparency (upper curve) and the hue (lower colorbar) of grid values as a function of $Ro(x, y, z)$.

$(\nabla \times \boldsymbol{u}(x, y, z, t) \cdot \hat{\boldsymbol{z}})/f$. To better show the vertical structure, we scale the vertical axis by N/f. The transparency curve in the upper left corner of the figure is designed such that the points with large absolute values are opaque and visible, whereas lower absolute magnitudes of vorticity are almost transparent. This technique emphasises the intense structures of the flow.

The coherent structures of QG flow have been extensively studied in the past, for example in McWilliams *et al.* (1994) and McWilliams *et al.* (1999). The results are similar to the vortex structures visible in the left panels of Figure 2–17 (low Ro). This is quite expected as the strongly rotating flow remains dynamically close to QG, especially at larger scales where coherent vortex structures appear. As explained in a number of previous studies, there are two important mechanisms in the formation and evolution of vortices in geostrophic flows: the merger of two like-sign vortices (Melander *et al.*, 1988; Polvani *et al.*, 1989) and vertical alignment (McWilliams, 1989; Polvani, 1991). Due to the weaker rotation, these structures are observed less at $Ro = 0.09$. The plots of Figure 2–17 also manifest the difference in slope of the energy spectrum at low and high Rossby numbers. At $Ro = 0.01$, where the energy spectrum is steeper, the vortex structures have distinct boundaries, whereas at $Ro = 0.09$ the vortex cores are smaller and more diffuse. There is also stronger mixing at the higher Ro due to the shallower spectrum. The coherent structures make the flow more anisotropic at low Rossby numbers. Hence, we see that the slope of the energy spectrum is steeper than predicted by isotropic theory. However, at the higher Ro of Figure 2–17, the flow looks more isotropic. Hence, the energy spectra of

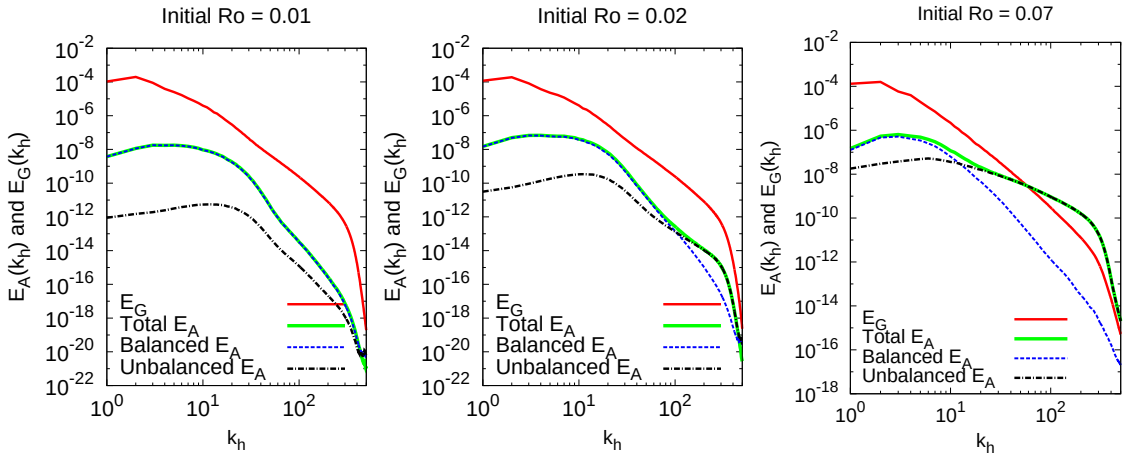


Figure 2–18: The balanced and unbalanced part of the ageostrophic energy spectrum. Geostrophic energy is shown for reference. The initial *Ro* is shown on the top of each panel.

weakly rotating flows scales with $k_h^{-5/3}$, which is similar to the scaling of 3D isotropic turbulence.

2.6.5 The persistence of balance

In this section, we try to examine the spontaneous imbalance generated in initially balanced flow as time grows. To that end we take the output of the Boussinesq simulation after $t = 10\tau$ and implement Baer-Tribbia initialization again to derive the balanced part of the motion. Then, we subtract the balanced field from the total field to obtain the unbalanced part of the motion.

Figure 2–18 shows the balanced and unbalanced as well as total ageostrophic spectrum. In the left panel, the lowest *Ro* is depicted. Since the rotation is very strong, the flow stays in balance. For this reason the balanced and total spectra lie on top of each other and the unbalanced spectrum is negligibly small. For the middle *Ro*, a kink emerges in the total ageostrophic spectrum. The balanced spectrum

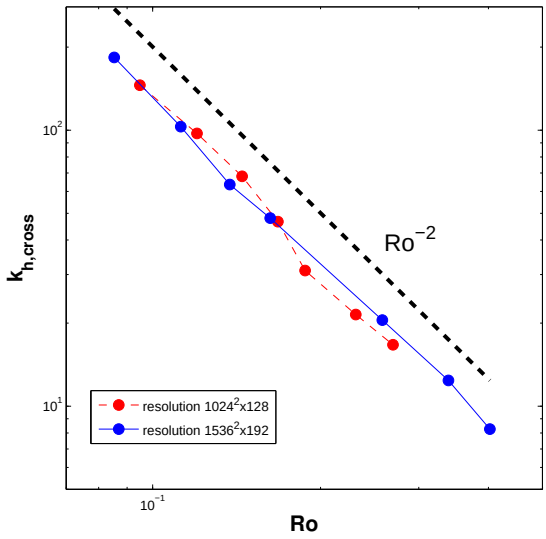


Figure 2–19: The wavenumber at which the balanced and unbalanced part of the ageostrophic energy spectrum cross versus (vorticity-based) *Ro*.

retains its steep slope. However, the shallow unbalanced part grows in amplitude and intersects the balanced spectrum. Hence, the total ageostrophic spectrum, which is the sum of the two, displays a kink near their intersection. At the highest *Ro*, the unbalanced spectrum grows further in magnitude and the crossover wavenumber moves to larger scales. Since a shallow range of the total ageostrophic spectrum emerges at larger scales, we can see that the geostrophic and ageostrophic spectra intersect as well.

To see how the onset of breakdown scales with *Ro*, we plotted the wavenumber at which balanced and unbalanced ageostrophic spectra cross ($k_{h,cross}$) as a function of *Ro* in Figure 2–19. This was extracted for a number of *Ro* and at two different

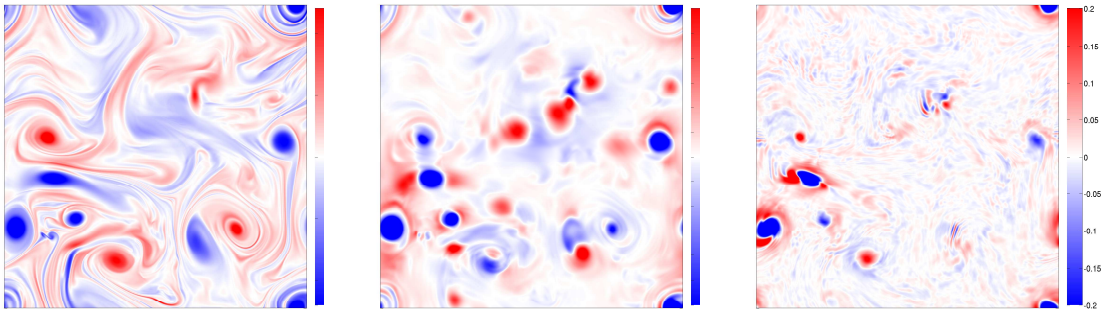
resolutions. It clearly shows that $k_{h,cross}$ scales with Ro^{-2}, just as the total balanced ageostrophic energy.

Considering all three panels of Figure 2–18 along with Figure 2–19, we can elucidate the mechanism of balance breakdown. The ageostrophic component consists of a balanced and unbalanced part

$$A_{\boldsymbol{k}} = A_{\boldsymbol{k}}^{balanced} + A_{\boldsymbol{k}}^{unbalanced}. \tag{2.21}$$

The balanced spectrum has a steep slope (close to -6) while the unbalanced slope is shallow. We speculate that the unbalanced slope is close to $-5/3$, given sufficient resolution, or equivalently high enough Reynolds numbers. As rotation weakens, the shallow unbalanced part of the spectrum increases in amplitude and hence extends to larger scales. The kink therefore appears at lower wavenumbers. This shallow part of the ageostrophic energy spectrum may intersect the steeper geostrophic one and can hardly be neglected in the dynamics at smaller scales. Therefore, the total energy consists of a steeper spectrum at large scales, dominated by geostrophic dynamics, and a shallow one at small scales, dominated by unbalanced ageostrophic dynamics. This is consistent with atmospheric data (Gage, 1979) and the previous unbalanced simulations of Bartello (2010). Further below this intersection scale, geostrophic modes also transition to a shallow spectrum as the linear decomposition loses its meaning.

The scaling of Figure 2–19 can be explained by considering (2.13), where the ageostrophic modes are expanded in terms of Ro and kept to first order. Hence, it can be concluded that $A^{balanced}/A^{unbalanced} = O(Ro^{-1})$ in our initialization scheme.



(a) ζ_{total} (b) $\zeta_{balanced\ ageo}$ (c) $\zeta_{unbalanced\ ageo}$

Figure 2–20: Horizontal slices of vertical vorticity ($\zeta = \nabla \times \boldsymbol{u}(x, y, z) \cdot \hat{\boldsymbol{z}}$) for (a) the total flow, (b) balanced ageostrophic and (c) unbalanced ageostrophic modes. The total flow has the initial Rossby number of 0.035.

Therefore, the balanced fraction of the ageostrophic energy is $O(Ro^{-2})$ larger than the unbalanced part. Hence, the change in the crossover wavenumber is expected to be proportional to Ro^{-2} as well. A similar argument can be made for the wavenumber at which geostrophic and ageostrophic spectra cross. It should be noted that other definitions of balance (at higher order, for example) produce different scalings for $A^{balanced}/A^{unbalanced}$ at large Ro.

A visualization of total, balanced and unbalanced parts of the vertical vorticity is displayed in Figure 2–20. The steep spectra of balanced ageostrophic modes and the shallow spectra of unbalanced modes are reflected in the real-space plots of panels 2–20b and 2–20c, respectively. The unbalanced part of the ageostrophic modes appears as small-scale structures surrounding the larger-scale vortices, whereas the balanced part embodies smooth vortical cores similar in scale to those of the total flow. These ageostrophic centres do not necessarily have the same signs as those of the total flow, which are predominantly geostrophic. Note that in figure 2–20c there are also some

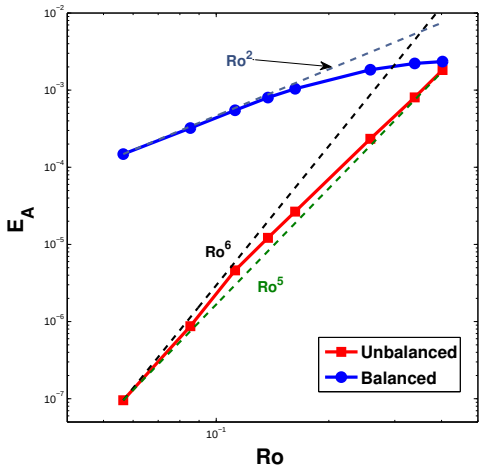


Figure 2–21: The unbalanced and balanced part of ageostrophic energy as a function of (vorticity-based) Ro at $t = 10\tau$.

large-scale extrema corresponding to those of the total flow. This could be due to the shortcoming of our balance model which is only second order in Ro. It may be that the vortical cores appearing there are even higher-order corrections to balance. An example of a dipole can be seen in the lower left part of the domain. There is a strong bundle of waves in the unbalanced panel travelling with this dipole. This is similar to the slow-moving inertia-gravity wave packets formed along dipoles in the study of Snyder *et al.* (2007) and Viúdez (2007, 2008). Note that these studies considered isolated dipoles, while we are considering a large distribution of vortices in homogeneous turbulence.

Figure 2–21 presents the balanced and imbalance ageostrophic energy summed over all wavenumbers as a function of Ro. According to (2.15a), when the flow is balanced we expect $A_{\boldsymbol{k}} \sim O(Ro)$. Hence, the ageostrophic energy asymptotes to Ro^2 as $Ro \to 0$. Figure 2–21 also shows that the unbalanced ageostrophic energy

increases faster as rotation gets weak. The unbalanced ageostrophic energy scales as Ro^{α}, with α between 5 and 6, whereas the balanced energy scales as Ro^2 at low Ro and becomes even weaker at larger Ro. At small Rossby numbers, different balance definitions should asymptote to the same behaviour as in Figure 2–21. However, as Ro increases the scaling of unbalanced ageostrophic energy is subject to the error in the definition of balance.

2.7 Conclusion

We have described the evolution of nonlinearly balanced initial conditions under the nonhydrostatic Boussinesq equations. To that end, we used the Baer-Tribbia scheme, which is a nonlinear normal mode initialization method enforcing balance at $O(Ro^2)$ by calculating ageostrophic modes that eliminate fast time derivatives, at least initially.

We investigated how initially balanced and energetically-equivalent unbalanced flows differ in their evolution in time and modal interplay. At strong rotation the time series of large-scale average quantities of initially balanced and unbalanced simulations do not show large differences. Nevertheless, there is a considerable difference in the frequency spectra over the range of their linear inertia-gravity wave frequencies. This shows that the initialization can lower the amplitude of these wave modes while not substantially affecting slower sub-inertial frequencies associated with the large-scale quasigeostrophic flow. Of course the effectiveness of this procedure is reduced as Ro increases.

The initialization produces an ageostrophic spectrum that is much steeper than the (already steep) geostrophic one. Hence, one can conclude that balance occurs

in conjunction with steep spectra and that balance dynamics cannot explain the observed sub-deformation scale -5/3 spectra in the atmosphere and ocean. If Ro is small enough, ageostrophic modes maintain their steep balanced spectrum at a given resolution and over a given integration period. As Ro is increased, a shallow tail develops in the ageostrophic spectrum after a short time, which we have identified as signalling the growth of unbalanced dynamics. The onset of this unbalanced range occurs at lower wavenumbers when Ro is increased further. It also becomes increasingly shallow and asymptotes to a slope consistent with $k^{-5/3}$, if the Reynolds number is high enough. In fact, we found the slope of the shallow range to be somewhat sensitive to viscosity and diffusion at our resolutions.

Our results for strong rotation confirm and complete the picture drawn by Dritschel & McKiver (2015) for the maintenance of balance. At very small Rossby, the ageostrophic spectrum stays balanced and steep in the resolutions considered in their study. However, our numerous simulations at different Ro and Re hint that any initially balanced rotating stratified flow develops an unbalanced shallow tail at sufficiently large times if the resolution is large enough (or equivalently dissipation is weak enough) to permit it. This is indeed an open question that can be addressed with higher resolution simulations. Even if future investigations show the existence of a shallow tail at very small Ro, it will clearly be at very small scales. Considering that its energy will then be very low, unbalanced energy is much smaller than balanced energy, once again consistent with Dritschel & McKiver (2015).

The shallow spectrum resulting from the growth of unbalanced ageostrophic modes is clearly consistent with the atmospheric data (Gage & Nastrom, 1986).

However, this similarity should be considered within the limitation of our idealized boundary conditions. Internal and external boundaries have been shown to play an important role in the breakdown of balance. Nevertheless, characteristics of unbalanced modes generated spontaneously can be analysed in a periodic configuration such as ours as a first step in understanding at least the internal dynamics far from boundaries.

When Ro is large enough that the ageostrophic spectrum admits an unbalanced shallow range, the corresponding frequency spectrum does not exhibit large peaks in the frequency band between N and f. This signals that the developing unbalanced part of the dynamics is not composed of quasi-linear high-frequency inertia-gravity waves but a transition to a more general form of turbulence involving both stratification and rotation.

Acknowledgements

As well as the three reviewers, the authors would like to thank Eyad Atallah, Jim McWilliams and Joe Tribbia for their valuable input. Financial support was received from the Natural Sciences and Engineering Research Council and computer resources were generously provided by Compute Canada/Calcul Québec.

CHAPTER 3

Rotating stratified turbulence and the slow manifold

We established our methodology in the previous chapter. In the current chapter, we expand our study by the post-processing of the results derived from this method. More particularly, we study the energy transfer and flux spectra of simulations at different flow parameters. By dividing the transfers into balance, imbalance and cross-terms, we calculate the contribution of each to the the total energy transfer at each wavenumber. We also portray the frequency spectra of geostrophic and ageostrophic modes at different Rossby numbers and discuss at which length scale the separation of time scales breaks down. We finish this chapter with a rather detailed discussion on the choice of vertical grid since there were two obvious choices and we wished to establish the robustness of our results.

This chapter looks at balance from a slow manifold perspective. As explained above, Leith (1980) pioneered the concept of the slow manifold as a geometric expression of balance in phase space. The invariance of trajectories on the slow manifold has been questioned by many previous studies (see the references in the introduction of this chapter). We quantify up to what scales this manifold stays approximately invariant at different Ro and Fr. Our initialisation procedure that was described

in the previous chapter is an approximate projection onto the slow manifold. The time-integration of the non-hydrostatic Boussinesq equations can perhaps be seen as following trajectories that start on the manifold. We then investigate how long these trajectories take to depart from the slow manifold. This is measured after the breakdown of balance by calculating the amount of imbalance generation.

This chapter is based on the following paper, which was pertinently published in a special issue in honour of **Cecil "Chuck" E. Leith**:

Kafiabad, H. A., & Bartello, P. (2017). Rotating stratified turbulence and the slow manifold. *Computers & Fluids*, 151, 23-34.

Leith peacefully passed away at the age of 93 during the course of my PhD. I dedicate this chapter to this great scientist whose contribution to the field was invaluable.

3.1 Abstract

Numerical simulations of decaying rotating stratified turbulence were performed from balanced initial conditions using the Baer-Tribbia initialisation scheme as an approximate projection onto the slow manifold of the non-hydrostatic Boussinesq equations. Starting with the low Rossby and Froude numbers characterising the quasigeostrophic limit, these were increased somewhat until small-scale balance appeared to break down. Following a previous paper by the authors, this occurred in conjunction with the emergence of a shallow range in the energy spectrum. The goal here is to work towards identifying the mechanism and characteristic scales of the resulting spontaneous imbalance. It is found that it originates at relatively large scales near the low-wave number end of the shallow spectral range. Unbalanced energy then cascades to small scales where it is efficiently dissipated in the decay case. The predominant interaction over the shallow range was seen to be a downscale cascade of unbalanced energy with little interaction with the balanced flow. At even smaller scales the decomposition loses its meaning and the spectrum remains shallow.

3.2 Introduction

It is a very interesting exercise in 2016 to look back to the early days of numerical modelling of both atmosphere and ocean in order to see how approximations to the dynamics were developed, implemented and verified against observations. Indeed, we find it rather strange today that, whereas the fully compressible Navier-Stokes equations were known in the late 19th century, the much simpler quasigeostrophic (QG) set was not derived until ca. 1950 (Sutcliff (1947), Obukhov (1949), Charney (1949) [1]). This was simply because we needed to know the characteristic scales of the flow and that had to wait until widespread commercial aviation could provide the observations, particularly of low Rossby, $Ro \sim U/fL$, and Froude, $Fr \sim U/NH$, numbers. Here, U is a characteristic velocity scale, H and L are vertical and horizontal length scales, respectively, while f is the modulus of twice the horizontal projection of the rotation vector and N is the stratification frequency. Following the successful implementation of the QG equations in a numerical model covering a limited mid-latitude domain (Charney *et al.*, 1950), computers and grids rapidly increased in power and size, respectively, eventually covering even tropical regions where the Rossby number could not possibly be small. Numerical specialists soon realised it was necessary to take a step back and employ the "primitive" (or hydrostatic) equations (PE), thereby undoing some of the assumptions made in deriving the QG set. In addition to motions that resembled synoptic-scale meteorological

[1] For an explanation of this reference list see Phillips (1982).

systems, the PE set contained additional degrees of freedom. Particularly troublesome were the high-frequency inertia-gravity waves, which were filtered from the QG equations by the assumption of slow variability. It was soon found that errors in observations projected onto all degrees of freedom of the primitive equations in a close-to-random fashion, causing unrealistic high-frequency motion.

Scale analysis had shown the high-frequency motion to be close to linear and the slowly-varying nonlinear motion to resemble the meteorological signal. As a result, there emerged a culture of dividing up the large-scale turbulence into high-frequency noise and low-frequency signal. The question was then how to constrain the forecast system to remain slowly-varying. One approach was already clear: analytical filtering leading to the QG set, but a tractable, globally-valid, slowly-varying analytical model remained elusive (and still does). Another approach was to somehow constrain the initial conditions such that the noise was not excited. It seems reasonable to speculate that these pioneers wondered why this high-frequency noise was so easily excited in the models and yet so rare in the real atmosphere. The principal difference was, of course, that new data with their associated error were periodically injected into the forecast system. Random observational error was seen as forcing the unphysical motion. Merely setting high-frequency modes to zero initially (linear balance) did not help, as the waves were quickly excited. Machenhauer (1977), Baer (1977), Baer & Tribbia (1977) and Browning & Kreiss (1982) developed nonlinear techniques to make small, systematic adjustments to initial conditions so that the initial tendency to generate high-frequency motion was minimised, at least to a given order.

It is at this point that Leith (1980) provided the seminal work that relates these techniques to the first approach, that of analytically deriving equations employing balances between terms that maintain the slow variability of their solution. Leith (1980) also made it clear that quasigeostrophy is the lowest order in a hierarchy of balance models and that, following Hoskins *et al.* (1978), the concept can be taken to higher order. In fact, while underlining the lack of proof of its existence, he pointed out that prevailing thinking at the time suggested an invariant slow manifold in the phase space of the dynamical system. If a phase point was initially on it, it would remain there and evolve slowly. If it were somewhere else, it would evolve towards it (likely in a very oscillatory way) eventually winding up on it, as the atmosphere had presumably done during the course of its existence. This work, along with Lorenz (1980), described mathematically the essential properties of this manifold.

The existence (really the invariance) of the slow manifold in this strict sense was first questioned by Warn in a study that was published much later as Warn (1997). Acknowledging this argument Lorenz (1986) showed that attempts to project initial conditions onto the slow manifold failed to converge, even in truncated low-order dynamical systems. Although the (approximate) invariance of the slow manifold has been the subject of discussion for many decades, it seems there is a consensus now that in the regimes of low Ro and Fr it still has some validity as unbalanced components are observed to be limited in amplitude (see e.g., Warn & Menard (1986), Lorenz & Krishnamurthy (1987), Bokhove & Shepherd (1996), Ford *et al.* (2000) and Saujani & Shepherd (2002)). A more extensive review of the literature on the slow manifold in reduced dynamical systems can be found in **?**.

Even though much attention was directed at low-order dynamical systems in order to simplify outstanding questions relating to the slow manifold, meteorologists clearly realised the atmosphere is a very turbulent flow spanning an enormous range of time and length scales. Charney (1971) had drawn the analogy between QG and 2-D turbulence, identifying inverse energy and direct potential enstrophy cascades in the QG limit. Although in a very different limit another study of interest is that of Bartello (1995) who showed that low-Ro, low-Fr turbulence that is initially very far from balance would become more balanced with time. Using the normal-mode decomposition of the nonhydrostatic Boussinesq equations he showed that the geostrophic modes decoupled from the ageostrophic modes. Thus, their energy was systematically transferred to larger scales as in Charney (1971), even in the presence of finite-amplitude waves. At the same time, the ageostrophic modes were predominantly transferred to smaller scales (with a more shallow spectrum) via a catalytic interaction involving two ageostrophic modes and a geostrophic mode, wherein the geostrophic mode remained unchanged. This interaction was originally described for shallow water by Warn (1986) and subsequently studied in detail for stratified turbulence by Lelong & Riley (1991). It has the effect of transferring the high-frequency energy towards the dissipation, yielding balanced large-scale turbulence at low Ro and Fr.

About the year 2000 it was apparent that, although the large-scale dynamics were restricted to "near" a slow manifold, there was no quantitative prediction of how near. Clearly this was a function of the Rossby number, which is the ratio of slow to fast time scales. In the largest scales this is very small, convergence of

perturbation series is rapid and quasigeostrophy had a fifty-year history of proving accurate, but as the horizontal length scale is diminished, time scales were smaller too and convergence is less rapid. In the very small scales simulations of atmospheric convection with and without earth's rotation proved to be very similar (e.g., Lilly *et al.* (1998)). It is in this limit one expects the manifold to lose its dynamical significance or equivalently its width to become large in some sense.

Another element of meteorological and oceanographic culture was that one loses the predominance of rotation before that of stratification as horizontal length scale is reduced. The ratio of stratification frequency (N) to rotation frequency (f) is often quoted as 100 to one. For this reason it is of interest to examine studies of stratified turbulence without rotation as a candidate for smaller scales in the atmosphere and ocean where the approximate slow manifold may become rather wide, or even break down. Stratified turbulence without rotation was studied by Lilly (1983) using scale analysis for both the advective, or vortical, time scale, $T_v \sim L/U$ and the wave time scale, $T_w \sim L/NH$. Their ratio is the Froude number, $Fr \sim T_w/T_v \sim U/NH$. Following Charney's lead, Lilly (1983) took this ratio to zero and derived equations describing near-linear gravity waves with T_w and layerwise 2-D flow using T_v.

Later, Werne & Fritts (1999) simulated a stratified shear layer and were surprised by the fact that the Froude number did not become small after the Kelvin-Helmholtz instability took place and their initial mean flow turned into decaying stratified turbulence. Instead, it resolutely maintained a Froude number of order unity. Billant & Chomaz (2001) showed that the nonhydrostatic Boussinesq equations were self-similar in Nz/U, at least to leading order. This implies that the vertical scale is

U/N, or equivalently that the Froude number is unity. Subsequent simulations by Laval *et al.* (2003) showed that the flow was consistent with Lilly's prediction only at low Reynolds numbers. Other simulations by Riley & deBruynKops (2003) showed a rapid decay of stratified turbulence until the Froude number was of order unity, followed by a much slower decay. Waite & Bartello (2004) were aware of these results when they systematically measured the tendency for stratified turbulence to maintain a Froude number of unity over a range of Brunt-Väisälä frequencies, N.

It must be stated that Lilly was aware of the inherent inconsistency of the logic behind the limit $Fr \to 0$. Since the resulting vortical dynamics can be described as layerwise 2-D flow, one can consider the flow at two different values of z to be like two realisations of the 2-D predictability problem (see Leith & Kraichnan (1972)), assuming there is a small 3-D perturbation confined to small scales. The differences in the mainly-2-D flow will grow with time until the two layers decouple. They can be very close together in z as long as they are separated by at least the scale of molecular diffusion and viscosity and therefore the ensuing vertical scales can be quite small. This implies the Froude number can become large, thereby invalidating Lilly's low-Fr analysis. This process was explained with considerably more rigour by Babin *et al.* (1996). In conclusion, the current thinking is that the vertical Froude number, and hence the wave/vortical time scale ratio, remains order unity if the vertical scale is set by the stratification. As a result, the very idea that the number of active degrees of freedom can be reduced by selecting only the slowly-varying ones is invalidated and there can be no slow manifold at these scales. This is not to contradict the presence of large-scale low-Ro, low-Fr rotating stratified flow that remains

approximately quasigeostrophic, hence balanced, since QG energy is transferred to larger scales and fast wave energy cascades down towards the dissipation. However, as we go down scale this eventually turns into completely unbalanced stratified turbulence, where there is a single time scale and ultimately to isotropic turbulence at the Ozmidov (1963) scale. In the view of the authors, the question now becomes how does balance break down at scales between the deformation scale and those that characterise stratified turbulence? Equivalently, the question can be posed in terms of the "width" of the approximate slow manifold, where it is clear that wider would imply the need for higher-order definitions of balance.

There are also a number of perplexing issues related to the cascade properties of the turbulence. Atmospheric data show a transition to a shallow -5/3 range at around a few hundred kilometres in the horizontal Gage & Nastrom (1986) over which an analysis of third-order structure functions showed that the energy cascade was toward smaller scales (Lindborg & Cho, 2000). Recent simulations have shown the energy cascade in stratified turbulence without rotation also to be towards smaller scales (Waite & Bartello, 2004, 2006*a*; Lindborg, 2006; Bartello & Tobias, 2013; Maffioli & Davidson, 2015). Given that stratification determines the vertical scale in maintaining $Fr \sim 1$, one might ask what happens in the horizontal. Lindborg (2006) speculated that it remains as it is in unstratified turbulence, i.e., as in Kolmogorov for the kinetic energy and Corrsin-Obukhov for the potential energy. Bartello & Tobias (2013) showed in DNS that changes that had previously been observed in the slope of the horizontal spectrum were the result of changes in the Reynolds number and not of changes in the stratification itself, suggesting that it may indeed be independent

of stratification as $Re \to \infty$. Recently Maffioli & Davidson (2015) used DNS to find that the energy dissipation, ϵ, followed the unstratified relation $\epsilon \sim U^3/L$, where L is the horizontal scale. It would therefore appear that as stratification is increased, the vertical scale becomes proportionately smaller while the horizontal spectrum remains unaffected and, most importantly, the cascade rate ϵ is constant.

These admittedly preliminary findings can be used to determine the scale at which rotation becomes unimportant using an Ozmidov-style argument. Recall that the small scales of turbulence vary with a frequency $[k^3E(k)]^{1/2}$ if $E(k)$ is sufficiently shallow. Ozmidov (1963) argued that the wavenumber above which stratification was unimportant, k_o, can be defined by the relation $N \sim [k_o^3E(k_o)]^{1/2}$. He used K41 for $E(k)$ and found $\ell_o \sim (\epsilon/N^3)^{1/2}$ for the length scale, arguing that below this scale the turbulence was fast enough compared to the stratification as to render the stratification dimensionally unimportant. Recent results on stratified turbulence without rotation suggest that the horizontal spectrum remains as in K41. The same logic therefore yields $\ell_f \sim (\epsilon/f^3)^{1/2}$ as the horizontal scale below which neglecting rotation would be justified. One can simply read the value off the Nastrom-Gage data as $\epsilon \approx 10^{-4}\mathrm{m}^2\mathrm{s}^{-3}$ and using $f \approx 10^{-4}\mathrm{s}^{-1}$, one obtains $\ell_f \sim 10$ km. The conclusion is therefore that the transition to a shallow spectrum in the Nastrom-Gage data occurs at a scale where it is not accurate to neglect rotation.

To us the outstanding question is therefore how do the large-scale rotating-stratified balance dynamics change at $O(100)$ km to yield the observed shallow spectrum? To begin this exploration Bartello (2010) forced large-scale geostrophic motion in a numerical domain whose aspect ratio followed quasigeostrophic scaling

as $H/L \sim f/N$. This was performed at various Rossby numbers. As Ro was increased the tail of the spectrum became increasingly shallow, as in the data (see also Nadiga (2014)). A decomposition into geostrophic and ageostrophic normal modes revealed that the geostrophic modal spectrum was steep at large scales, while the ageostrophic energy spectrum, though far below the geostrophic modes, was much more shallow, following an approximate $k_h^{-5/3}$ form. At larger k_h the two crossed, implying the total energy became more shallow. At still larger k_h both geostrophic and ageostrophic contributions became equally shallow as the linear terms on which the decomposition is based became less predominant. While the results of this study were consistent with the observed spectrum, it could be argued that forcing only large-scale geostrophic modes, although linearly balanced, would not produce the nonlinear balance described in Leith (1980). In fact, it was well known to NWP practitioners that this would directly generate high-frequency waves. On the other hand, since the ocean is forced by wind stress resulting from atmospheric variability on a wide range of time scales, these simulations may be more applicable to that case. Waves generated at the low-Ro, low-Fr forcing scale would cascade down scale via the catalytic interaction described in Bartello (1995) and later observed at much higher resolution by Marino *et al.* (2015)). If so, the quasigesotrophic part of the flow would be unaffected by the finite-amplitude waves.

Kafiabad & Bartello (2016) employed a nonlinear initialisation scheme proposed by Baer & Tribbia (1977) along with the geostrophic modes from a preliminary QG simulation to project initial conditions onto the slowly-varying degrees of freedom, at least in an approximate sense. They found that the balanced flow constructed by

this nonlinear initialisation yielded steep spectra, both for geostrophic modes and more surprisingly their high-order ageostrophic corrections. The power-law range at which balanced ageostrophic modes admit such a steep spectra starts from wavenumbers larger than the peak of the geostrophic spectrum and extends to the dissipation range. More importantly, this initial condition was found to be unstable below a certain Ro-dependent scale if the Reynolds number was large enough, and the spectrum became more shallow. In fact, the spectral transition was entirely consistent with the atmospheric data of Gage & Nastrom (1986) and the oceanic spectra of Klymak & Moum (2007*a*,*b*). Employing the rotational modes after $O(10)$ eddy turnover times of evolution in the nonhydrostatic Boussinesq model and calculating the balanced ageostrophic modes from the Baer-Tribbia scheme, Kafiabad & Bartello (2016) were able to separate the ageostrophic motion into balanced and unbalanced components. The shallow spectral tail at large horizontal wave numbers was found to be entirely unbalanced. The goal here is to continue this study by examining how initially balanced flows transfer energy into unbalanced small-scale modes. The subject has potential implications for the energy budget of the ocean and for parameterising small-scale mixing. It may also be important for designing future numerical approaches in terms of the ratio of horizontal to vertical resolution, temporal resolution, etc.

After stating the theoretical basis of the approach in Section 2 and our numerical configuration in Section 3, we present our main result that the unbalanced motion is generated by interaction with balanced degrees of freedom at relatively large scales with our parameters. This then becomes the large-scale end of the shallow spectral

range, where there is a diminishing interaction with the balanced motion at smaller scales. Over this range the Rossby and Froude numbers are increasing with wave number and a transition to stratified turbulence, where rotation has a negligible influence, would be expected at smaller scales if the Reynolds number is large enough, or equivalently here, the numerical resolution. Some speculation on the consequences of these facts is offered in the conclusions.

3.3 Theoretical background

3.3.1 Governing equations and normal modes

We consider the three-dimensional non-hydrostatic Fourier-space Boussinesq equations

$$\frac{\partial u_i(\boldsymbol{k})}{\partial t} + \epsilon_{mji} f u_j(\boldsymbol{k})\delta_{m3} + ik_i\phi(\boldsymbol{k}) + b(\boldsymbol{k})\delta_{i3} = \sum_{\boldsymbol{p}+\boldsymbol{q}=\boldsymbol{k}} ip_j u_i(\boldsymbol{p})u_j(\boldsymbol{q}) + D_{u_i}, \tag{3.1a}$$

$$ik_i u_i(\boldsymbol{k}) = 0, \tag{3.1b}$$

$$\frac{\partial b(\boldsymbol{k})}{\partial t} + N^2 u_j(\boldsymbol{k})\delta_{j3} = \sum_{\boldsymbol{p}+\boldsymbol{q}=\boldsymbol{k}} ip_j b(\boldsymbol{q})u_j(\boldsymbol{p}) + D_b, \tag{3.1c}$$

where u_i is the velocity field, b is the buoyancy fluctuation and ϕ is the pressure perturbation divided by the constant reference density, ρ_0. The operator D_q represents the dissipation of quantity q and the Coriolis parameter, f, and Brunt-Väisälä frequency, N, are constants. $\mathbf{k} = (k_1, k_2, k_3)$ denotes the wavevector. We take the boundary conditions as periodic, implying wave vectors have discrete components and are henceforth represented using vector indices.

As mentioned in Section 3.2, Leith (1980) presented a decomposition of state variables using the normal modes of the linearised hydrostatic Boussinesq equations, later generalised to the nonhydrostatic set by Bartello (1995). The appropriate variables are those that make the matrix of linear terms Hermitian, yielding real eigenvalues and orthogonal eigenvectors that can be normalised. Under these circumstances (3.1) admits three eigenfrequencies

$$\lambda_{\mathbf{k}}^{(0)} = 0, \quad \lambda_{\mathbf{k}}^{\pm} = \pm\sigma_{\mathbf{k}} = \frac{(f^2 {k_z}^2 + N^2 {k_h}^2)^{1/2}}{k}, \tag{3.2}$$

where horizontal and vertical wavenumbers are defined as $k_h = (k_1^2 + k_2^2)^{1/2}$ and $k_z = k_3$, respectively. These frequencies indicate that there are two time scales in the linearised Boussinesq equations. The normal modes can then be derived by projecting the state variables onto the eigenvectors associated with $\lambda^{(0)}$ and $\lambda^{(\pm)}$. We denote the modes corresponding to $\lambda^{(0)}$ by $G_{\mathbf{k}}$, known as geostrophic or rotational modes. They show slow nonlinear variability when f and N are large and contribute to the linear potential vorticity. The other modes, denoted by $A_{\mathbf{k}}^{(\pm)}$, have a smaller time scale since they correspond to larger frequencies, $\lambda^{(\pm)}$. They are referred to as ageostrophic or gravitational modes and have no linear PV. For notational economy we drop the superscript $(\pm)$ for the ageostrophic modes, but it should be kept in mind that there are two of opposite sign for each wavevector.

In terms of the normal-mode equations (3.1) takes the following form in a triply-periodic domain

$$\frac{dG_{\mathbf{k}}}{dt} = \gamma_{\mathbf{k}}(G, A) + D_G \tag{3.3a}$$

$$\frac{dA_{\mathbf{k}}}{dt} + i\sigma_{\mathbf{k}} A_{\mathbf{k}} = \psi_{\mathbf{k}}(G, A) + D_A, \tag{3.3b}$$

where γ and ψ are quadratic nonlinear convolution terms.

3.3.2 Separation of time scales

Based on (3.2), there are two time scales in the flow when the nonlinear terms are small compared to the linear ones. The time scale of the set of G modes is $T \sim L/U$, which is determined by the nonlinear terms since $\lambda_{\mathbf{k}}^{(0)} = 0$, and the time scale of the set of A modes are the inverses of their eigenfrequencies, $t^* = 1/\sigma_{\mathbf{k}}$. Therefore, the ratio of the two is

$$\frac{t^*}{T} = \frac{1/\sigma}{L/U} \leq \frac{U}{fL} = \varepsilon = Ro, \tag{3.4}$$

This separation enables us to split the time derivative into two parts at small Ro

$$\frac{d()}{dt} = \frac{d()}{dt^*} + \varepsilon \frac{d()}{dT} \tag{3.5}$$

The nonlinear terms are then $O(Ro)$ smaller than the linear terms. Therefore, equation (3.3) can be written in non-dimensional form as

$$\varepsilon \frac{dG_{\mathbf{k}}}{dT} = \varepsilon \Gamma(G, A) + D_G \tag{3.6a}$$

$$\frac{dA_{\mathbf{k}}}{dt^*} + \varepsilon \frac{dA_{\mathbf{k}}}{dT} \pm i\sigma A_{\mathbf{k}} = \varepsilon \Psi(G, A) + D_A. \tag{3.6b}$$

It should also be noted that $dG/dt^* = 0$, as the geostrophic modes vary only on the slow nonlinear time scale.

3.3.3 High-order balance model

Leith (1980) and Lorenz (1980) put forward the concept of the slow manifold, which is strictly defined as an invariant manifold in phase space on which the dynamics are devoid of any high-frequency variability (Leith, 1980; Lorenz, 1980). This definition was relaxed in subsequent studies such as Warn & Menard (1986); Warn (1997); Lorenz (1992); Ford *et al.* (2000). Its existence, or essentially how invariant it is in the relaxed sense, would imply that the evolution of the fast modes (A) in time is approximately "slaved" to the slow modes, G, i.e.

$$A = F(G, \varepsilon), \tag{3.7}$$

referred to by Van Kampen (1985) and Warn *et al.* (1995) as the slave relation. Since the derivative of the geostrophic modes with respect to the fast time scale is zero, this relation filters out the natural high-frequency variation of the ageostrophic modes, thereby maintaining slowly-varying balance dynamics.

We employ a high-order normal mode initialization scheme developed by Baer & Tribbia (1977) to define balance, or equivalently, our working slave relation. In this scheme the ageostrophic terms in the scaled equation (3.6) are expanded in terms of ε. Then the derivative of ageostrophic modes with respect to the fast time scale is forced to zero at the given time, $dA/dt^*|_{t_0} = 0$. By using perturbation techniques different terms in the expansion of A can be derived in terms of G and ε up to any desired order. In this study, we expanded ageostrophic modes up to $O(\varepsilon^2)$,

which is admittedly only one order higher than quasigeostrophy. The details of this initialisation scheme and its numerical implementation in triply-periodic flows can be found in Baer & Tribbia (1977) and Kafiabad & Bartello (2016).

3.3.4 Balanced/unbalanced decomposition

The Baer-Tribbia scheme was formulated as an initialisation method for numerical weather prediction and can be used to derive the balanced part of any flow at any time. To accomplish this we first calculate the geostrophic modes, G, by normal-mode projection, explained in Section 3.3.1. Then, we use the method described in Section 3.3.3 to obtain the balanced ageostrophic modes, A^b, via this definition of the slave relation (3.7). Once the balanced ageostrophic modes are calculated, they can be subtracted from the simulation's total ageostrophic modes to obtain the unbalanced part

$$A^u_{\mathbf{k}}(t) = A_{\mathbf{k}}(t) - A^b_{\mathbf{k}}(t) \tag{3.8}$$

We denote the total flow by $X = B+U$, where each term is a vector of state variables, while noting that B is composed of both geostrophic and balanced ageostrophic modes.

It should be stressed that, while the normal-mode decomposition is in terms of orthogonal basis functions, our balance and unbalanced components are more complicated. If the total energy is given by $E = \frac{1}{2}\sum_{\mathbf{k}} X_{\mathbf{k}} \cdot X^*_{\mathbf{k}}$, then in addition to the usual balanced and unbalanced energy, $E^b = \frac{1}{2}\sum_{\mathbf{k}} B_{\mathbf{k}} \cdot B^*_{\mathbf{k}}$ and $E^u = \frac{1}{2}\sum_{\mathbf{k}} U_{\mathbf{k}} \cdot U^*_{\mathbf{k}}$, there are the cross terms $E^c = \frac{1}{2}\sum_{\mathbf{k}} U_{\mathbf{k}} \cdot B^*_{\mathbf{k}} + c.\ c.$ ($c.\ c.$ denoting the complex conjugate).

The horizontal energy spectrum and its components are obtained by summing over cylindrical shells in Fourier space in the usual way;

$$E(k_h) = \frac{1}{2} \sum_{|\boldsymbol{k'} - \boldsymbol{k'}\cdot\hat{\boldsymbol{z}}| = k_h} X_{\mathbf{k'}} . X^*_{\mathbf{k'}}, \tag{3.9a}$$

$$E^b(k_h) = \frac{1}{2} \sum_{|\boldsymbol{k'} - \boldsymbol{k'}\cdot\hat{\boldsymbol{z}}| = k_h} B_{\mathbf{k'}} . B^*_{\mathbf{k'}}, \tag{3.9b}$$

$$E^u(k_h) = \frac{1}{2} \sum_{|\boldsymbol{k'} - \boldsymbol{k'}\cdot\hat{\boldsymbol{z}}| = k_h} U_{\mathbf{k'}} . U^*_{\mathbf{k'}}, \tag{3.9c}$$

$$E^c(k_h) = \frac{1}{2} \sum_{|\boldsymbol{k'} - \boldsymbol{k'}\cdot\hat{\boldsymbol{z}}| = k_h} U_{\mathbf{k'}} . B^*_{\mathbf{k'}} + c.\ c.. \tag{3.9d}$$

Similar to the above, the vertical energy spectrum, $E(k_z)$, can be defined when the summation is carried over k_z.

3.4 Numerical configuration

In this section we briefly describe the numerical setting of our simulations. The periodicity of our boundary conditions enables us to employ the de-aliased pseudo-spectral method to solve the Boussinesq equations. Centred second-order finite differences are employed for time stepping. We set our domain to be $[0, 2\pi]^2 \times [0, 2\pi f/N]$, in which the vertical domain size is smaller than in the horizontal. Most of our simulations employ grids that are equally spaced in physical coordinates, i.e. $\Delta x = \Delta y = \Delta z$. However, we investigate how the results change if $\Delta x = \Delta y = (f/N)\Delta z$ in several simulations. All simulations in this paper are of decaying turbulence as we were unsure how to implement nonlinearly balanced forcing. The form

$$D_\xi = -\nu k^8 \xi, \tag{3.10}$$

Ro_u	$N_h \times N_z$	N	f	Ro_ω	Fr	τ	ν	Re
0.01	2048×256	339.53	42.44	0.140	0.050	6.1	5×10^{-20}	750
0.02	2048×256	169.77	21.22	0.288	0.112	6.1	5×10^{-20}	1220
0.03	2048×256	113.18	14.15	0.456	0.207	6.45	5×10^{-20}	1270
0.04	2048×256	84.88	10.61	0.661	0.358	7.04	5×10^{-20}	1410

Table 3–1: The values of the rms velocity-based Ro_u, rms vorticity-based Ro_ω and horizontal vorticity-based Fr in the Boussinesq runs are presented at $t = 0$. τ is the initial turn over time based on r.m.s. vorticity of the initial condition after Baer-Tribbia initialisation.

is employed for the dissipation, where ξ is one of the flow variables. ν is set equal for velocity and buoyancy.

To produce an initial condition for our Boussinesq simulations, we first run a decaying QG model starting from an initial energy distribution which peaks at $k = 20$ with random Fourier phases. The QG model is run until $t = 48.8$, which assures that the quasigeostrophic turbulence is fully developed. From this the geostrophic modal data are extracted and we use the Baer-Tribbia initialization scheme to find a set of balanced ageostrophic modes represented above by (3.7). These, together with the geostrophic modes, form our initial conditions for the nonhydrostatic Boussinesq simulation. Bearing in mind that the Baer-Tribbia initialization depends on Ro, separate initialisation is required for different Boussineq simulations with different initial Rossby numbers. We refer the reader to Kafiabad & Bartello (2016) for more detail on the procedure.

Starting from these we performed full Boussinesq simulations with the different parameters listed in Table 3–1. As can be seen, the ratio N/f is set to 8 in all our simulations, which is a substantial numerical compromise on exact atmospheric

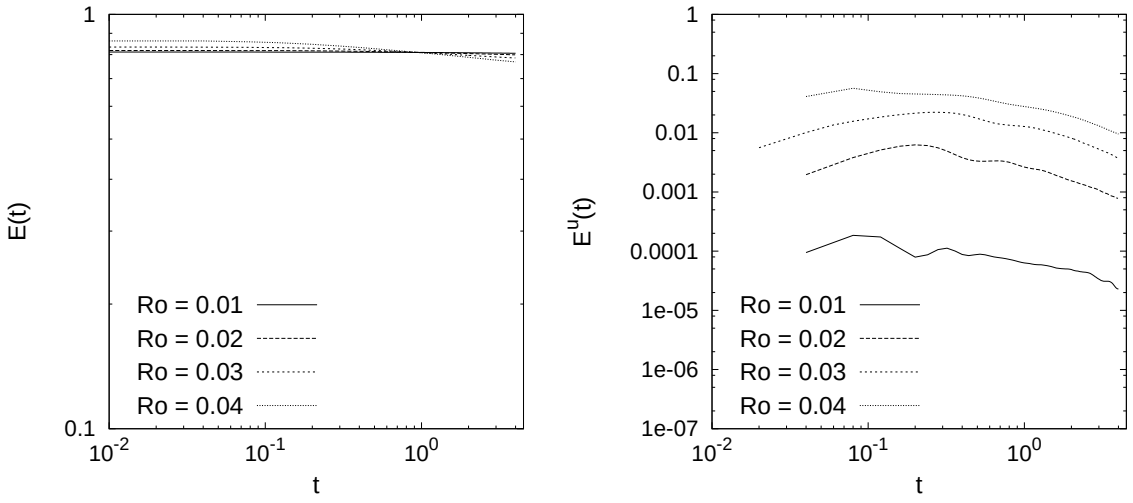


Figure 3–1: Time series of total energy, $E(t)$, (left) and unbalanced energy, $E^u(t)$, (right) at the four Rossby numbers shown in the legend.

or oceanic scaling. It was required primarily because of our desire to employ an isotropic grid while maintaining QG scaling, i.e., $L/H \sim N/f$, in the domain size. Again, more detail can be found in Bartello (2010) and Kafiabad & Bartello (2016).

We also presented the Reynolds number in Table 3–1 as a function the scale ratio between the integral scale, k_i, and the dissipation scale, k_η. For isotropic turbulence with a Laplacian of order one (regular viscosity), $Re \sim (k_\eta/k_i)^{4/3}$. We use this expression to derive Re; however, we derive k_η by computing the peak of the energy dissipation spectrum in our hyperdissipation simulations and k_i by computing the peak of the energy spectrum. The Reynolds thus obtained depends on the ratio of the inner to outer scale in the standard way, while its value is not unreasonably large and hence remains comparable to other simulations with Laplacian viscosity.

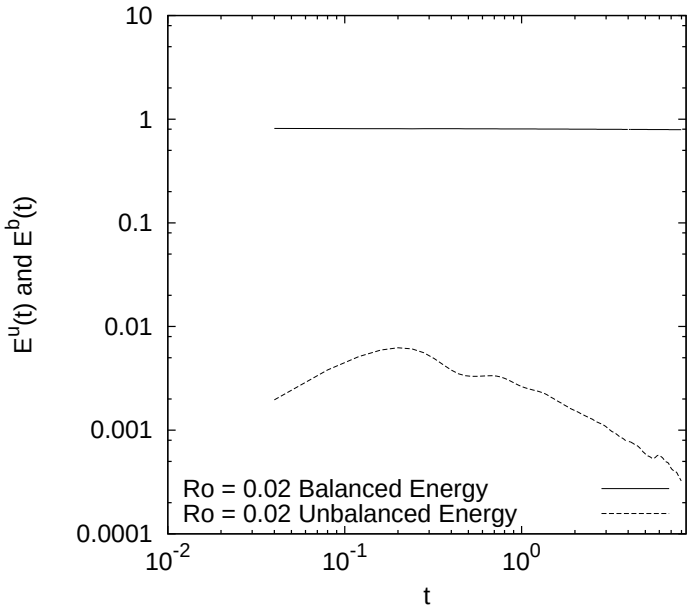


Figure 3–2: Time series of balanced energy, $E^b(t)$, and unbalanced energy, $E^u(t)$, for $Ro = 0.02$. The time is extended to $t = 8\tau$ in this figure.

3.5 Results

3.5.1 Energy time series

Figure 3–1 shows the total energy, which is the energy spectrum summed over all wavenumbers, as a function of time. The total energy (left panel) shows little decay, since the flow is predominantly quasigeostrophic, and increasingly so at lower Rossby numbers. For instance, at $Ro = 0.01$ the decay of energy is only 0.006 of the initial energy after 4τ. However, unbalanced energy is in all cases present. This may be significant or possibly an indication of the imprecision of our definition of balance. In any event, it increases at small times, reaches a maximum and then decays, suggesting that balance dynamics break down at least at some scales. As shown in figure 3–2, the late-time decay of unbalanced energy is faster than that

of balanced energy. Consequently, the flow once again approaches balance. Recall that in decaying high-Re QG flow, the Rossby number decreases with time as kinetic energy is not dissipated, f is constant and the integral scale grows.

As Ro decreases, the total energy decays less, signalling that in strongly rotating regimes the flow maintains its QG-like nature, where there is a forward cascade of potential enstrophy but no significant cascade of energy to the small-scale dissipation. On the other hand, as Ro increases the level of unbalanced energy increases as well (by orders of magnitude here), and the total energy decays more with time. When the rotation is decreased the maximum of E^u is also seen to move to later times with a higher amplitude, showing that the flow generates more unbalanced energy over a longer period. Below we will look at the energy and transfer spectra close to the time when the unbalanced energy reaches its maximum in order to examine how and at what scales the unbalanced energy is generated.

3.5.2 Energy spectra

Figure 3–3 portrays the horizontal energy spectra defined in (3.9) for three different Rossby numbers at later times near the end of the simulations. At the lowest Rossby number, $Ro = 0.01$, the total and balanced spectra are on top of one another. The unbalanced spectra is orders of magnitude smaller but has a shallower slope at large k_h. By extending the balanced and unbalanced inertial-range spectra to unresolved smaller scales, one can speculate that at higher Re they would eventually cross at some large wavenumber, above which unbalanced energy will predominate.

When the Ro is increased to 0.02, the balanced energy spectrum looks quite similar, but the unbalanced spectrum is more shallow. Indeed, the two can be seen to

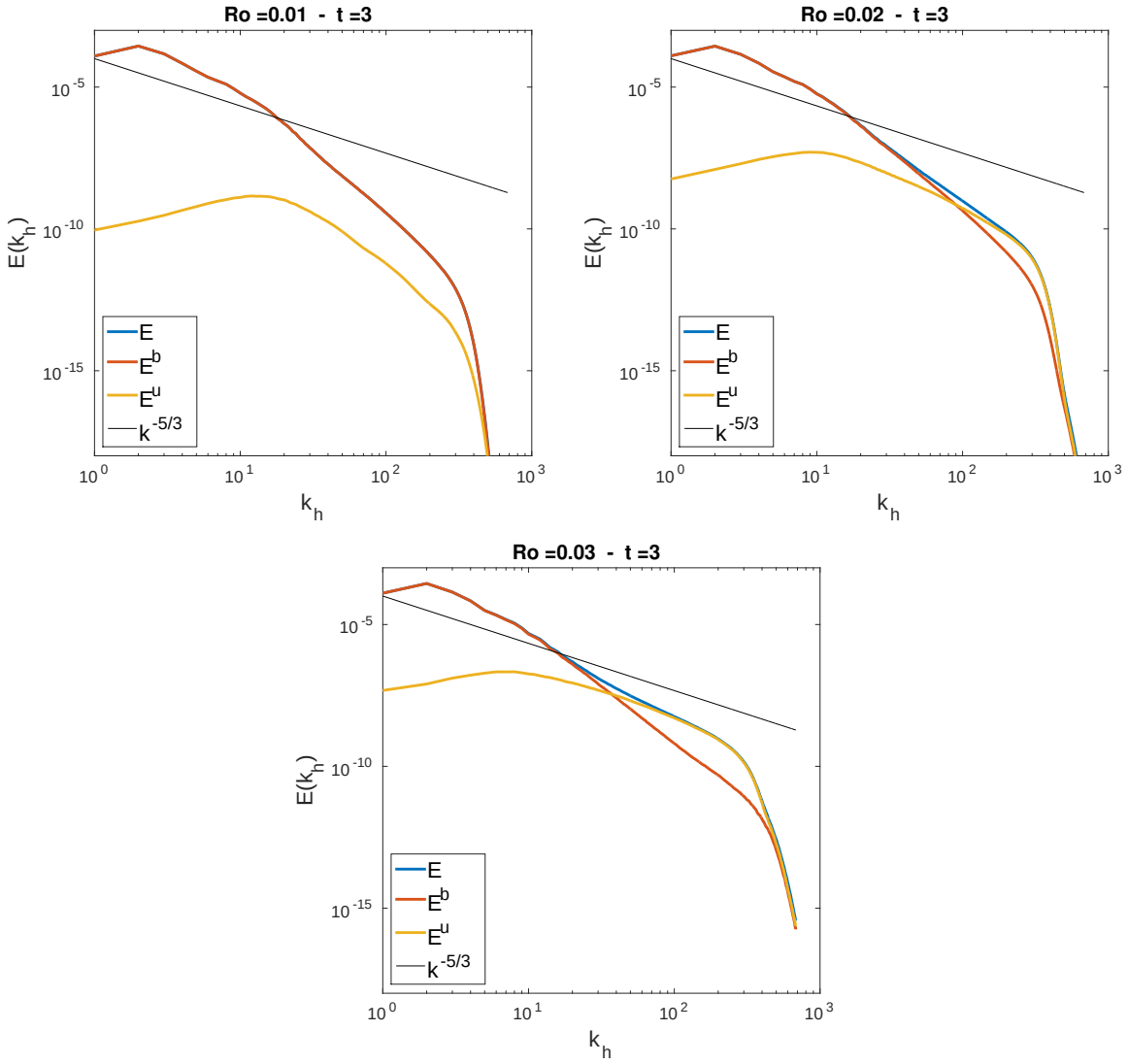


Figure 3–3: The horizontal energy spectra of total, balanced and unbalanced flows at $t = 3$ for three different Rossby numbers.

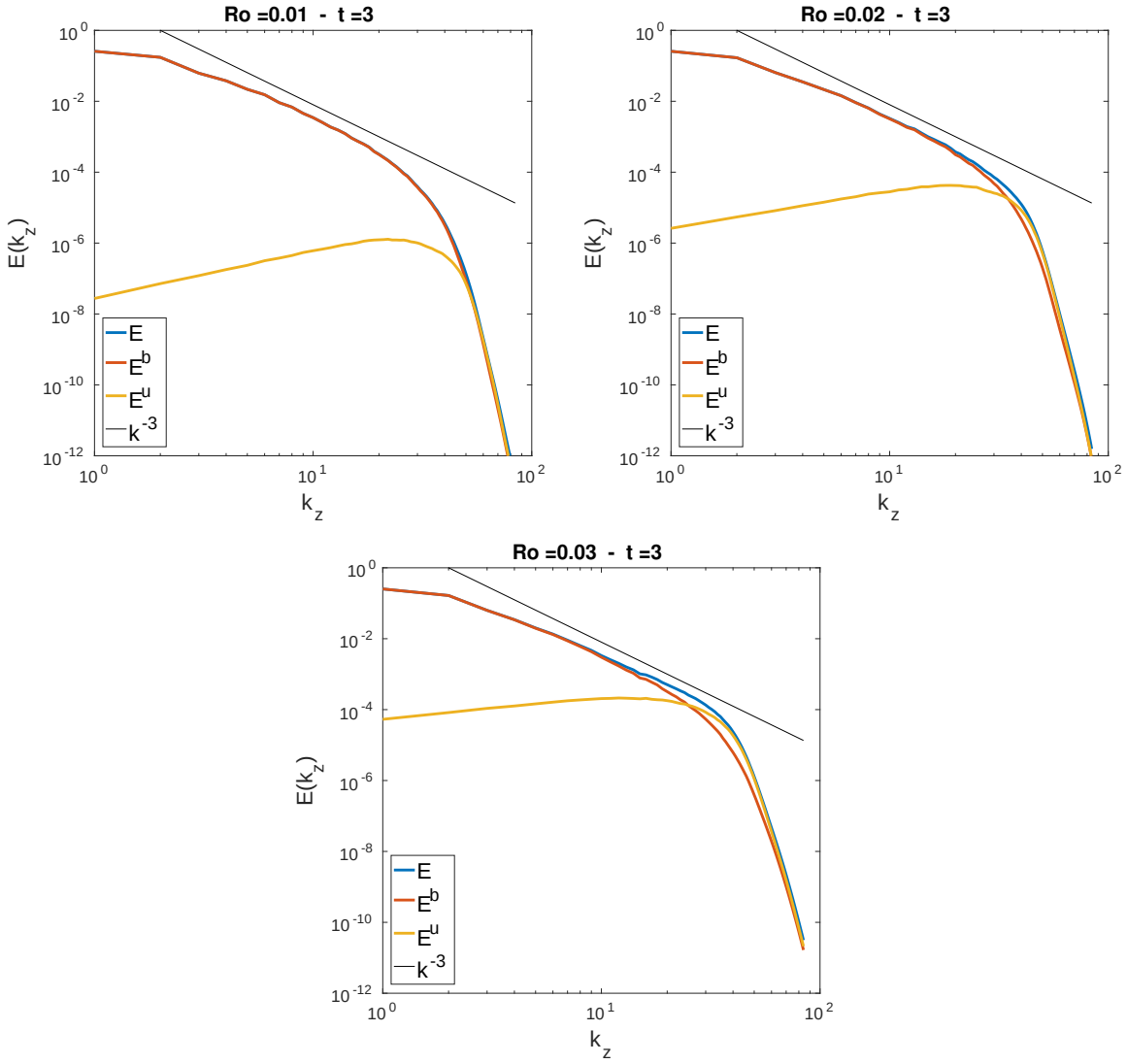


Figure 3–4: The vertical energy spectra of total, balanced and unbalanced flows at $t = 3$ for three different Rossby numbers.

cross near $k_h = 90$. Below this wavenumber the total energy spectrum is dominated by the balanced flow and above it the situation is reversed. Compared to the $Ro = 0.01$ curve, the unbalanced spectrum at $Ro = 0.02$ has a higher amplitude and a more shallow slope. At the highest Rossby number, $Ro = 0.03$, the unbalanced energy is even more energetic with an even more shallow slope close to $-5/3$. Consequently, the kink of the total energy spectrum moves to smaller wavenumbers, a fact that was observed by Nadiga (2014) as well. Kafiabad & Bartello (2016) found the spectral slope of unbalanced ageostrophic modes to be somewhat sensitive to the dissipation. A similar sensitivity was observed in the spectra of Marino *et al.* (2015). We therefore expect that at much larger Reynolds numbers the slopes of unbalanced energy would approach -5/3, but its amplitude would vary with Rossby number as displayed here. In any event, these simulations are felt to cover a range of Ro with balanced flow in the large scales and a transition to unbalanced small-scale dynamics in a way that is broadly consistent with atmospheric and oceanic data.

Finally, it is readily seen from the figure that the cross-term contribution is far smaller than that from the balanced flow at large scales and far smaller than that from the unbalanced flow at small scales, as the total spectrum is seen to be very close to either one or the other at almost all scales. In addition, it is also apparent that the wavenumber where balanced and unbalanced spectra cross is proportional to Ro^{-2} (see Fig. 3–5) over this rather limited range. Recall that the balanced ageostrophic modes must scale with Ro and therefore the balanced horizontal ageostrophic energy spectrum scales as Ro^2 (see Kafiabad & Bartello (2016)). In that study the

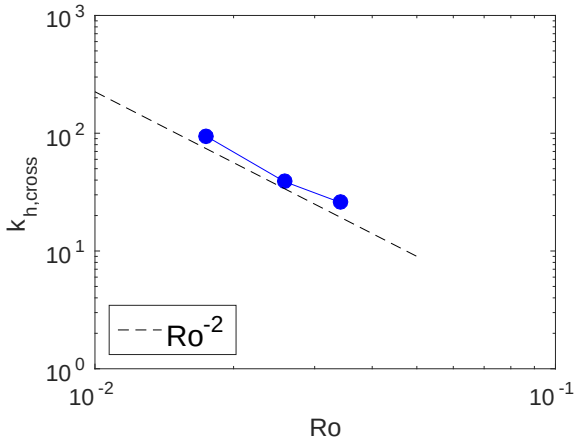


Figure 3–5: The wavenumber at which E^u and E^b cross, $k_{h,cross}$, as a function of instantaneous $Ro(t)$. A line proportional to Ro^{-2} is also shown as a guide.

total ageostrophic contribution to the horizontal energy spectrum was plotted after the generation of considerable unbalanced motion. When scaled by $Ro(t)^2$, the ageostrophic spectra collapsed onto the same steep curve at larger, balanced scales, but departed from this steep envelope at smaller scales, where the spectrum was much more shallow. In fact, that large-k_h shallow range approached a logarithmic slope of -5/3 at the larger Rossby numbers and the wavenumber at which it left the steep balanced envelope increased as the Rossby number was decreased.

In our view a reasonable simplification would be to represent the unbalanced spectrum by $E^u(k_h) \sim C_u Ro^{-\alpha} k^{-5/3}$, where C_u is an as yet unknown dimensional quantity encompassing all of the other unknowns, and the balanced spectrum by $E^b(k_h) \sim C_b k^{-\beta}$, which in the first instance is independent of the Rossby number, as expected if it is low enough. Then, the two spectra would cross at $k_{h,cross} \sim Ro^{\gamma}$, where $\gamma = 3\alpha/(3\beta - 5) \approx 2$, or $\alpha \approx 2(3\beta - 5)/3$. Therefore, given our balanced spectra have a logarithmic slope around $\beta \approx 4.3$ near $k_{h,cross}$, we can infer that the

unbalanced range is growing in amplitude at the rate of $Ro^{5.3}$. If we take instead the Charney (1971) potential-enstrophy cascade range with $\beta = 3$, we obtain unbalanced energy growth proportional to $Ro^{2.7}$. Both are unsurprisingly faster than the Ro^2 growth in the balanced ageostrophic energy, producing a breakdown of QG at larger Ro.

The vertical energy spectra are plotted in figure 3–4. Since we are resolving a smaller range of vertical wavenumbers, we do not see the flattening of the tail for the vertical spectra as clearly as we see it in the horizontal. However, we can still observe that the level of unbalanced energy increases with the Rossby number. At $Ro = 0.03$, we see a hint of a shallow tail in the small scales, and we speculate that it would extend to even smaller scales if the Reynolds number were increased. All the vertical spectra presented in figure 3–4 show a well resolved balanced inertial range with a slope of -3.

3.5.3 Energy transfers and energy fluxes

The evolution of the horizontal energy spectrum is governed by

$$\frac{\partial E(k_h,t)}{\partial t} = T(k_h,t) - 2\nu k_h^8 E(k_h,t), \tag{3.11}$$

where $T(k_h)$ represents the nonlinear transfer spectrum and is the result of triad interaction between three modes

$$T(k_h,t) = \sum_{|\boldsymbol{k'}-\boldsymbol{k'}\cdot\hat{\boldsymbol{z}}|=k_h} \quad \sum_{\boldsymbol{k'}=\boldsymbol{p}+\boldsymbol{q}} N_{\boldsymbol{pqk'}} X_{\boldsymbol{p}} X_{\boldsymbol{q}} X^*_{\boldsymbol{k'}} + c.c., \tag{3.12}$$

with $N_{\boldsymbol{pqk'}}$ as the interaction coefficient. We will refer to the triad interaction corresponding to the various terms of $T(k_h)$ as XXX. Keeping in mind that $X = B + U$, these can be divided into several sets of interactions.

Similarly, the evolution of $E^b(k_h)$ and $E^u(k_h)$ can be written as

$$\frac{\partial E^b(k_h,t)}{\partial t} = T^b(k_h,t) - 2\nu k_h^8 E^b(k_h,t), \tag{3.13a}$$

$$\frac{\partial E^u(k_h,t)}{\partial t} = T^u(k_h,t) - 2\nu k_h^8 E^u(k_h,t),. \tag{3.13b}$$

The terms in the sums of $T^b(k_h,t)$ and $T^u(k_h,t)$ are composed of only BBB and UUU interactions, respectively. Considering the definition of $E^c(k_h,t)$ in (3.9d), the subtraction of (3.13a) and (3.13b) from (3.11) gives

$$\frac{\partial E^c(kh,t)}{\partial t} = T^c(k_h,t) - 2\nu k_h^8 E^c(k_h,t), \tag{3.14}$$

where $T^c(k_h,t) = T(k_h,t) - T^b(k_h,t) - T^u(k_h,t)$. This term is the contribution to the nonlinear transfer due to the interaction between the balanced and unbalanced modes, via UBB and BUU triads.

Given the transfer functions, one normally integrates to obtain energy fluxes. In our discrete numerical work this was done as

$$\Pi(k_h) = -\sum_{k_h'<k_h} T(k_h') \tag{3.15}$$

By replacing $T(k_h')$ with $T^b(k_h')$, $T^u(k_h')$ and $T^c(k_h')$, corresponding fluxes dentoed by $\Pi^b(k_h')$, $\Pi^u(k_h')$ and $\Pi^c(k_h')$ are obtained. The transfer spectra of total, balanced, unbalanced flows as well as the cross-term contribution are depicted in Fig. 3–6.

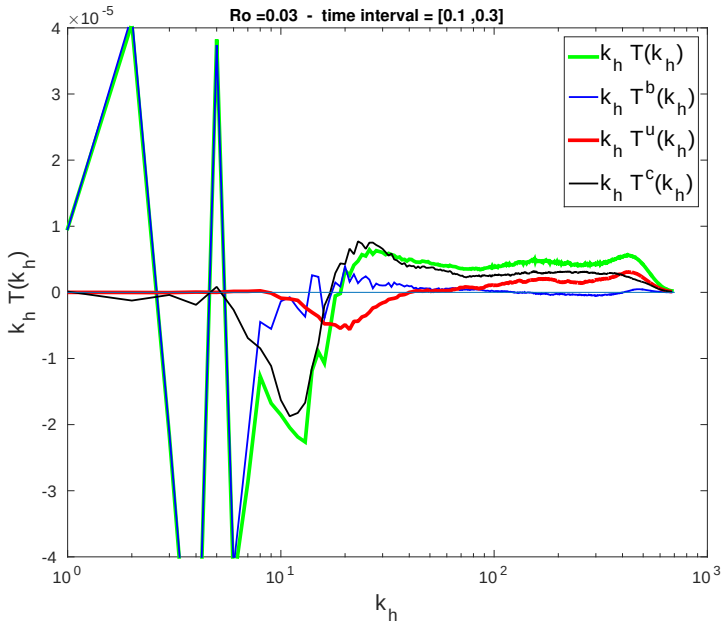


Figure 3–6: The transfer spectra for different contributions to the energy (shown in the legend) at $Ro = 0.03$. The curves are averaged over the time interval of $[0.1, 0.3]$, which is around the time of maximum unbalanced energy.

These curves are averaged over $t \in [0.1, 0.3]$ which is around the time of maximum unbalanced energy shown in Fig. 3–1. The Rossby number we consider in detail is 0.03, at which E^u and E^b cross at $k_h \approx 30$ at the time considered. Note that the transfer spectrum of the balanced energy (blue curve) is very close to zero above the wavenumber where the two spectra cross ($k_h = 30$). Knowing that the balanced flow is predominantly quasigeostrophic, this follows the Charney (1971) phenomenology in that there is no significant forward cascade of balanced energy in the potential enstrophy cascade range. However, $T^b(k_h)$ is fairly large and negative at the largest scales, indicating an upscale cascade by these interactions.

Proceeding down scale from the energy peak, the balanced part of the transfer (blue) ceases to dominate slightly above $k_h \approx 10$, where the total transfer (green curve) follows the contribution from the cross terms (black curve) quite closely. These terms are removing energy at this scale and pumping it into smaller scales with a peak occurring near $k_h \approx 30$, which is the crossover wave number. At larger wave numbers a wide flat positive range is produced by these cross terms. The reader is reminded that only these triads can effect transfer from balanced to unbalanced components. Finally, the interactions between unbalanced modes (red curve) remove energy from the negative peak around $k_h \approx 20$ and move it to higher wavenumbers over the broad range above $k_h \approx 70$, producing a forward flux towards the dissipation. Recall that in a true inertial range the transfer is zero and the flux is constant. Clearly, we have an insufficient scale range to achieve this here, but the curves do approach smaller values at larger k_h.

It is tempting to draw the analogy between these results and those of Bartello (1995) who found that a catalytic interaction between two ageostrophic modes and the geostrophic mode was primarily responsible for the down scale cascade of fast-mode energy at low Ro and low Fr. Unlike that study, here the ageostrophic modes are decomposed to a balanced and an unbalanced part. The balanced ageostrophic modes in general do not cascade downscale, as they are slaved to the geostrophic modes following 3.7. However, in certain conditions the unbalanced ageostrophic energy can cascade forward in a scenario similar to the catalytic interaction described in Bartello (1995). In that study the initial conditions were completely unbalanced giving rise to two time scales in the resulting solution. The argument rested on both the predominance of resonant interactions and the fact that only the geostrophic mode contributed to the linear potential vorticity. Here we have tried to keep the initial conditions as close as possible to the (at least approximate) slow manifold, so weighting the importance of triads based on resonance seems ill advised. Still, the generation of unbalanced energy must occur in the form of higher frequencies at low enough Ro and Fr (explored more fully below). If the the two-time scale approach retains some validity, the catalytic interaction is included in the present decomposition as part of the set of BUU triads. On the other hand, since the balanced motion is heavily dominated by the geostrophic normal modes as $Ro \to 0$ and the PV is increasingly linear in this limit, it is still expected that the balanced motion contains virtually all of the PV.

The preceding argument can also be examined by studying the fluxes of energy in Fig. 3–7. The total flux of energy is negative at low $k_h \leq 6.5$ and consists almost

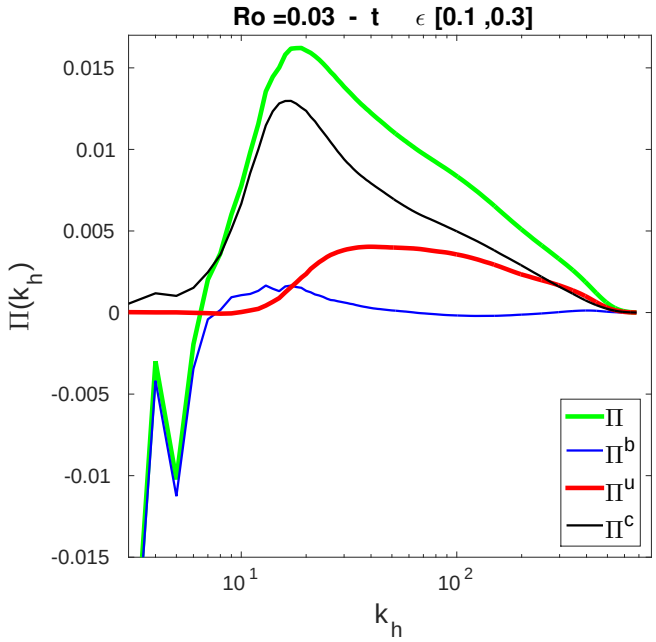


Figure 3–7: The fluxes of energy at $Ro = 0.03$. The curves are averaged over the time interval of $[0.1, 0.3]$ near the time of maximum unbalanced energy.

entirely of balanced energy in an inverse cascade. More importantly to our goal here, above $k_h = 6.5$, Π^b makes a contribution to the total flux that is far smaller than the other terms. At the wave number where the total flux changes signs, $\Pi^u \approx 0$ and the largest contribution to the downscale flux comes from Π^c, which is due to the interaction of the balanced and unbalanced modes. It reaches a peak near $k_h \approx 20$. At larger k_h, when enough energy has been transferred to the unbalanced motion, Π^u increases from zero, reaches a broad peak and enhances the forward cascade of energy. Since this graph is merely the negative integral of the curves of Fig. 3–6, it does not tell a different story, but the reader may benefit from seeing both representations.

We have examined the same transfer statistics at these early times from simulations at other initial Rossby numbers as well. They display no real surprises and are therefore not shown. At very low Ro the balanced transfer is large, negative and Ro-independent at scales larger than the energy-containing scale, as expected (Charney, 1971). At smaller scales all the other curves are negligibly different from zero. At larger Rossby numbers the terms discussed above, that are participating in the downward flux of energy, are all larger, increasingly so as the Rossby number increases.

At later times considerable unbalanced energy has been generated, at least at the larger Rossby numbers simulated. In this respect, from this point on these simulations have more in common with previous work such as Bartello (1995) starting from unbalanced initial conditions. Figure 3–8 shows the various flux terms for three different Rossby numbers well after the time of maximum unbalanced energy. Note that the balanced flux is approximately the same for all Ro and that the unbalanced

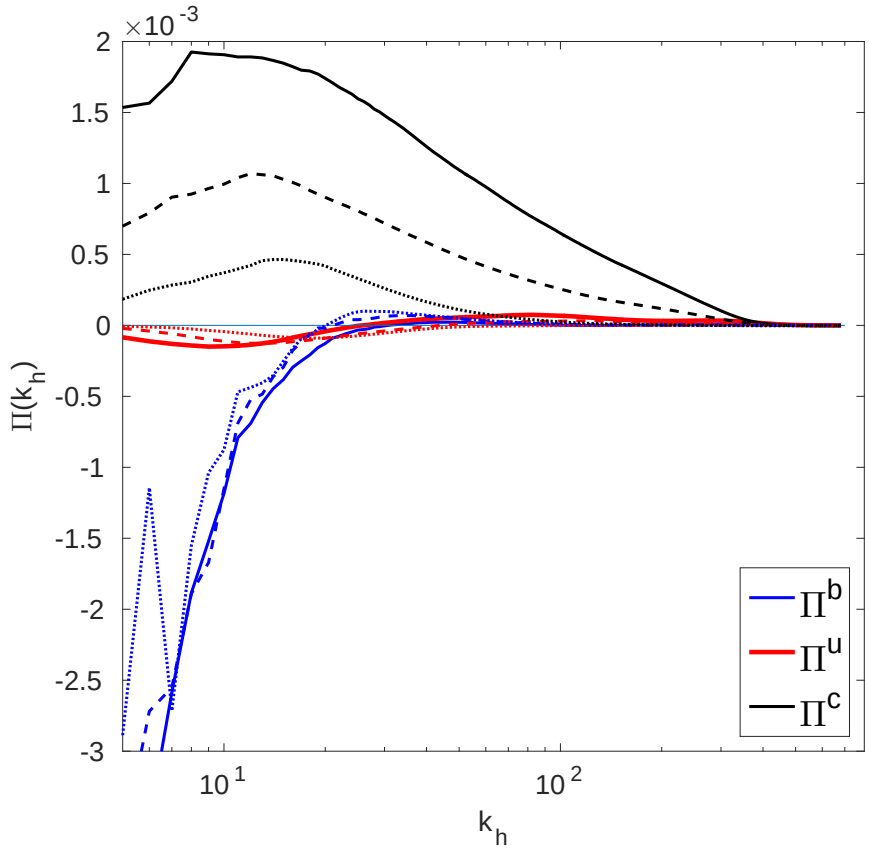


Figure 3–8: The flux spectra of balanced, unbalanced and cross terms at later times. The simulations for three Rossby numbers are shown; $Ro = 0.02$ (dotted line), $Ro = 0.03$ (dotted line) and $Ro = 0.04$ (dotted line).

contribution is quite small at all horizontal wave numbers. The predominant feature of this graph is clearly the cross-term contribution, involving interactions between balanced and unbalanced motion. It is large and positive and increases with the Rossby number as Ro^2. The downscale flux of energy grows to rival the upscale transfer from the balanced motion, implying a significant damping on the balanced motion.

3.5.4 Frequency spectra

Since the purpose of the nonlinear balance scheme applied to the initial conditions is to remove high-frequency motion, and since simulations show evolution away from this balance, we examine the degree to which departures from balance display large frequencies. This may seem paradoxical at first. If the Rossby and Froude numbers are small at all scales, then the unbalanced motion must, by definition, be high-frequency motion. However, we are dealing with a wide range of scales here and in the atmosphere and ocean. Since the nonlinear frequency of the turbulence increases at large wavenumbers, the small scales do not necessarily have low Rossby and Froude numbers. In the high-Re limit, we presumably have 3-D turbulence below the Ozmidov scale. While completely unbalanced, there is no significant wave motion.

To produce statistically reliable frequency spectra we used horizontal isotropy in that the complex Fourier coefficients of geostrophic and ageostrophic modal energies were saved for sixteen modes within circular rings defined by the values of $|k_z|$ and k_h. The Fourier transform of each of these time series was then taken and averaged over the sixteen modes to yield the curves in the figures. We rather arbitrarily chose the barotropic mode, but other modes at low $|k_z|$ were qualitatively similar.

At $k_h = 8$, which is in the range of scales where Ro is small and the Baer-Tribbia scheme is expected to be effective, the geostrophic energy dominates, but a sharp peak in an otherwise red spectrum is observed between N and f in the ageostrophic energy, which clearly indicates some (very small) imbalance in the simulation at this time. Proceeding down scale to $k_h = 20$, which is near the horizontal scale at

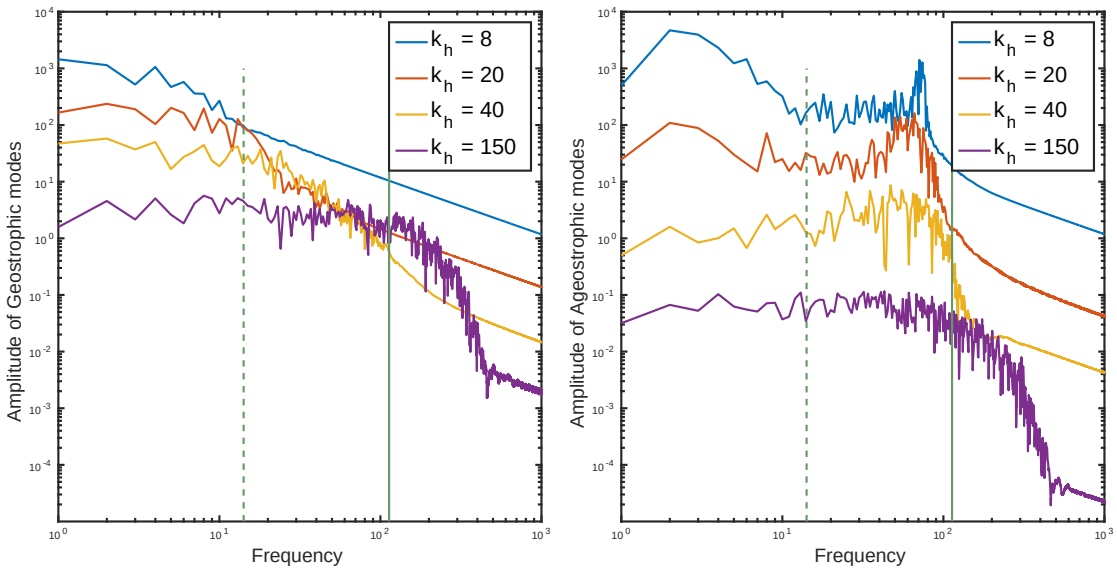


Figure 3–9: The frequency spectra of geostrophic (left) and ageostrophic (right) modes. The simulation is run at $Ro = 0.03$, and 16000 points are sampled from $t = 0$ to $t = 4\tau$. The dashed and the solid vertical lines represent f and N, respectively. For clarity the ageostrophic spectra of $k_h = 40$, $k_h = 20$ and $k_h = 8$ are vertically offset by one, two and three decades, respectively.

which the balanced and unbalanced energy spectra cross, we see that the peak in the ageostrophic frequency spectrum is less sharp, but still occurs between N and f. At even smaller scales (see $k_h = 40$) within the shallow range of the energy spectrum, this is still the case, but the peak continues to broaden, becoming rather flat at the small-scale end of our simulated range ($k_h = 150$). Indeed, at these scales the slow geostrophic modes display a frequency spectrum which parallels that of the supposedly fast ageostrophic modes, indicating there is no time scale disparity at all in this range.

From these results and the analysis of the transfer above, it seems reasonable to conclude that the generation of unbalanced motion is occurring at the large-scale end of the shallow spectral range. At these scales the unbalanced degrees of freedom do have a smaller time scale than the balanced flow and so the catalytic interaction may be more efficient than other interactions. Previous work suggests that if these triads were acting alone at low Ro and Fr, they would yield a shallow spectrum with very little energy transfer with the geostrophic modes, a result that is entirely consistent with our transfer spectra. However, we are hardly in the asymptotic regime since we increased Ro and Fr until a shallow spectral tail emerged in our simulations. The frequency spectra indicate that the time scale disparity between geostrophic and ageostrophic modes gradually disappears as one proceeds to smaller scales in the shallow range. The image is therefore not of fast waves moving between slow vortices, but rather of a more general type of turbulence in which there is a single, although broad, distribution of time scales.

3.5.5 The effect of vertical resolution

To conclude the results section we briefly investigate how the ratio of $\Delta x/\Delta z$ can affect the results, keeping the horizontal grid the same, i.e. $\Delta x = \Delta y$. As mentioned in Kafiabad & Bartello (2016), historically two choices of grid have been considered; grid 1) $\Delta x/\Delta z = N/f$, and grid 2) $\Delta x/\Delta z = 1$, used in the preeceeding simulations. Grid 1) has an equal spacing in the Charney stretched coordinate (in which our physical domain becomes a cube of $[0, 2\pi]^3$), whereas on grid 2) the grid points are equally spaced in physical space. We discussed in Kafiabad & Bartello (2016) that QG simulations have similar spectra at the resolutions used; nevertheless, grid 2) cannot capture the details of QG dynamics as accurately as grid 1). We argue that the higher vertical resolution of QG simulations does not affect the balance dynamics in our problem, since we consider the statistics of geostrophic modes mostly at large scales. On the other hand, grid 2) provides unbiased numerics for the dynamics in the small-scale transition range, which are far from geostrophic. In hindsight, the results of our Boussinesq simulations show that the small scale dynamics appear to be close to isotropy after the breakdown of balance, which further justifies the choice of grid 2). This is shown in figure 3–10. At $Ro = 0.01$ in this figure, the vertical integral scale is indeed much smaller than the horizontal scale; however at the higher Rossby numbers, at which balance breaks down, the turbulent regions have similar vertical and horizontal scales. Another advantage of grid 2) is allowing a wider range of horizontal scales with fixed computational resources.

In this paper, we present simulations with both grids using the full Boussinesq equations. We performed two sets of simulations at the resolutions 1024^3 and $1024^2 \times$

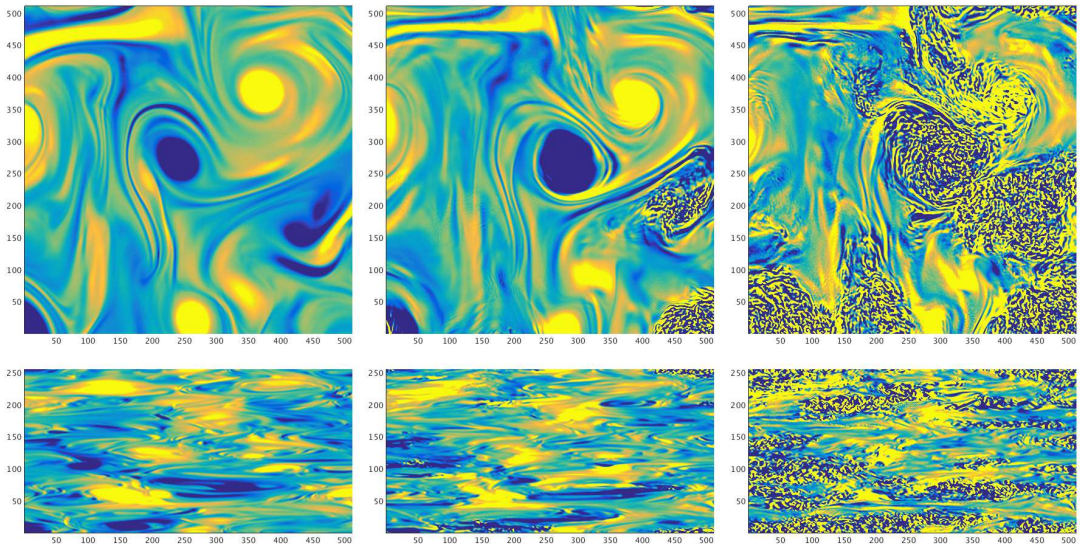


Figure 3–10: The vertical vorticity field after 2τ for different Rossby numbers: $Ro = 0.01$ (left), $Ro = 0.05$ (middle), and $Ro = 0.09$ (right). Top plots are horizontal slices normal to the z-axis and the bottom ones are vertical slices at the same scale. Note that only part of the domain is depicted at the resolution $2048^2 \times 256$.

128, using grid 1) and grid 2), respectively. The latter is a low-resolution equivalent of the other simulations in this paper in that $N/f = 8$ and $\Delta x = \Delta z$. Figure 3–11 shows the spectra of simulations employing both grids 1) and 2) at $Ro = 0.03$. At this relatively early time, Grid 2) does not develop a shallow range at the resolved scales, whereas grid 1) admits a shallow range in vertical and horizontal spectra. In our view this suggests that our grid does not generate spurious imbalance. On the contrary, it shows less tendency to become unbalanced. When the vertical resolution is increased, more small scales are resolved. This is where unbalanced dynamics may take place. The vertical spectra of figure 3–4 are also consistent with this explanation. At the highest Rossby number shown in figure 3–4, a hint of a shallow tail is seen in the very small scales. If grid 1) is used at the same Rossby number, the shallow

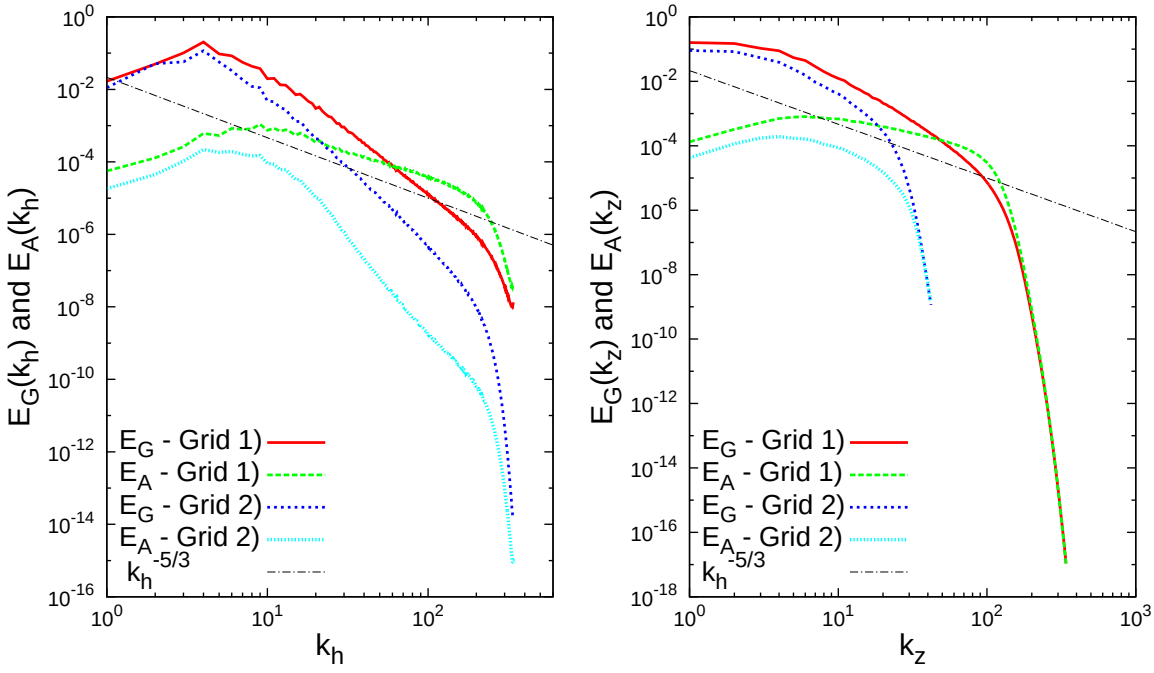


Figure 3–11: Horizontal (left) and vertical (right) spectra of total energy for Boussinesq simulations at $R = 0.03$. The spectra were sampled after one eddy turnover time.

range in the vertical energy spectrum would be better resolved, thereby influencing the horizontal spectrum. On the other hand, there is also the possibility that grid 1) falsifies the results by forcing non-isotropic features in the small-scale dynamics. With the same Rossby number, if more scales are resolved and the simulation is run for a longer time, eventually grid 2) also develops a shallow range in the ageostrophic spectrum, as was shown in figure 3–3.

One may argue that grid 2) is incapable of resolving QG structures with aspect ratio f/N. It remains to be seen whether QG dynamics consists of a hierarchy of structures at all scales, each with aspect ratio f/N. Although it is clear the invariants are isotropic in Charney coordinates, the QG equations are not since the advection term only contains horizontal velocity components. Therefore, the Charney scaling holds for the integral scale, but not necessarily at all smaller scales down to the dissipation range. Hence, we feel grid 2) can properly resolve large-scale QG structures if it has enough vertical resolution.

It is noteworthy that there are studies showing that coarse vertical resolution can lead to spurious wave generation. For instance, Snyder *et al.* (1993) observed that by changing $\Delta x/\Delta z$ from 100 to 16, more gravity waves are produced in frontogenesis. If these studies are compared to ours, two points need to be kept in mind. First, Snyder *et al.* (1993) solved a two dimensional problem assuming that the frontal flow is independent of the alongfront coordinate. Our turbulence simulations on the other hand, are three dimensional. Second, we are studying homogeneous turbulence at small scales without vertical boundaries unlike Snyder *et al.* (1993), where the initial condition was not statistically homogeneous in the vertical.

3.6 Conclusions

We have performed simulations of decaying rotating stratified turbulence in which the initial conditions have been formulated to keep the solution free from high-frequency motion, at least in an approximate sense. Several values of the initial Rossby and Froude numbers were simulated that yield a shallow small-scale tail in the energy spectrum, as in previous studies. We have examined simulation data in order to determine the extent to which Leith's work on balance and the slow manifold holds.

In Kafiabad & Bartello (2016) it was shown that balanced flow has a very steep spectrum and that the shallow -5/3 range that emerges at later times if Ro and Fr are sufficiently large, was entirely dominated by unbalanced motion. These results were extended here by examining the energy transfer, dividing the flow into balanced and unbalanced components using the Baer-Tribbia scheme. It was found that the interaction converting balanced to unbalanced energy occurred at relatively large scales near the low-wavenumber end of the shallow range in the energy spectrum. Examination of the frequency spectra of the Fourier amplitudes characterising this scale showed a broad peak in the range covered by linear wave frequencies. As such there is still some validity to the two-timescale approach, which is at the heart of Leith's work, although our simulations are far from the asymptotic limit.

The two distinct time scales were seen to merge into a single broad, flat peak in the frequency spectrum when small-scale Fourier amplitudes were considered. Here, the entire basis of the classical analysis is undermined and the notion of co-existing waves and turbulence is seen to be false. To the authors this implies that the breadth

of the approximately invariant slow manifold increases to encompass the whole of phase space and the concept is not useful on these scales.

Since the exchange and consequent generation of unbalanced energy peaks at relatively large scales. i.e., larger than the wavenumber characterising the shallow spectral zone, there is hope it may be possible to parameterise the drag on the balanced modes due to spontaneous imbalance. Since at least some time-scale separation was observed at these horizontal length scales, higher-order theoretical approaches may prove successful. In this light more work on vertical variability would seem to be called for.

Acknowledgements

This work would not have been possible without the financial support of the Natural Sciences & Engineering Research Council. The computational resources were generously provided by Compute Canada/Calcul Québec.

CHAPTER 4

Spontaneous imbalance in the non-hydrostatic Boussinesq equations

In the previous chapters, we produced a high-order balanced initial condition. We then studied its evolution in the non-hydrostatic Boussinesq equations. By implementing the balance initialization scheme after time integration, we decomposed the flow into a balanced and an unbalanced part. This decomposition enabled us to investigate the characteristics and the statistics of the balanced and the unbalanced modes separately.

The mere balance-imbalance decomposition cannot measure the nonlinear interactions between the decomposed components if the balanced and unbalanced modes are not orthogonal. In this chapter we lay out a new method to calculate these interactions in a flow that is predominantly balanced. More particularly, we quantify spontaneous imbalance generation, which is the energy transfer between the balanced and unbalanced modes. In addition to the quantification of imbalance generation, we present a comprehensive investigation of characteristic timescales at different spatial scales.

This chapter is based on the following paper:

Kafiabad, H. A., & Bartello, P. (2017). Spontaneous generation of imbalance in the non-hydrostatic Boussinesq equations . Submitted to Journal of Fluid Mechanics.

4.1 Abstract

Whereas high-frequency waves are valid solutions to the Boussinesq equations in certain limits, their amplitudes are generally observed to be small in large-scale atmospheric and oceanic data. Traditionally, this has led to the development of balance models, reducing the dynamics to only the slow subset. Their solutions, however, can spontaneously generate imbalance in the context of the full equations. To quantify this, we calculate how much energy is transferred from the balanced to the unbalanced part of a turbulent rotating stratified flow that has been initialised to remove high frequencies. We lay out an approach to derive the time evolution of the balanced modes in which their interactions with unbalanced modes are taken into account. This enables us to calculate the budget of balanced (and unbalanced) energy. Our results show that imbalance generation occurs at scales where the Froude and Rossby numbers are still small and the energy spectrum is steep. We find that the scale at which maximum imbalance is generated depends on the Rossby radius of deformation and is invariant to the strength of rotation over the range examined. The unbalanced energy, after being transferred from the balanced component of the flow at larger scales, is cascaded forward and forms a shallow energy spectrum. The steep balanced subrange of the energy spectrum and the shallow subrange cross and form a kink in the total energy spectrum consistent with observed atmospheric and oceanic data. A frequency analysis at different wavenumbers shows the separation of time scales breaks down at wavenumbers larger than those of maximum imbalance generation, but smaller than the kink of the energy spectrum. At these scales there is a single turbulent distribution of frequencies.

4.2 Introduction

The large-scale dynamics of the atmosphere and ocean are dominated by motion that is slow compared to a day. These dynamics are nearly geostrophic and referred to as balance dynamics, where dominant forces such as the Coriolis and pressure gradient forces approximately balance each other to the extent that high-frequency motion is not generated. Depending on the problem as well as the level of desired simplicity and accuracy, various balance models such as quasigeostrophy have been offered to reduce the dynamics to a slow subset (see McIntyre & Norton, 2000; Mohebalhojeh & Dritschel, 2001, as examples of balance models and their hierarchy). Concurrent with balance dynamics, there are fast components in these flows such as inertia-gravity waves (IGWs), which are observed to have relatively small amplitudes. These components may be referred to as imbalance. Inspired by the observed dominance of balance dynamics, initialisation schemes (such as Machenhauer, 1977; Baer & Tribbia, 1977; Leith, 1980; Kreiss, 1979, 1980; Van Kampen, 1985) produce initial conditions (ICs) that to varying degrees limit the initial growth or amplitude of the fast modes. The existence of a small imbalance in predominantly balanced flows raises the question of whether the balanced modes are affected by the unbalanced modes as time evolves. The degree to which balance dynamics remains robust depends on the parameters of the flow. In geophysical fluids, the two important dimensionless parameters are the Rossby and Froude numbers, defined as

$$Ro = \frac{U}{fL}, \quad Fr = \frac{U}{NH}, \tag{4.1}$$

in which U is the characteristic velocity and L and H are the characteristic horizontal and vertical length scales, respectively, while f is the Coriolis parameter and N is the stratification frequency. Therefore, Ro and Fr indicate the strength of rotation and stratification, respectively.

Two classes of studies have previously shown that balance dynamics breaks down. The first group considered low-order dynamical systems, which were often derived by modal truncation of the governing partial differential equations (PDEs). Some of these demonstrated that perturbation methods, which reduce the dimensionality of the dynamics, are not convergent, while others concluded that balance relations do not admit an exact solution (see Lorenz, 1986; Warn, 1997; Warn & Menard, 1986; Vautard & Legras, 1986; Lorenz & Krishnamurthy, 1987). In particular, Warn & Menard (1986) argued that any non-steady solution of their balance model generates some level of IGW activities. Because of these unbalanced IGWs, the solution lies within a "fuzzy" stochastic layer, the thickness of which decreases with Rossby number. Vanneste & Yavneh (2004) studied a low-order model derived from the Boussinesq equations and showed that in their model exponentially small IGWs were generated. They also offered an asymptotic estimate for the amplitude of these waves. Their work shows slow manifolds can be defined that are invariant to exponential accuracy, but not beyond.(for a comprehensive review of spontaneous generation see Vanneste, 2013).

The other group of studies considered the full PDEs to study the breakdown of balance. Ford *et al.* (2000) showed that, in the regime of low Froude number (Fr) and Rossby numbers of order unity, balance breaks down by generating IGWs of

$O(Fr^2)$. In a comment Saujani & Shepherd (2002) pointed out that this regime is different from quasigeostrophic balance, where both the Rossby and Froude numbers are small. Recently, Kafiabad & Bartello (2016) studied the breakdown of balance from a turbulence perspective. They showed that if rotation and viscosity are weak enough, a high-order balanced IC develops a shallow tail in its energy spectrum, which signals the breakdown of balance.

The general conclusion from the preceding studies is that reduced balance models cannot self-sufficiently describe the entire dynamics. Nevertheless, their interaction with the unbalanced part of the flow is relatively small such that it makes them good approximations at sufficiently small Froude and Rossby numbers. The main goal of this study is to quantify the interaction between the balanced and unbalanced motion when the flow is balanced initially. To achieve this we show how the unbalanced energy scales with the balanced energy and at what scales the transfer between balance and imbalance takes place. We also demonstrate how spontaneous imbalance, which is the energy transfer from balance to imbalance, is affected by changing the flow parameters such as rotation and stratification.

The quantification of imbalance generation can also shed light on the mechanisms that transition the large-scale balanced dynamics to the unbalanced mesoscale dynamics in the atmosphere. This transition was observed in atmospheric data by Nastrom & Gage (1985) and Lindborg (1999). According to these observations the atmospheric energy spectrum has a slope of -3 at synoptic scales. This power-law scaling is explained by the theory of geostrophic turbulence (Charney, 1971). However, at the mesoscales the energy spectrum develops a shallow tail with a slope

of -5/3. The steep part of the spectrum at large scales can be viewed as balanced and much evidence implies that the shallow tail is unbalanced (e.g. see Molemaker *et al.*, 2010; Bartello, 2010; Nadiga, 2014; Lindborg, 2015; Kafiabad & Bartello, 2016). Many turbulence studies have been performed to understand the physics behind this transition. Some, such as Lindborg (2005) and Waite & Bartello (2006*b*), focused on strongly stratified turbulence. For instance, Waite & Bartello (2006*b*) started from weak rotation with a forward cascade of energy and then decreased Ro until an inverse energy cascade appeared. By contrast, more recent rotating stratified turbulence studies have focused on small Ro with no significant forward cascade of energy and by weakening the rotation investigated the onset of the forward cascade (Bartello, 2010; Vallgren *et al.*, 2011; Deusebio *et al.*, 2013; Marino *et al.*, 2013; Pouquet & Marino, 2013). The latter approach seems to be more relevant to the development of the shallow mesoscale spectrum. Some of these studies such as Marino *et al.* (2013), Pouquet & Marino (2013) and Marino *et al.* (2015) used random forcing in their simulation, whereas others such as Bartello (2010), Vallgren *et al.* (2011) and Deusebio *et al.* (2013) only forced geostrophic modes. Since geostrophic forcing is linearly balanced, it can more realistically model the large-scale dynamics. Taking balance to higher orders, Kafiabad & Bartello (2016) initialized the flow with nonlinearly balanced conditions that included both geostrophic and ageostrophic parts. Unlike linear balance that is void of ageostrophic energy, higher-order balance displayed a steep ageostrophic spectrum. Knowing that large-scale geophysical flows are predominantly balanced, one may speculate that the ageostrophic part of the

atmospheric and oceanic energy spectrum is steep as well. These results are consistent with recent high-resolution Global Climate Model (GCM) and Numerical Weather Prediction (NWP) simulations. In addition to a shallow energy spectrum at mesoscales, the trace of a steep large-scale ageostrophic spectrum can be seen in these studies. Among others, Hamilton *et al.* (2008) and Evans *et al.* (2013) presented the divergent part of the atmospheric spectrum. They observed that at large wavenumbers the divergent energy spectrum is shallow, whereas at small wavenumbers it is steep. Bearing in mind that in our simulations ageostrophic modes contain all of the divergence, a steep ageostrophic spectrum implies a steep divergent spectrum. Therefore, one may see a similarity between the results of Kafiabad & Bartello (2016) based on nonlinearly balanced ICs and high-resolution GCM/NWP studies. This similarity is reliable at larger scales, but it should be interpreted with caution at smaller scales since the vertical resolution of GCM/NWP simulations is usually not sufficient to produce reliable turbulence statistics (see the discussion in Waite & Snyder, 2009; Waite, 2016; Cullen, 2017). In this study, after comparing imbalance generation at different orders of balance, we conclude that many features of numerical studies with linear balance remain valid when nonlinearly balanced ICs are employed. Hence, we feel the qualitative results of previous studies that employed linear balance, such as Bartello (2010), Vallgren *et al.* (2011) and Deusebio *et al.* (2013), can reliably describe the transition from the synoptic to the mesoscale spectrum from a turbulence point of view.

This study, as well as most turbulence studies that have investigated the mesoscale energy spectrum, relies on periodic geometry to have statistical homogeneity. Knowing that atmospheric data were collected close to the tropopause, one may cast doubt on the relevance of triply periodic simulations. In other words, the internal boundary-induced mechanisms may play a more important role in generating imbalance than the fluid interior. A recent study by Asselin *et al.* (2017) investigated the comparative significance of these two mechanisms at different Rossby numbers. The authors preformed two sets of simulations: one with constant stratification and the other with a rapid change in the background stratification (which models the tropopause). Starting with QG initial conditions they found that at characteristic velocities around 1 $m.s^{-1}$ the former maintains a steep spectrum, while the latter develops a shallow spectrum with a slope of -5/3. This transition from a -3 to a -5/3 slope is consistent with surface quasigeostrophy (SQG), as suggested by Tulloch & Smith (2006). However, for more typical atmospheric Rossby numbers (characteristic velocities of 10 $m.s^{-1}$), the dynamics with and without a tropopause were reported to be similar. The results of this study highlight the importance of internal processes in the breakdown of balance and the shallowing of the mesoscale spectrum, which is the focus of the current work.

To reach our goal, which is the quantification of imbalance generation, we face several challenges. The equations that balance models offer for the budget of the balanced energy are devoid of any interactions with the unbalanced modes. Therefore, they cannot be used to derive spontaneous imbalance, which is the transfer from

balance to imbalance. On the other hand, more complete systems such as the primitive or Boussinesq equations contain the convoluted evolution of all modes together, where balance and imbalance dynamics cannot be separated easily. To explain this problem more clearly, we look at a generic form of the governing equations in Fourier space. With the assumption of periodicity, the wavevector $\mathbf{k}$ becomes discrete and can be considered as an index for the flow variables such as velocity or buoyancy. If we present them in the state vector $\mathbf{X_k}$, the governing equations can be written as

$$\frac{d\mathbf{X_k}}{dt} = -i\mathbf{L_k}\mathbf{X_k} + \mathbf{N_k}(\mathbf{X}), \tag{4.2}$$

where the matrix $\mathbf{L_k}$ represents the linear terms. The nonlinear terms are denoted by $\mathbf{N_k}$, in which $\mathbf{X} = \{\mathbf{X_k}\}$ is the set of all state vectors with different wavevectors. If we denote balanced modes by $\mathbf{B_k}$ and unbalanced modes by $\mathbf{U_k}$, the state vector can be written as $\mathbf{X_k} = \mathbf{B_k} + \mathbf{U_k}$. Balance models can generically be presented as

$$\frac{d\mathbf{B_k}}{dt} = \tilde{\mathbf{L}}_\mathbf{k}\ \mathbf{B} + \tilde{\mathbf{N}}_\mathbf{k}(\mathbf{B}). \tag{4.3}$$

We define the scalar product as $\langle X, Y\rangle = X_i Y_i^*$, in which the repeated indices, i, are summed and * denotes the complex conjugate. One may take the scaler product of (4.3) with $\mathbf{B_k}$ and integrate over cylindrical or spherical shells in spectral space to derive the Lin equation for the balanced energy. In this equation the balanced energy is not affected by unbalanced modes, as there is no $\mathbf{U_k}$ in the RHS of (4.3). Hence, to include the interaction between balance and imbalance, the full equation (4.2) should be used instead of (4.3). However, it is not obvious how to derive the budget of balanced and unbalanced energy from (4.2). In previous studies, two approaches

have been used. The first approach relies on a specific low-order definition of balance that makes $\mathbf{B_k}$ and $\mathbf{U_k}$ orthogonal. Studies such as Leith (1980), Bartello (1995), Ngan *et al.* (2008), Bartello (2010) and Deusebio *et al.* (2013) used the normal modes of the linear matrix $\mathbf{L_k}$ to define balance and imbalance. Since these are orthogonal $\langle \mathbf{B_k}, \mathbf{U_k} \rangle = 0$. Therefore, taking the scalar product of (4.2) with $\mathbf{B_k}$ eliminates $\mathbf{U_k}$ in the LHS but keeps it in the nonlinear terms in the RHS. Hence, it successfully derives the budget of balanced energy in which the interaction between $\mathbf{B_k}$ and $\mathbf{U_k}$ is taken into account. However, this approach is only applicable to linear balance. Therefore, it is of interest to find a more general approach for higher-order balanc, where $\mathbf{B_k}$ and $\mathbf{U_k}$ are not necessarily orthogonal.

Kafiabad & Bartello (2017) decomposed the transfer into three terms; transfer in the balanced flow, transfer in the unbalanced flow, and the transfer cross-terms. The first two terms are derived by setting $\mathbf{U_k} = 0$ and $\mathbf{B_k} = 0$ in the total transfer term, respectively. The last term is calculated as a residual. The physical interpretation of these terms is not clear. For instance, the transfer cross-terms do not show to which modes (balanced or unbalanced) the energy is transferred.

Considering the limitations and shortcomings of previous approaches we propose a new perturbation method to re-write the non-hydrostatic Boussinesq equations in terms of the budget of the balanced and unbalanced modes. Unlike its precedents, this method does not rely on the orthogonality of $\mathbf{B_k}$ and $\mathbf{U_k}$. Hence, it can be applied to higher-order balance. The fundamental assumption we use is the proximity of the solution to balance dynamics. Hence, its utility is in the study of the early stages of spontaneous imbalance from balanced ICs. According to the results

presented below, starting from an initial low-Ro and low-Fr balanced condition, imbalance generation occurs over a range of wavenumbers and times where the flow is predominantly balanced. Therefore, the use of methods based on the smallness of imbalance is justified.

This study can also be viewed as a response to a call for the investigation of balanced turbulence. As mentioned before, most previous studies of balance dynamics either relied on low-order models or special solutions of the governing equations. In his comprehensive review, Vanneste (2013) expressed a need to extend balanced methods "to treat PDE models and clarify how the full spatiotemporal structure of the balanced motion controls IGW generation". In the current study, we investigate imbalance generation from a rotating stratified turbulence perspective to address this need.

It is hoped that our study can be employed to formulate sub-grid parametrisation schemes that judiciously remove balanced energy and dissipate it without necessarily resolving the small-scale unbalanced dynamics. Moreover, our results can be used to determine the resolution that realistic models require to reliably describe large-scale atmospheric and oceanic dynamics.

This paper is organized as follows. After this introduction, we demonstrate our approach to derive the transfer between balance and imbalance. This is followed by a section that lays out the specifications of our system as well as the numerical configuration. In the results section, we first compare imbalance generation for different orders of balance. We then analyse the spectrum of transfer to unbalanced motion to find the scales at which most of the balanced energy leaks out. After this, we

investigate the effect of different parameters (such as Ro and N/f) on spontaneous imbalance. We also present a comprehensive section on time scales which investigates the frequency spectra of the different flow regimes. The conclusions are offered in §4.6.

4.3 Theory

4.3.1 Normal mode decomposition

For statistical homogeneity, we restrict our attention to a flow with periodic boundaries. With this assumption, (4.2) describes the evolution of flow variables. If the variables in $\mathbf{X}$ are appropriately chosen, the matrix of linear terms ($\mathbf{L_k}$ in (4.2)) will be Hermitian, with corresponding real eigenvalues and orthogonal eigenvectors. For the Boussinesq equations, there are a set of zero eigenvalues $\lambda_{\mathbf{k}}^{(0)} = 0$ and pairs of non-zero eigenvalues with opposite signs $\lambda_{\mathbf{k}}^{(\pm)} = \pm\sigma_{\mathbf{k}}$. If we present the unit eigenvectors associated with $\lambda_{\mathbf{k}}^{(0)}$ and $\lambda_{\mathbf{k}}^{(\pm)}$ by $\hat{\mathbf{g}}_{\mathbf{k}}$ and $\hat{\mathbf{a}}_{\mathbf{k}}$ respectively, the geostrophic modes $G_{\mathbf{k}}$ and the ageostrophic modes $A_{\mathbf{k}}$ are then defined by projecting $\mathbf{X_k}$ onto $\hat{\mathbf{g}}_{\mathbf{k}}$ and $\hat{\mathbf{a}}_{\mathbf{k}}$

$$G_{\mathbf{k}} = \langle \mathbf{X_k}, \hat{\mathbf{g}}_{\mathbf{k}} \rangle, \quad A_{\mathbf{k}}^{(\pm)} = \langle \mathbf{X_k}, \hat{\mathbf{a}}_{\mathbf{k}}^{(\pm)} \rangle, \quad \mathbf{X_k} = G_{\mathbf{k}}\hat{\mathbf{g}}_{\mathbf{k}} + A_{\mathbf{k}}^{(+)}\hat{\mathbf{a}}_{\mathbf{k}}^{(+)} + A_{\mathbf{k}}^{(-)}\hat{\mathbf{a}}_{\mathbf{k}}^{(-)} \tag{4.4}$$

We drop the superscript $(\pm)$ in $\hat{\mathbf{a}}_{\mathbf{k}}$ and $A_{\mathbf{k}}$ to simplify the notation (for more detail on normal mode decomposition see Leith, 1980; Bartello, 1995). Noting that $\hat{\mathbf{g}}_{\mathbf{k}}$ and $\hat{\mathbf{a}}_{\mathbf{k}}$ form an orthogonal basis, the projection of (4.2) onto $\hat{\mathbf{g}}_{\mathbf{k}}$ gives the time-evolution of $G_{\mathbf{k}}$

$$\frac{dG_{\mathbf{k}}}{dt} = \langle \mathbf{N_k}(\mathbf{X}), \hat{\mathbf{g}}_{\mathbf{k}} \rangle = \langle \mathbf{N_k}(G\hat{\mathbf{g}} + A\hat{\mathbf{a}}), \hat{\mathbf{g}}_{\mathbf{k}} \rangle = M_{\mathbf{k}}(G, A) \tag{4.5}$$

In the Boussinesq equations, the nonlinear terms are quadratic. Therefore, the RHS of (4.5) can be expanded as

$$M_{\mathbf{k}}(G,A) = \zeta_{GG}(G,G) + \zeta_{GA}(G,A) + \zeta_{AA}(A,A), \tag{4.6}$$

where

$$\zeta_{GG}(G,G) = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} \Gamma_{\mathbf{k},\mathbf{p},\mathbf{q}} G^*_{\mathbf{q}}\, G^*_{\mathbf{p}} + c.c. \tag{4.7a}$$

$$\zeta_{GA}(G,A) = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} \Psi_{\mathbf{k},\mathbf{p},\mathbf{q}} G^*_{\mathbf{q}}\, A^*_{\mathbf{p}} + c.c. \tag{4.7b}$$

$$\zeta_{AA}(A,A) = \sum_{\mathbf{k}+\mathbf{p}+\mathbf{q}=0} \Phi_{\mathbf{k},\mathbf{p},\mathbf{q}} A^*_{\mathbf{q}}\, A^*_{\mathbf{p}} + c.c. \tag{4.7c}$$

(*c.c.* denoting the complex conjugate). $\Gamma_{\mathbf{k},\mathbf{p},\mathbf{q}}$, $\Psi_{\mathbf{k},\mathbf{p},\mathbf{q}}$ and $\Phi_{\mathbf{k},\mathbf{p},\mathbf{q}}$ are coefficients of the convolution sums and functions of $\mathbf{k}$, $\mathbf{p}$ and $\mathbf{q}$. Note that ζ_{GG}, ζ_{GA} and ζ_{AA} can be made symmetric with respect to their first and second arguments.

Similarly, we can derive the time-evolution of $A_{\mathbf{k}}$ by projecting (4.2) onto $\hat{\mathbf{a}}_{\mathbf{k}}$

$$\frac{dA_{\mathbf{k}}}{dt} - i\lambda_{\mathbf{k}} A_{\mathbf{k}} = Q_{\mathbf{k}}(G,A) = \eta_{GG}(G,G) + \eta_{GA}(G,A) + \eta_{AA}(A,A). \tag{4.8}$$

4.3.2 Balance dynamics

The eigenfrequencies introduced in the previous section indicate there are two time scales in the flow if Ro is small enough. Therefore, we can rewrite the time derivative in terms of a fast time t^* and a slow time $T = \epsilon t^*$ as below

$$\frac{d()}{dt} = \frac{d()}{dt^*} + \epsilon \frac{d()}{dT}, \tag{4.9}$$

in which ϵ is a small parameter proportional to the Rossby number. This separation enables us to decompose the flow into a balanced and an unbalanced part. The unbalanced part has fast dynamics; whereas the balanced part is a slowly-varying. We assume the $G_{\mathbf{k}}$ are slowly-varying when ϵ is small enough. Hence, the geostrophic modes are considered balanced. The ageostrophic modes, however, can admit a slow and a fast time scale. Hence, they can have a balanced, $A^b_{\mathbf{k}}$, and an unbalanced part, $A^u_{\mathbf{k}}$. Assuming the separation of time scales is large enough, one may try to reduce the dynamics to only slow motion with an independent set of equations. As explained in Warn *et al.* (1995), this can be done by defining a balance relation that filters the fast oscillation and a balance equation that describes the slow-time evolution. Since only the geostrophic modes contribute to the linear potential vorticity and are more energetic than the ageostrophic modes (when Ro is small enough), one can write the balance equation for $G_{\mathbf{k}}$, by setting $A = 0$ in (4.5)

$$\frac{dG_{\mathbf{k}}}{dT} = \zeta_{GG}(G, G), \tag{4.10}$$

which is known as QG dynamics and describes the geostrophic part of these balance models. The ageostrophic modes are derived from the (diagnostic) balance relation (also known as the slave relation)

$$A^b_{\mathbf{k}} = F_{\mathbf{k}}(G, \epsilon), \tag{4.11}$$

where $G = \{G_{\mathbf{k}'}\}$ denotes the set of all geostrophic modes. The entire balanced flow can then be written as $\mathbf{B}_{\mathbf{k}} = G_{\mathbf{k}}\hat{\mathbf{g}}_{\mathbf{k}} + A^b_{\mathbf{k}}\hat{\mathbf{a}}_{\mathbf{k}}$. The solutions of governing equations can be considered as trajectories in the phase space, and (4.11) can be seen as the slow

manifold that they lie on (see Leith, 1980; Lorenz, 1980, for the precise definition). If the fast oscillations do not affect the slow motion, this manifold should be invariant, meaning that trajectories which start on it stay there. With this assumption, the evolution of A^b is derived by taking the time derivative of (4.11) and applying the chain rule

$$\frac{dA^b_{\mathbf{k}}}{dT} = \sum_{\mathbf{k}'} \frac{dG_{\mathbf{k}'}}{dT} \frac{\partial F_{\mathbf{k}}(G,\epsilon)}{\partial G_{\mathbf{k}'}} = \sum_{\mathbf{k}'} \zeta_{GG}(G,G) \frac{\partial F_{\mathbf{k}}(G,\epsilon)}{\partial G_{\mathbf{k}'}}, \tag{4.12}$$

in which $dG_{\mathbf{k}'}/dT$ is obtained from (4.10) and the summation is carried over all wavenumbers.

As mentioned in §4.2, if balance breaks down, (4.10) and (4.12) cannot describe the time evolution of the flow accurately. More particularly, the RHS of (4.10) does not include the nonlinear interaction between balanced and unbalanced parts of the flow, whereas (4.5) does. Despite the inaccuracy of (4.10) compared to (4.5), the slave relation (4.11) can still be a good approximation for the relationship between G and A^b, assuming the solution is predominantly balanced. Considering this fact, to derive a better approximation for $dA^b_{\mathbf{k}}/dt$, one may combine (4.5) and (4.11) to obtain

$$\frac{dA^b_{\mathbf{k}}}{dt} = \sum_{\mathbf{k}'} \frac{dG_{\mathbf{k}'}}{dt} \frac{\partial F_{\mathbf{k}}(G,\epsilon)}{\partial G_{\mathbf{k}'}} = \sum_{\mathbf{k}'} M_{\mathbf{k}}(G,A) \frac{\partial F_{\mathbf{k}}(G,\epsilon)}{\partial G_{\mathbf{k}'}}, \tag{4.13}$$

in which $dG_{\mathbf{k}'}/dt$ is calculated from the RHS of (4.5). Since the RHS of (4.5) consists of both balanced and unbalanced terms, we can derive the transfer between balance and imbalance from (4.13) (at least to leading order). This approach is particularly helpful when imbalance generation from an initially balanced flow is considered.

To proceed we must first provide a working definition of balance. This definition determines $F_{\mathbf{k}}$ in (4.11) and therefore $A^b_{\mathbf{k}}$ and $A^u_{\mathbf{k}}$. To that end, we employ the Nonlinear Normal Mode Initialization (NNMI) scheme introduced by Baer & Tribbia (1977), as described in Kafiabad & Bartello (2016) (see Appendix A).

Assuming m to be the total number of modes in the system, $\partial F_{\mathbf{k}}/\partial G_{\mathbf{k}'}$ in the RHS of (4.13) will have $m \times m$ elements. Finding a general expression for this large matrix is a tedious task. Instead of calculating this term explicitly, we use an indirect numerical approach based on the perturbation of $dA^b_{\mathbf{k}}/dt$ in terms of ϵ. To that end, we use (4.5) to calculate the time derivative of geostrophic modes. The details of our approach are explained in the following section.

4.3.3 The transfer to/from imbalance

The following scaled equations describe the evolution of $A^b_{\mathbf{k}}$, in which the fast time derivatives are set to zero

$$\frac{dA^b_{\mathbf{k}}}{dT} + \frac{i\sigma A^b_{\mathbf{k}}}{\epsilon} = \eta_{GG}(G,G) + \eta_{GA}(G,A^b) + \eta_{AA}(A^b,A^b). \tag{4.14}$$

In order to calculate (4.13), we use (4.14) to derive $\partial F/\partial G$, while we use (4.5) to obtain dG/dt. We expand the balanced ageostrophic modes in terms of ϵ as

$$A^b_{\mathbf{k}} = \alpha + \alpha^{'}_{\mathbf{k}}\epsilon + \alpha^{''}_{\mathbf{k}}\epsilon^2 + O(\epsilon^3), \tag{4.15}$$

and substitute them into (4.14). By expanding both sides of (4.14) at order $O(\epsilon^0)$, one can deduce that $\alpha = 0$, which yields

$$O(A^b) = \epsilon O(G) \tag{4.16}$$

The expansion at $O(\epsilon)$ yields

$$\alpha_{\mathbf{k}}^{'} = \frac{1}{i\sigma_{\mathbf{k}}}\eta_{GG}(G,G), \tag{4.17}$$

which is the first iteration of Machenhauer's (1977) initialisation scheme. By substituting (4.17) into (4.15) the balanced ageostrophic modes are derived at first order. As mentioned before, our goal is to derive dA^b/dt. In so doing, we take d/dt of (4.17)

$$\frac{d\alpha_{\mathbf{k}}^{'}}{dt} = \frac{2}{i\sigma_{\mathbf{k}}}\eta_{GG}(G,\frac{dG}{dt}) = \frac{2}{i\sigma_{\mathbf{k}}}\eta_{GG}(G,M), \tag{4.18}$$

where in the last term, dG/dt is replaced by the RHS of (4.5) (denoted by $M = M(G,A)$). Note that the quadratic form of η_{GG} is used to derive (4.18).

(4.18) yields the first approximation of dA^b/dt, which is relatively straightforward to calculate numerically. In the results section, we show that the first order approximation is good enough in the regimes of low and intermediate Rossby numbers, which are the subject of our study.

To find the next order, we expand (4.14) at $O(\epsilon^2)$

$$\frac{d\alpha_{\mathbf{k}}^{'}}{dt} + i\sigma_{\mathbf{k}}\alpha_{\mathbf{k}}^{''} = \eta_{GA}(G,\alpha^{'}) \tag{4.19}$$

Using (4.18) we can substitute for $d\alpha^{'}/dt$ in (4.19) and derive an expression for $\alpha_{\mathbf{k}}^{''}$

$$\alpha_{\mathbf{k}}^{''} = \frac{1}{i\sigma_{\mathbf{k}}}\left(\eta_{GA}(G,\alpha^{'}) - \frac{2}{i\sigma_{\mathbf{k}}}\eta_{GG}(G,M)\right). \tag{4.20}$$

Taking the time derivative of this equation and replacing dG/dt with M yields

$$\frac{d\alpha_{\mathbf{k}}^{''}}{dt} = \frac{1}{i\sigma_{\mathbf{k}}}\left(\eta_{GA}(M,\alpha^{'}) + \eta_{GA}(G,\frac{d\alpha^{'}}{dt}) - \frac{2}{i\sigma_{\mathbf{k}}}\left(\eta_{GG}(M,M) + \eta_{GG}(G,\frac{dM}{dt})\right)\right), \tag{4.21}$$

in which $d\alpha^{'}/dt$ can be derived by using (4.19). Finally, (4.18) and (4.21) can be used to derive the time derivative of the balanced ageostrophic flow up to $O(\epsilon^2)$

$$\frac{dA_{\mathbf{k}}^{b}}{dt} = \frac{d\alpha_{\mathbf{k}}^{'}}{dt}\,\epsilon + \frac{d\alpha_{\mathbf{k}}^{''}}{dt}\,\epsilon^2 + O(\epsilon^3) \tag{4.22}$$

Once dA^b/dt is calculated, it can be subtracted from the time derivative of the total ageostrophic modes in (4.8) to yield $dA_{\mathbf{k}}^{u}/dt$.

4.4 Governing equations and numerical settings

The approach that was described in the previous section can be applied to different governing equations. In this study we use the non-hydrostatic Boussinesq equation on an f-plane written as

$$\frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + f\hat{\boldsymbol{z}}\times\boldsymbol{u} = -\nabla p + b\hat{\boldsymbol{z}} + \boldsymbol{D_u}, \tag{4.23a}$$

$$\nabla\cdot\boldsymbol{u} = 0, \tag{4.23b}$$

$$\frac{\partial b}{\partial t} + \boldsymbol{u}\cdot\nabla b = -N^2 w + D_b, \tag{4.23c}$$

where $\boldsymbol{u}$ is the velocity vector and b is the buoyancy perturbation. The operator D_q represents the dissipation of quantity q. The Brunt-Väisälä frequency, N, is assumed to be constant. Setting the boundaries to be periodic, (4.23) can be transformed to Fourier space and written in the form described in (4.2). Following the normal

mode decomposition explained in §4.3.1, the linear dispersion relation of the non-hydrostatic Boussinesq equation is

$$\lambda_{\mathbf{k}}^{(0)} = 0, \quad \lambda_{\mathbf{k}}^{\pm} = \pm\sigma_{\boldsymbol{k}} = \pm\frac{(f^2 {k_z}^2 + N^2 {k_h}^2)^{1/2}}{k}, \tag{4.24}$$

where $k_h = (k_x^2 + k_y^2)^{1/2}$ and $k = (k_h^2 + k_z^2)^{1/2}$.

The partial differential equations in (4.23) need a set of ICs to be integrated. As expressed in previous sections, we want the flow to be predominantly balanced. Therefore, we seek ICs that are nonlinearly balanced at high order. To produce such conditions we used the NNMI scheme of Baer & Tribbia (1977). In so doing, we first generate geostrophic data and then use (4.15) to include the balanced ageostrophic modes in the IC. We carry out the expansion in (4.15) up to $O(\epsilon^2)$ in all the simulations of this paper. To produce the geostrophic data, we run a QG model that has the same resolution and dissipation as the corresponding Boussinesq simulation. More detail can be found in Kafiabad & Bartello (2016).

In our Boussinesq simulations we employ the de-aliased pseudo-spectral method with centred time stepping. The size of our domain is set to 2π in the horizontal and to $2\pi \times (f/N)$ in the vertical. In this configuration the aspect ratio, $\alpha = H/L$, is equal to f/N, which makes the domain a cube of $(2\pi)^3$ in the classical QG coordinates (Charney, 1971). We employ equally spaced grids in physical space, i.e. $\triangle x = \triangle y = \triangle z$. An alternative grid is equally spaced in the scaled coordinate, i.e. $\triangle x = \triangle y = N/f \triangle z$. We compared the results on both grids in Kafiabad & Bartello (2017) and expressed the benefits and shortcomings of each. We concluded

the former grid might better suit our study that is focused on the breakdown of balance. Therefore, all simulations in this paper are performed on the isotropic grid.

The simulations presented in the following sections are listed in table 4–1. They employ either 2048 or 1536 grid points in the horizontal. The ratio of horizontal to vertical resolution is N/f to keep $\triangle x = \triangle z$. For most of our simulations $N/f = 8$. However, we performed several simulations with $N/f = 16$ and 32 to investigate the effect of N/f on the breakdown of balance.

All the simulations in this paper are of decaying turbulence. We use the following form of hyper dissipation for $D_{\boldsymbol{u}}$ and D_b in (4.23)

$$D_{\boldsymbol{u},b} = -\nu_h \nabla_h^8(\boldsymbol{u}, b) - \nu_z \frac{\partial^8}{\partial z^8}(\boldsymbol{u}, b), \tag{4.25}$$

We assume the same dissipation for buoyancy and velocity. Since the flow is expected to be close to isotropic in the dissipation range (and knowing that $\triangle x = \triangle z$ in our grid), we set $\nu_h = \nu_z = \nu$. All the times reported in this paper are normalized with the rms geostrophic contribution to the vertical vorticity, τ. This value is measured at the beginning of each simulation and presented in table 4–1. Hence, the reported values of time can be considered as initial eddy turnover times.

The non-dimensional parameters in table 4–1 are calculated as

$$Ro = \frac{< \omega_z^2 >^{1/2}}{f}, \quad Fr = \frac{< \omega_x^2 + \omega_y^2 >^{1/2}}{N} \tag{4.26}$$

in which $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ is the vorticity vector and $< >$ denotes a spatial average.

Simulation	N_h	N_z	f	N	Ro	Fr	τ	ν
B2048R.288	2048	256	21.22	169.77	0.288	0.112	6.11	5×10^{-20}
B1536R.314	1536	192	16.98	135.84	0.314	0.112	6.11	5×10^{-19}
B1536R.533	1536	192	10.19	81.50	0.533	0.204	6.11	5×10^{-19}
B1536R.650	1536	192	8.49	67.92	0.650	0.262	6.11	5×10^{-19}
B1536R.772	1536	192	7.28	58.22	0.772	0.330	6.11	5×10^{-19}
B1536R.199	1536	192	16.98	135.84	0.199	0.075	3.40	5×10^{-19}
B1536R.579	1536	192	5.66	45.28	0.579	0.268	3.40	5×10^{-19}
B1536Ki160	1536	192	3.92	31.34	0.501	0.077	1.97	5×10^{-19}
B1536Ki49	1536	192	5.33	42.62	0.509	0.167	2.72	5×10^{-19}
B1536Ki20	1536	192	7.81	62.46	0.517	0.199	4.06	5×10^{-19}
B1536Ki10	1536	192	10.67	85.39	0.514	0.188	5.46	5×10^{-19}
B2048N/f8	2048	256	11.63	93.04	0.58	0.256	6.79	9×10^{-20}
B2048N/f16	2048	128	9.54	152.64	0.58	0.223	5.50	7×10^{-20}
B2048N/f32	2048	64	11.43	365.70	0.58	0.169	6.58	5×10^{-21}

Table 4–1: The non-hydrostatic Bousinesq simulations presented in this paper. The values of Ro, Fr and τ are calculated at $t = 0$.

4.5 Results

4.5.1 Comparing different orders of balance

To derive the transfer between balance and imbalance, we re-write the full Boussinesq equation defined in (4.23) in terms of $\mathbf{B}_i$ and $\mathbf{U}_i$, in which i denotes the order of balance. As explained in §4.3.3, the definition of balance and imbalance depends upon the order to which the ageostrophic modes are expanded in (4.15), as does the transfer between balance and imbalance. In this section, we study how this can affect the results of imbalance generation.

Zeroth Order

The zeroth order is derived by keeping terms up to $O(\epsilon^0)$ in (4.15). Knowing that $\alpha = 0$, $\mathbf{B}_0 = G\hat{\mathbf{g}}$ and $\mathbf{U}_0 = A\hat{\mathbf{a}}$. This order of balance is referred to as linear

balance, since the nonlinear interaction between geostrophic and ageostrophic modes are not taken into account. Unlike higher-order balance, $\mathbf{B}_0$ and $\mathbf{U}_0$ are orthogonal. Hence, the Boussinesq equations in terms of $\mathbf{B}_0$ and $\mathbf{U}_0$ are readily derived by using (4.5) and (4.8)

$$\frac{d\mathbf{B}_{0\mathbf{k}}}{dt} = M_{\mathbf{k}}(A, G)\hat{\mathbf{g}} \tag{4.27a}$$

$$\frac{d\mathbf{U}_{0\mathbf{k}}}{dt} = \big(+ i\lambda_{\mathbf{k}}^{(\pm)} A_{\mathbf{k}} + Q_{\mathbf{k}}(A, G)\big)\hat{\mathbf{a}}. \tag{4.27b}$$

Here, balance and imbalance are merely geostrophic and ageostrophic modes, and we may use G and A interchangeably with $\mathbf{B}_0$ and $\mathbf{U}_0$.

First Order

Keeping terms up to $O(\epsilon)$ in (4.15) leads to a balance that includes an ageostrophic contribution, $\mathbf{B}_1 = G\hat{\mathbf{g}} + \epsilon\alpha'\hat{\mathbf{a}}$ and $\mathbf{U}_1 = (A - \epsilon\alpha')\hat{\mathbf{a}}$. Note at this order, $\mathbf{B}_1$ and $\mathbf{U}_1$ are no longer orthogonal as they both have a component in the $\hat{\mathbf{a}}$ direction. This balance is the first-order approximation in the Baer & Tribbia (1977) initialisation scheme, as well as the first iteration of Machenhauer's (1977) scheme. Leith (1980) discussed that at this order, the balance dynamics are quasigeostrophic for the f-plane model, imlying the linear balance in §4.5.1 is even less accurate than quasigeostrophy, since it has no ageostrophic contribution to geostrophy. Defining $\mathbf{B}_1$ and $\mathbf{U}_1$ in this way, the time evolutions in the Boussinesq equation are

$$\frac{d\mathbf{B}_{1\mathbf{k}}}{dt} = M_{\mathbf{k}}(A,G)\hat{\mathbf{g}} + \epsilon\frac{d\alpha_{\mathbf{k}}^{'}}{dt}\hat{\mathbf{a}} \tag{4.28a}$$

$$\frac{d\mathbf{U}_{1\mathbf{k}}}{dt} = \big(i\lambda_{\mathbf{k}}^{(\pm)}A_{\mathbf{k}} + Q_{\mathbf{k}}(A,G) - \epsilon\frac{d\alpha_{\mathbf{k}}^{'}}{dt}\big)\hat{\mathbf{a}}. \tag{4.28b}$$

Second Order

The second-order balance, based on Baer & Tribbia (1977), yields $\mathbf{B}_2 = G\hat{\mathbf{g}} + \big(\alpha^{'}\epsilon + \alpha^{''}\epsilon^2\big)\hat{\mathbf{a}}$ and $\mathbf{U}_2 = \big(A - \alpha^{'}\epsilon - \alpha^{''}\epsilon^2\big)\hat{\mathbf{a}}$. This formulation can also be related to the bounded derivative method proposed by Kreiss (1979) (see Leith, 1980).

The time evolution of $\mathbf{B}_2$ and $\mathbf{U}_2$ based on the Boussinesq equations is

$$\frac{d\mathbf{B}_{2\mathbf{k}}}{dt} = M({}_{\mathbf{k}}A,G)\hat{\mathbf{g}} + \big(\epsilon\frac{d\alpha_{\mathbf{k}}^{'}}{dt} + \epsilon^2\frac{d\alpha_{\mathbf{k}}^{''}}{dt}\big)\hat{\mathbf{a}} \tag{4.29a}$$

$$\frac{d\mathbf{U}_{2\mathbf{k}}}{dt} = \big(i\lambda_{\mathbf{k}}^{(\pm)}A_{\mathbf{k}} + Q_{\mathbf{k}}(A,G) - \epsilon\frac{d\alpha_{\mathbf{k}}^{'}}{dt} - \epsilon^2\frac{d\alpha_{\mathbf{k}}^{''}}{dt}\big)\hat{\mathbf{a}}. \tag{4.29b}$$

In these equations, $d\alpha^{'}/dt$ and $d\alpha^{''}/dt$ are calculated from (4.18) and (4.21), respectively.

4.5.2 Balanced and unbalanced energy spectra

Regardless of the order of balance, the budget of $\mathbf{B}$ and $\mathbf{U}$ defined in (4.27-4.29) can be written in the following generic forms

$$\frac{d}{dt}\mathbf{B}_{\mathbf{k}} = N_{B\mathbf{k}}(\mathbf{B},\mathbf{U}) + L_{B\mathbf{k}}\mathbf{B}_{\mathbf{k}} \tag{4.30a}$$

$$\frac{d}{dt}\mathbf{U}_{\mathbf{k}} = N_{U\mathbf{k}}(\mathbf{B},\mathbf{U}) + L_{U\mathbf{k}}\mathbf{U}_{\mathbf{k}}, \tag{4.30b}$$

in which the first terms in the RHS are the nonlinearities and the second terms are linear. We emphasise again that (4.30a) does not describe balance dynamics; it describes the evolution of the balance component in a flow that is close to balance.

Defining the horizontal balanced and unbalanced energy spectra as

$$E_{\mathbf{B}}(k_h) = \sum_{|\mathbf{k}'-\mathbf{k}'\cdot z|=k_h} \langle \mathbf{B}_{\mathbf{k}'}, \mathbf{B}_{\mathbf{k}'} \rangle, \quad E_{\mathbf{U}}(k_h) = \sum_{|\mathbf{k}'-\mathbf{k}'\cdot z|=k_h} \langle \mathbf{U}_{\mathbf{k}'}, \mathbf{U}_{\mathbf{k}'} \rangle, \tag{4.31}$$

the following expressions are obtained for the Lin equation governing the evolution of these spectra using (4.30)

$$\frac{d}{dt}\, E_{\mathbf{B}}(k_h) = T_B(k_h), \tag{4.32a}$$

$$\frac{d}{dt}\, E_{\mathbf{U}}(k_h) = T_U(k_h), \tag{4.32b}$$

where $T_B(k_h)$ and $T_U(k_h)$ are the horizontal transfer spectra to the balanced and unbalanced energy. These terms can be written in terms of nonlinearities as below

$$T_B(k_h) = \sum_{|\mathbf{k}'-\mathbf{k}'\cdot z|=k_h} \langle N_{B\mathbf{k}}(\mathbf{B},\mathbf{U}), \mathbf{B}^*_{\mathbf{k}} \rangle, \quad T_U(k_h) = \sum_{|\mathbf{k}'-\mathbf{k}'\cdot z|=k_h} \langle N_{U\mathbf{k}}(\mathbf{B},\mathbf{U}), \mathbf{U}^*_{\mathbf{k}} \rangle. \tag{4.33}$$

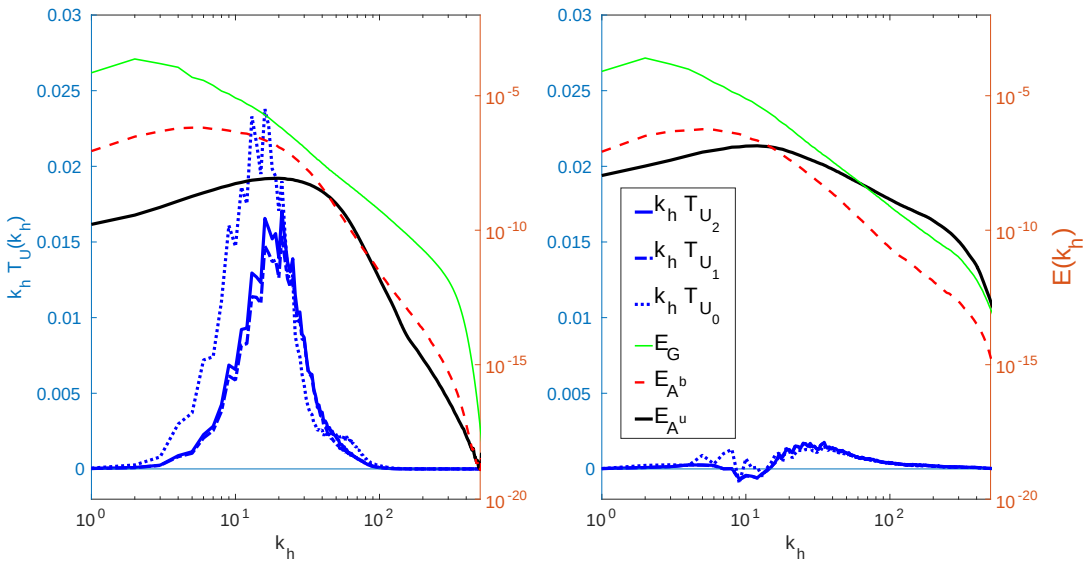


Figure 4–1: The unbalanced transfer spectra, $T_U(k_h)$, at early times in the left panel and at late times in the right panel. For the early times the spectra are averaged over the time interval of $[0, 0.1\tau]$, and for the late times over the interval of $[0.4\tau, 0.5\tau]$. The transfer to imbalance is calculated for different orders of balance marked in the legend. The horizontal energy spectra of the same simulation at the same times are also plotted: E_G (green thin line), E_{A^b} (red dashed line) and E_{A^u} (black thick line). The left axis is associated with the transfer spectra and the right axis with the energy spectra. $Ro = 0.29$ (Simulation B2048R.288).

Figure 4–1 portrays $T_U(k_h)$ at $Ro = 0.288$, for three different orders of balance described in §4.5.1-4.5.1. The left plot corresponds to the transfer at the beginning, and the right plot corresponds to the end of the integration period. The balanced and unbalanced ageostrophic as well as the geostrophic energy spectra are also portrayed. For this flow the shallow unbalanced energy spectrum surpasses E_G at the end of the integration time. Hence, the total energy spectrum admits a kink at $k_h \simeq 60$. At early times, the energy is transferred to imbalance at scales larger than the kink in the energy spectrum ($k_h \simeq 20$) where the flow is still predominantly balanced. This transfer generates an unbalanced energy spectrum that peaks at the same wavenumber as $T_U(k_h)$ does. This spectrum has a steep slope initially, but it shallows in time, and its peak moves to larger scales. The transfer decreases considerably in time and its peak moves toward smaller scales. At the later time there is a negative transfer at very large scales, followed by a positive transfer. This indicates that the unbalanced part of the flow cascades the energy (that it received initially from the balanced modes) forward toward the dissipation range. Note that the lin-log $k_h T_U(k_h)$ plot is area-preserving, which indicates that the integral representing the net transfer to imbalance is negligible.

Comparing different orders, one finds that the zeroth order is somewhat different from the higher orders. T_{U0} (which is simply transfer to ageostrophic modes) is larger than T_{U1} and T_{U2} and peaks at smaller wavenumber at the beginning of the simulation. Later, the difference between T_{U0} and higher orders decreases, since Ro decreases in our decaying simulations and the expansion in (4.15) converges. Comparing the first and the second orders at the early time shows they are very similar.

They marginally differ at the peak of the transfer spectrum and at wavenumbers slightly smaller than the peak. The difference between T_{U1} and T_{U2} is not observable at the later time as both curves lie on top of each other at considerably lower amplitude. This comparison demonstrates that first order balance (in which nonlinear interactions are considered to the leading order) is a good approximation for the relatively low Rossby number of 0.29.

4.5.3 The contribution of different triads

Once the balanced and unbalanced modes are defined, the transfer to balance and imbalance, T_B and T_U, are due to the interactions between three modes, or triads. We denote these based on the type of their interactions. For instance, **BBU** represents the interactions between one unbalanced and two balanced modes. Separating the contributions to T_B and T_U according to the type of interaction leads to the following equations

$$T_B(k_h) = T_{BB\to B}(k_h) + T_{BU\to B}(k_h) + T_{UU\to B}(k_h) \tag{4.34a}$$

$$T_U(k_h) = T_{BB\to U}(k_h) + T_{BU\to U}(k_h) + T_{UU\to U}(k_h), \tag{4.34b}$$

where, for instance, $T_{BU\to B}(k_h)$ is the net transfer of all **BBU** interactions to or from **B** at wavenumber k_h.

Before analysing the numerical results, we try to draw some analytical conclusions on triad interactions. By considering (4.18), one can conclude that **BUU** do not transfer energy to A_b at first order, because if we set B equal to zero, G must

be zero. Therefore, the RHS of (4.18) becomes zero as well. This conclusion holds at any time as long as the flow is slowly-varying.

If we consider an initially balanced flow at early times, we can derive a more general statement. In this case, we can assume that $\mathbf{U} = O(\tilde{\epsilon}\ \mathbf{B})$ for small time, simply because $\mathbf{U} = 0$ at $t = 0$. Note that $\tilde{\epsilon}$ is a small parameter which is not necessarily equal to $\epsilon = Ro$ but is related to ϵ since the breakdown of balance can be delayed by decreasing Rossby number. We expand the governing equations in terms of $\tilde{\epsilon}$ and keep terms up to $O(\tilde{\epsilon})$. With some algebra (shown in Appendix C), we derive the following equation for the budget of the balanced energy spectrum at asymptotically small times

$$\frac{d}{dt}E_{\mathbf{B}}(k_h) = \sum_{|\mathbf{k}'-\mathbf{k}'\cdot z|=k_h} \big(\langle \xi_{BB}(\mathbf{B},\mathbf{B}), \mathbf{B}^*_{\mathbf{k}}\rangle + \langle \xi_{BU}(\mathbf{B},\mathbf{U}), \mathbf{B}^*_{\mathbf{k}}\rangle + \langle \xi_{BB}(\mathbf{B},\mathbf{B}), \mathbf{U}^*_{\mathbf{k}}\rangle + c.c.\big), \tag{4.35}$$

in which, ξ_{BB} and ξ_{BU} are quadratic functions (see Appendix C). This equation holds regardless of balance model or its order, as long as t is small enough. The RHS of (4.35) demonstrates that when balance breaks down, the first order interaction between balance and imbalance takes place via **BBU** triads.

We can numerically calculate the different nonlinear terms defined in (4.34). $T_{BB\to U,B}$ is determined by filtering all the unbalanced terms. Similarly $T_{UU\to U,B}$ is calculated by setting the balanced part of the flow to zero. $T_{BU\to U,B}$ is the residual derived by subtracting $T_{BB\to U,B}$ and $T_{UU\to U,B}$ from the total transfer $T_{B,U}$. To simplify the computations, we only consider first-order balance for the calculations of this section.

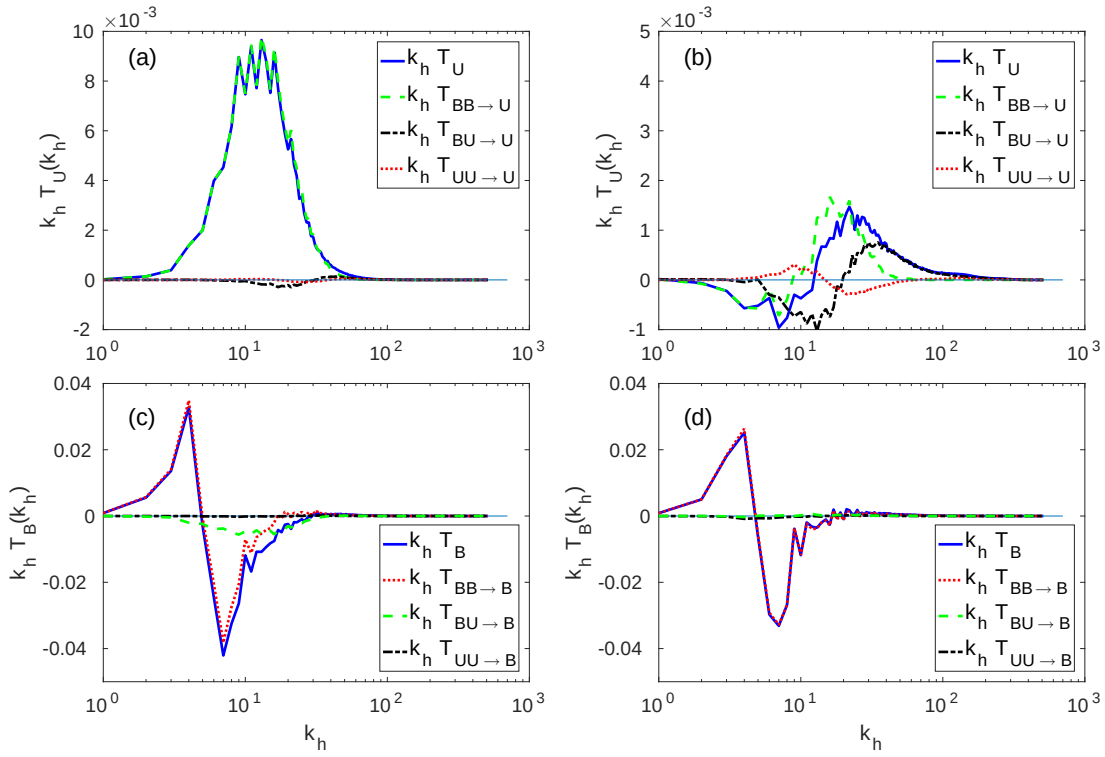


Figure 4–2: Top row: unbalanced transfer spectra. Bottom row: balanced transfer spectra. $Ro = 0.53$ (Simulation B1536R.533). The left column corresponds to the early time (averaged over $t \in [0, 0.1\tau]$), and the right column to the later time (averaged over $t \in [0.4\tau, 0.5\tau]$). The total transfer to balance/imbalance is shown by solid (blue) lines and it is decomposed into its constituent triads: the triads of the same kind ($BB \to B$ and $UU \to U$) shown with dotted (red) lines, the triads of **BBU** (either in the form of $BU \to B$ or $BB \to U$) shown with dashed (green) lines and the triads of **BUU** (either in the form of $UU \to B$ or $BU \to U$) shown with dash-dotted (black) lines.

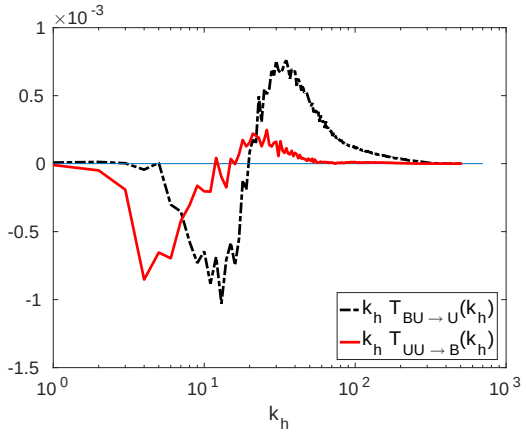


Figure 4–3: $T_{BU\to U}$ and $T_{UU\to B}$ averaged over $[0.4\tau, 0.5\tau]$

The results of these calculations are shown in figure 4–2. The first row displays the transfer to unbalanced modes and the second row the transfer to balanced modes. The left panels correspond to transfer at early time when balance breaks down, and the right panels correspond to later time when the shallow tail is developed in the energy spectrum. At early time, the major interactions between **B** and **U** is via **BBU** triads, as predicted by (4.35). Hence, T_U follows $T_{BB\to U}$. These triads transfer energy from balance to imbalance modes. Due to NNMI, there is no fast time scale initially. Therefore, resonant interactions between two opposite-signed fast modes, as described by Bartello (1995), do not prevail at early times in our simulations.

Figure 4–2.b shows the transfers to imbalance at the end of the integration time, in which $T_{BB\to U}$ decreases and $T_{BU\to U}$ increases significantly. Therefore, the two terms become comparable but peak at different wavenumbers. Based on this plot, **BBU** triads are still more important at larger scales where they cascade energy forward toward smaller scales. At smaller scales **BUU** triads become more

important and T_U follows $T_{BU\to U}$. These triads transfer the balanced energy, that was channelled to the unbalanced flow, to the dissipation range. Figure 4–3 displays $T_{BU\to U}$ and $T_{UU\to B}$ together averaged over the interval of $[0.4\tau, 0.5\tau]$. At the scales that **BUU** triads are more important, $T_{BU\to U}$ is substantially larger than $T_{UU\to B}$, signalling that **B** acts somewhat like a catalyst facilitating the transfer between the two unbalanced modes and remains approximately unchanged. Since the fast time scale is now present, resonant catalytic interactions may occur. Bartello (1995) demonstrated that two opposite-signed ageostrophic modes in GAA triads can resonate and transfer significant amounts of energy. The geostrophic mode in this resonant interaction plays the role of a catalyst and remains rather unchanged, as long as the PV is close to linear.

Figure 4–2.d shows that in time the flow transforms to another state of balance that it maintains. This can be considered as a nonlinear geostrophic adjustment (Zeitlin *et al.*, 2003). At the end of the integration time, T_B is approximately equal to $T_{BB\to B}$, indicating that the interactions with imbalance are too small to affect the balanced flow. At very small wavenumbers in figure 4–2.b, the total transfer to imbalance is negative followed by a positive transfer at higher wavenumbers. This is consistent with a forward cascade of unbalanced energy without significant interaction with the balanced part. $T_{UU\to U}$, which represents the **UUU** interactions, is very small compared to all other terms. At the end of the integration time, this term is positive at large scales and then becomes negative at smaller scales and then positive again, although this structure is uncertain, possibly due to insufficient averaging.

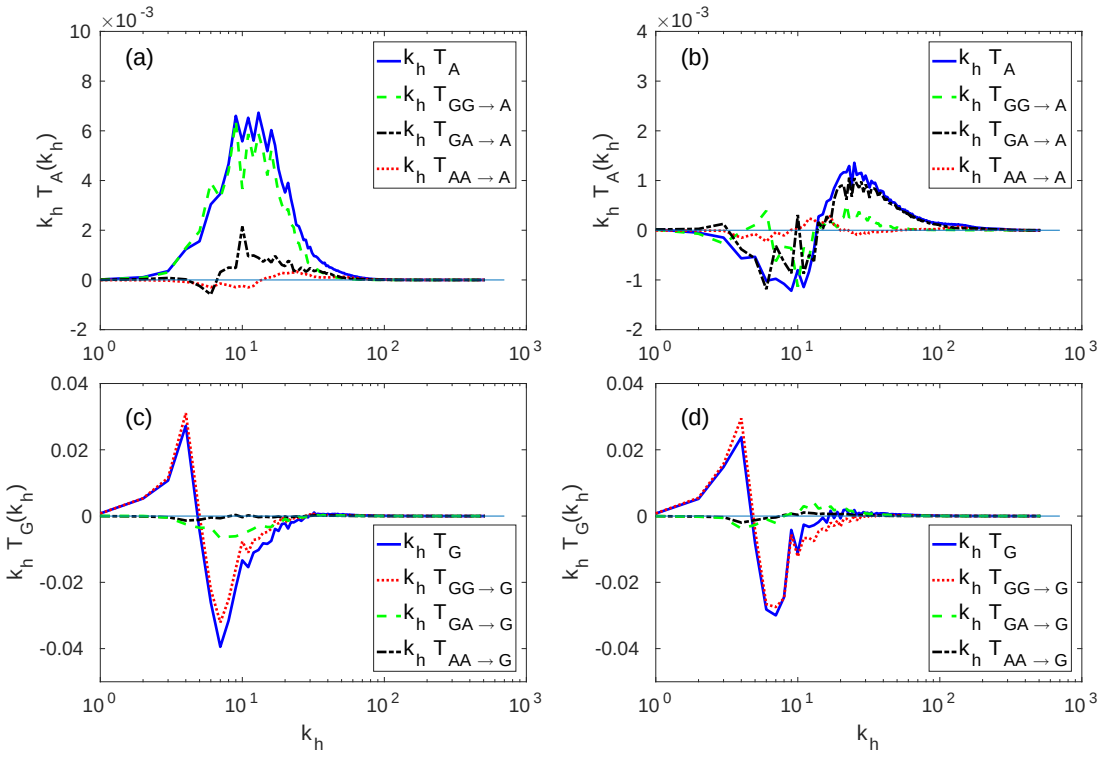


Figure 4–4: The same features as figure 4–2 for the same simulations, except $\mathbf{B}$ is replaced by G and $\mathbf{U}$ by A.

Several studies such as Bartello (1995), Ngan *et al.* (2008) and Deusebio *et al.* (2013) investigated the nonlinear transfer by decomposing it into the interactions between geostrophic and ageostrophic modes (i.e. based on the triads of type GGG, GGA, GAA and AAA). Unlike these studies, we employ a higher-order balance to classify the types of nonlinear interactions. To show how this improves the analysis, we compare figure 4–2 with its zeroth-order counterpart in which **B** is replaced by geostrophic modes and **U** by ageostrophic modes (figure 4–4). The same simulation is used to derive the spectra of figure 4–2 and 4–4. All other features of these figures are kept the same to have a meaningful comparison.

Despite their similarities, the transfer to ageostrophic modes in figure 4–4.a has some differences from the transfer to unbalanced modes in figure 4–2.a. The total transfer to T_A is slightly smaller than T_U, indicating that a part of the transfer to imbalance comes from balanced ageostrophic modes. T_U is approximately equal to $T_{BB\to U}$ at the early time in figure 4–2.a, whereas in figure 4–4.a there is a more distinguishable difference between T_A and $T_{GG\to A}$. This is because the argument behind (4.35) that proves the dominance of the **BBU** triads does not hold for the GGA triads since $A \neq 0$ initially. There is a noticeable amount of $T_{GA\to A}$ initially, which can be associated with transfer from balanced ageostrophic to unbalanced ageostrophic modes knowing that transfers due to **BUU** are negligible initially. At later times, $T_{BB\to U}$ and $T_{BU\to U}$ peak at two different wavenumbers in figure 4–2.b demonstrating the importance of **BBU** and **BUU** in different ranges of scales. This cannot be seen clearly in figure 4–4.b as $T_{GG\to A}$ is smaller than $T_{BB\to U}$ and does not have a distinct peak.

Transfers to balanced and geostrophic modes are alike at early time, as the panels 4–4.c and 4–2.c are the most similar ones. However, at later times the panels 4–4.d and 4–2.d are not as similar. As mentioned, T_B is almost equal to $T_{BB\to B}$ in figure 4–2.d showing that the balance flow becomes self-contained in time. Nonetheless, T_G is not equal to $T_{GG\to G}$ implying that the geostrophic and ageostrophic modes interact with each other while the properly defined balanced and unbalanced modes do not. This highlights that the balanced ageostrophic motion plays a non-negligible role in the new state of balance that the flow reaches in time. Recalling that in QG dynamics geostrophic modes are not affected by ageostrophic modes, this non-negligible role of ageostrophic modes shows that the flow adjusts to higher-order balance.

4.5.4 The effect of Rossby number

Previous studies such as Nadiga (2014) and Kafiabad & Bartello (2016) observed that the kink of the energy spectrum moves to smaller scales as Ro is decreased. In other words, the scale at which balance breaks down depends on Ro. Kafiabad & Bartello (2017) investigated this dependence further and showed that the wavenumber at which E_B and E_U cross scales with Ro^{-2}. This may raise the question of whether the peak of T_U depends on Ro as well.

To address this question, we present the spectrum of transfer to imbalance for four different Rossby numbers in the left panel of figure 4–5. The corresponding energy spectra are also plotted in the right panel. The transfer spectra are calculated at the beginning of the integration interval, as the most significant energy transfer takes place early in the simulations. It leads to a more shallow tail of $E(k_h)$ later, as shown in the right panel. In agreement with previous studies such as Nadiga

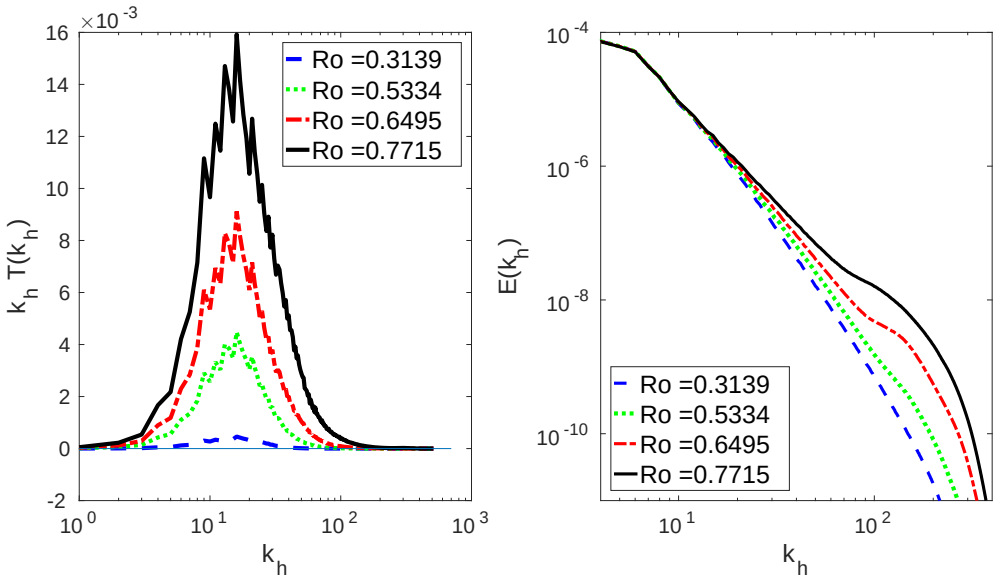


Figure 4–5: Left panel: $T_U(k_h)$ averaged over the time interval of [0, 0.1] for $Ro = 0.314$, 0.533, 0.650 and 0.772 marked in the legend (simulations: B1536R.314, B1536R.533, B1536R.650 and B1536R.772) . Right panel: total horizontal energy spectra of the Rossby numbers shown in the left panel at $t = 0.5$. Only a part of the energy spectra are shown to magnify the differences between the kinks at different Ro

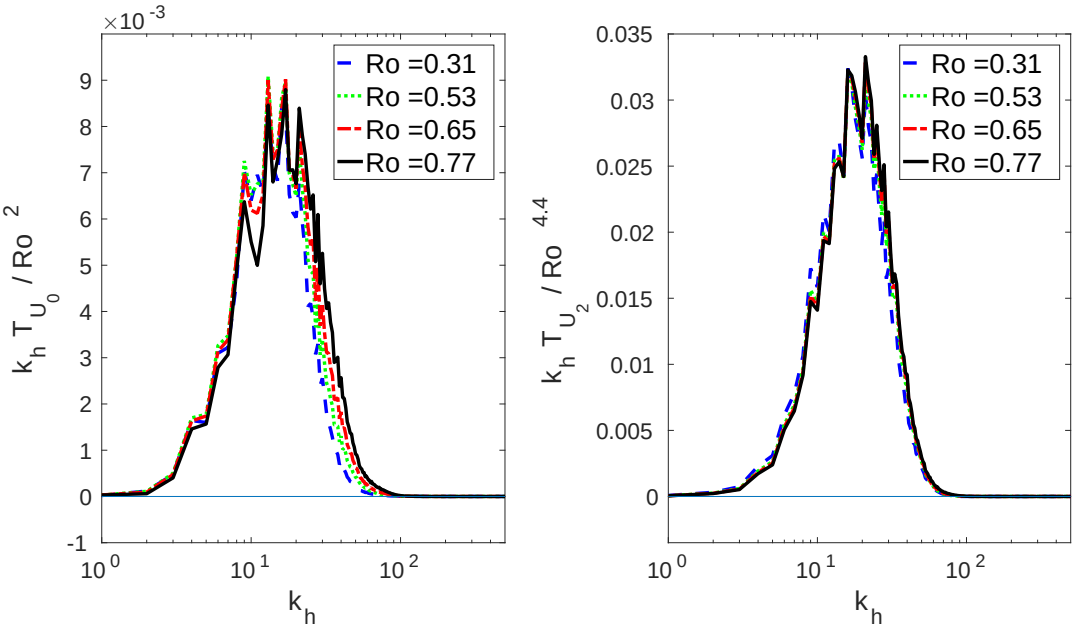


Figure 4–6: Left panel: $k_h T_{U_0}(k_h)$ divided by Ro^2. Right panel: $k_h T_{U_2}(k_h)$ divided by $Ro^{4.4}$ Each curve corresponds to the initial Rossby number shown in the legend and all the spectra are averaged over the interval of $[0,\ 0.02\tau]$.

(2014) and Kafiabad & Bartello (2016), the kink of the energy spectrum changes considerably with Ro for these simulations, as does the amplitude of transfer from balanced to unbalanced components. However, T_U peaks at the same wavenumber ($k_h \simeq 20$) for all Rossby numbers. Only the magnitude of the maximum T_U increases as rotation weakens. This demonstrates that the scale at which the balanced energy flows into the unbalanced motion is invariant to the strength of rotation. This also implies that the breakdown of balance begins in the larger scales where Ro and Fr are small.

As will be discussed in §4.5.6, the unbalanced modes can contain IGWs as well as other dynamics. However, after investigating the frequency spectra of the scales

at which T_U peaks initially ($k_h \simeq 20$) , we find that the unbalanced modes are predominantly IGWs at these wavenumbers. Therefore, by studying the initial T_U, we derive how the generated IGWs scale with Ro. Figure 4–6 portrays the unbalanced transfer of four different Rossby numbers scaled with Ro^n for two different orders of balance, where n is the best-fit exponent. The left panel indicates that the transfer at zeroth order (which is merely transfer to ageostrophic modes) scales with Ro^2, and the transfer at second order with $Ro^{4.4}$. Although figure 4–1 implies that the wavenumber at which T_U peaks does not change considerably with the order of balance, the results of figure 4–6 demonstrates that the scaling of its magnitude does change substantially. As will be shown in §4.5.6, ageostrophic modes can have a balanced part that behaves like geostrophic modes. Therefore, a substantial part of T_{U_0} does not display wave-like dynamics. Higher orders of balance such as the right panel of figure 4–6 are more indicative of IGW generation. In both panels of this figure, the curves have almost the same profile after scaling. The reason behind the scaling on the right panel ($Ro^{4.4}$) is an open question to the authors.

Vanneste & Yavneh (2004) presented an example of IGW generation from a particular balanced IC. They showed numerically and analytically that the amplitude of IGWs scale exponentially with the inverse Rossby number. Several other studies, such as Vanneste (2013) and Nadiga (2014), reported similar scalings for low-order dynamics or particular solutions of the primitive equations. Unlike the examples of Vanneste (2013), the scaling of T_U with Ro in figure 4–6 is power-law and not exponential. This is mainly because the turbulent flow has a spectrum of scale-dependent Ro (and Fr) numbers as opposed to one distinct value. Vanneste

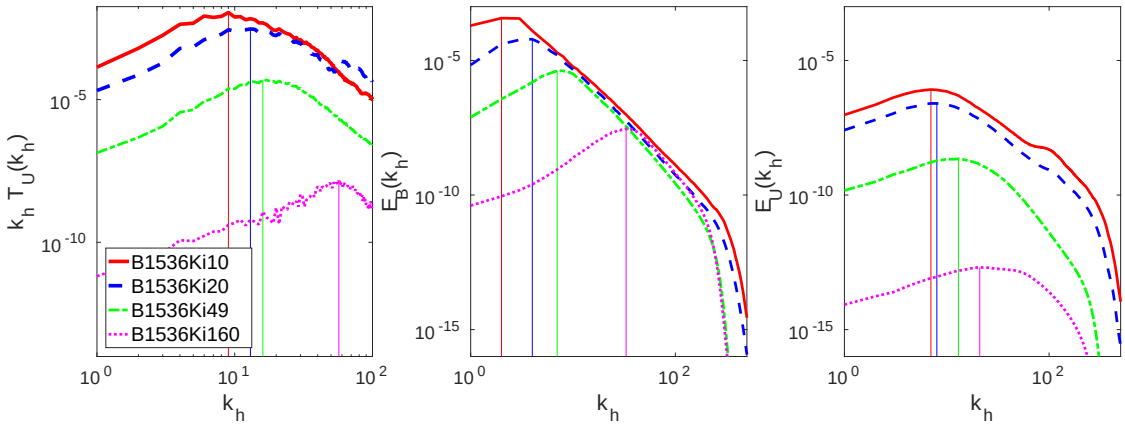


Figure 4–7: The spectra of transfer to imbalance (left panel), balance energy (middle panel), and unbalanced energy spectra (right panel) for simulations B1536Ki10, B1536Ki20, B1536Ki49 and B1536Ki160. The energy spectra are calculated at $t = 2.5$ eddy turnover times, and the transfer spectra at $t = 0.5$ eddy turnover times. The vertical lines mark the discrete wavenumber at which each curve reaches their maximum.

(2013) himself expressed that "the power-law wavenumber and frequency spectra of turbulence suggest that the exponential dependence of IGW amplitudes on ϵ found in smooth flows is replaced by a power-law dependence". Nadiga (2014) studied the evolution a baroclinic wave in a high-resolution Boussinesq model and observed the exponential scaling as well. His IC had only limited number of geostrophic modes at zero-order balance (geostrophic thermal-wind balance), whereas ours is a broad distribution that is the output of a QG turbulence simulation at second-order balance.

4.5.5 The effect of the deformation radius

It can be inferred from figure 4–5 that imbalance generation depends mostly on the larger scales that are predominantly balanced. One of the characteristic parameters of the large-scale dynamics is the wavenumber at which the horizontal

energy spectrum peaks. In our simulations, the inverse of this wavenumber may be shown to be proportional to the Rossby radius of deformation, L_R. At large scales, our simulations are predominantly QG, where $L/H \sim N/f$ (Charney, 1949). Therefore,

$$L_R = L \sim \frac{NH}{f}, \tag{4.36}$$

To derive initial conditions with different deformation radii, we performed different preliminary QG simulations. These decaying QG simulations start with a distribution of energy oproportional to $e^{(k-k_i)^2}$. Simulations B1536Ki10, B1536Ki20, B1536Ki49 and B1536Ki160 are based on ICs produced by the QG runs with $k_i =$ 10, 20, 49, 160, respectively. After running the QG models for $t = 5\tau$, these simulations reach a fully-developed state and the peak of the energy spectrum moves toward smaller wavenumbers. Then, we derived the balanced ageostrophic modes associated with the output of the QG simulations and use them as ICs in B1536Ki10, B1536Ki20, B1536Ki49 and B1536Ki160. We adjust f such that the Rossby number based on initial geostrophic vertical vorticity remains the same ($Ro = 0.5$).

The middle panel in figure 4–7 portrays $E_B(k_h)$ of these simulations. As was the goal, each curve has a peak at a different wavenumber. The initial $T_U(k_h)$ depicted in the left panel of this figure has the same sequence of peaks as those of the balanced energy. As discussed in §4.5.3, the initial T_U is mostly the transfer from balance to imbalance. Therefore, it can be concluded that by decreasing the Rossby radius of deformation, the scale at which imbalance is generated decreases as well. For the unbalanced energy spectrum in the right panel, the maxima occur in the same sequence as those of the other panels. Knowing that the unbalanced energy is

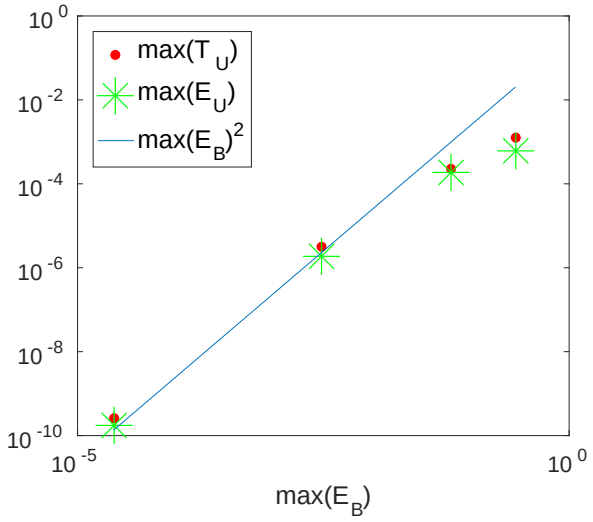


Figure 4–8: The maximum values of $T_U(k_h)$ and $E_U(k_h)$ as functions of maximum values of $E_B(k_h)$ for the curves depicted in figure 4–7. The solid blue line represents the slope of 2.

formed by the initial transfer from the balanced modes, it is expected that the right and left panels of figure 4–7 have the same sequence of maxima. Considering the similar sequences of maxima for T_U, E_B and E_U, one may try to quantify how these wavenumbers scale with respect to each other. This is hard to do in our simulations for two reasons. First, the energy spectrum peaks at very low wavenumbers for some of them. Since these wavenumbers are discrete values, the difference between them cannot be measured accurately. Second, we are suspicious of the spurious effect of the box size in B1536Ki10 and B1536Ki20, because the peak of their energy spectrum is close to the scale of the domain. Further investigation of the relationship between the wavenumbers of maximum T_U, E_B and E_U requires employing a larger computational domain. The scaling of the maximum values of these functions, on the other hand, can be studied. Figure 4–8 displays the maximum values of T_U and

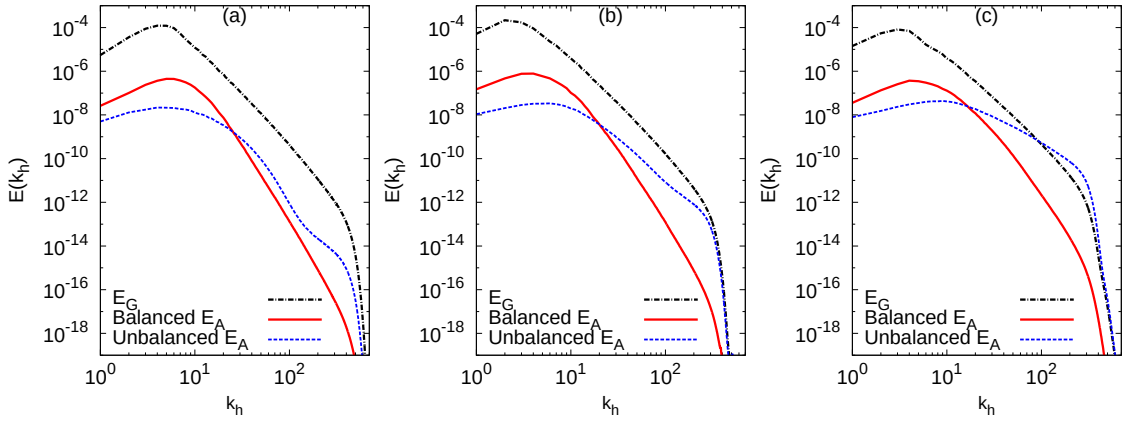


Figure 4–9: E_G, E_{A^b} and E_{A^u} at $Ro = 0.58$ after 27 eddy turnover times for a) $N/f = 32$ (resolution $2048^2 \times 64$), b) $N/f = 16$ (resolution $2048^2 \times 128$) and c) $N/f = 8$ (resolution $2048^2 \times 256$).

E_U as a function of maximum E_B for the corresponding simulation. The maxima of T_U and E_U lie on top of each other, which indicates that the peak of the unbalanced energy depends only on the initial transfer to imbalance. Note that T_U is measured at the beginning of the integration interval. We looked at the maxima of E_U at later times and the scalings of figure 4–8 remained unchanged. The maxima of T_U and E_U are proportional to E_B^2. This scaling holds more accurately for B1536Ki49 and B1536Ki160 than for B1536Ki10 and B1536Ki20.

In the simulations that are presented so far, N/f has been kept the same and equal to 8. However, this ratio is measured to be around 30 in the ocean and 100 in the atmosphere. To project the results to higher N/f, we present two other sets of simulations with $N/f = 16$ and 32. The simulations B2048N/f8, B2048N/f16 and B2048N/f32 in table 4–1 represent $N/f = 8$, 16 and 32, respectively. As explained in §4.4, the aspect ratio of the domain has to change proportionally to N/f to keep

the grid points equally spaced. Hence, to produce ICs for each simulation with different N/f, different QG simulations are performed. The outputs of these QG simulations have slightly different levels of energy, which leads to a small modification of f to keep the Rossby number the same (see table 4–1). Since f is kept similar, N changes substantially, leading to a smaller Froude number at higher N/f. The nearly geostrophic balanced flow at large scales is characterised by small Ro and Fr numbers. If Fr is smaller for a flow, balance is expected to prevail up to smaller scales. Therefore, as shown in figure 4–9 the shallow tail of the energy spectrum develops at smaller scales when N/f is increased.

4.5.6 Time scales and frequency spectra

As mentioned in §4.3.2, the concept of balance is entwined with the separation of time scales. The filtering of the fast motion in balance dynamics is possible only when there is a considerable disparity between time scales. At smaller scales, the time scale decreases in isotropic turbulence. Therefore, the turbulent time scale changes across length scales. In this section, we study the time scales of balanced and unbalanced modes at different wavenumbers to shed more light on the interaction between them and whether they can be viewed as quasi-linear modes. To that end, we examine the frequency spectra of geostrophic as well as balanced and unbalanced ageostrophic modes for different wavenumbers. In addition to these frequency spectra, we study the (spatial) spectrum of characteristic frequencies.

The normal mode decomposition that was reviewed in §4.3.1 has been extensively used in geophysical fluid dynamics. This decomposition, as explained, is based on the linear part of the equations. The ageostrophic modes defined there are also

frequently referred to as *wave modes* (e.g. Lelong & Riley, 1991; Bühler *et al.*, 2014). This is because they exhibit wave dynamics if the nonlinear terms are neglected, not because they have wave-like properties in the fully nonlinear context. This gives rise to the question of the extent to which these modes behave like IGWs. The answer, of course, depends on how small the nonlinear terms are compared to the linear terms in different regimes. The recent applications of normal mode decomposition to observed data raises the importance of addressing this question. More specifically, Bühler *et al.* (2014) suggested a method to apply the wave-vortex (geostrophic-ageostrophic) decomposition to the one-dimensional spectra of the observed horizontal velocity and buoyancy. They applied their method to ship-track (Bühler *et al.*, 2014) and aircraft (Callies *et al.*, 2014, 2016) data. They concluded that the shallow part of the atmospheric mesoscale spectrum is dominated by IGWs as the energy spectrum of the wave modes exceeds the energy spectrum of the vortical modes in these scales. Without concluding that the mesoscale spectrum is dominated by IGWs, idealized turbulence simulations also reported that linear wave modes have larger amplitudes than vortical modes do in this shallow range (e.g. Bartello, 1995, 2010; Deusebio *et al.*, 2013). Using the Weather Research and Forecasting (WRF) model, Waite & Snyder (2009, 2013) also investigated the shallow part of the mesoscale spectrum from a turbulence perspective by simulating an idealized baroclinic wave. In their dry simulations, they observed that in the upper troposphere the rotational energy is much larger than the divergent energy. However, by including moist processes, they reported that the divergent energy increases. Knowing that only wave modes have a divergent part, they expressed that "the divergence field in physical space

is consistent with IGWs being generated in regions of latent heating". Lindborg (2015) argued that if IGWs prevail, in theory the ratio of rotational to divergent structure functions must be smaller than unity. However, he showed that this ratio was somewhat larger than 1 for the observed data. Hence, he concluded that IGWs cannot be the principal explanation of the shallow mesoscale energy spectrum. In response, Callies *et al.* (2016) mentioned that structure functions suffer from the aliasing of large-scale rotational energy to smaller scales. Therefore, the rotational structure function in the mesoscale can have spuriously large values. Most recently, Asselin *et al.* (2017) employed a rapidly-varying stratification profile to account for the tropopause in their turbulence simulations. They performed a detailed frequency analysis and concluded that it is unlikely to observe wave-like dynamics in the shallow mesoscale range for the regime of intermediate and high Rossby numbers. However, they reported significant wave activity in the scales larger than the shallow range, where the slope of the spectrum was steep (close to -3). In light of these arguments, we found it insightful to thoroughly investigate the frequency spectrum of balanced and unbalanced modes. This analysis can delineate at which scale and to what extent wave-like dynamics are present. Knowing the limitations of our idealized simulations compared to the real ocean and atmosphere, it is hoped this frequency analysis adds to the picture drawn by the aforementioned studies.

We use horizontal isotropy of the flow to ensure the frequency spectra of figures 4–11 and 4–12 are statistically reliable. In so doing, we save the time series of the complex variables $G_{\mathbf{k}}$, $A^b_{\mathbf{k}}$ and $A^u_{\mathbf{k}}$ for all wavenumbers that lie on half rings ($k_x > 0$) with particular values of k_h and k_z. The set of $\{5, 10, 25, 97, 197\}$ is selected for

horizontal wavenumbers and marked by vertical lines in figure 4–10. These values are chosen such that $(|k_x|, |k_y|, |k_h|)$ can form pythagorean triples to increase the number of discrete wavenumbers on each half ring. Then, we take the Fourier transform of these time series and average them over all the modes with the same k_h and k_z. We implement this procedure on the real and imaginary parts of these modes to double the number of samples. This leads to 12 independent modes overall for $k_h \in \{5, 10, 97, 197\}$ and 20 modes for $k_h = 25$. For each mode, 9650 points are sampled from $t = 0$ to $t = 3.14\tau$, and the Hann window is employed to minimize the spectral leakage due to the non-periodicity of the signals (Oppenheim, 1999). We carried out our analysis for barotropic modes as well as a baroclinic mode ($k_z = 3 \times N/f = 24$) which is marked in the inset of figure 4–10.

We illustrate how the time scales of different regimes of flow change by considering the frequency spectra of two Rossby numbers at different scales. E_G, E_{A^b} and E_{A^u} of these Rossby numbers are shown in figure 4–10 to clarify which regime of the flow is analysed. At the low Rossby number, E_G and E_A do not cross each other in the resolved range of scales. However, E_{A^b} and E_{A^u} cross at $k_h \simeq 60$. Below this wavenumber, E_{A^b} is larger than E_{A^u}, marking the regime of balance. At $Ro = 0.579$, the unbalanced modes exceed the balanced modes for wavenumbers $k_h > 37$ which marks the regime that contains some imbalance. From the inset plots, it is clear that the difference between balanced and unbalanced modes decreases when k_z is increased from 0 to 24.

Figure 4–11 displays the frequency power spectrum of the geostrophic as well as the balanced and unbalanced ageostrophic modes. The top left panel represents the

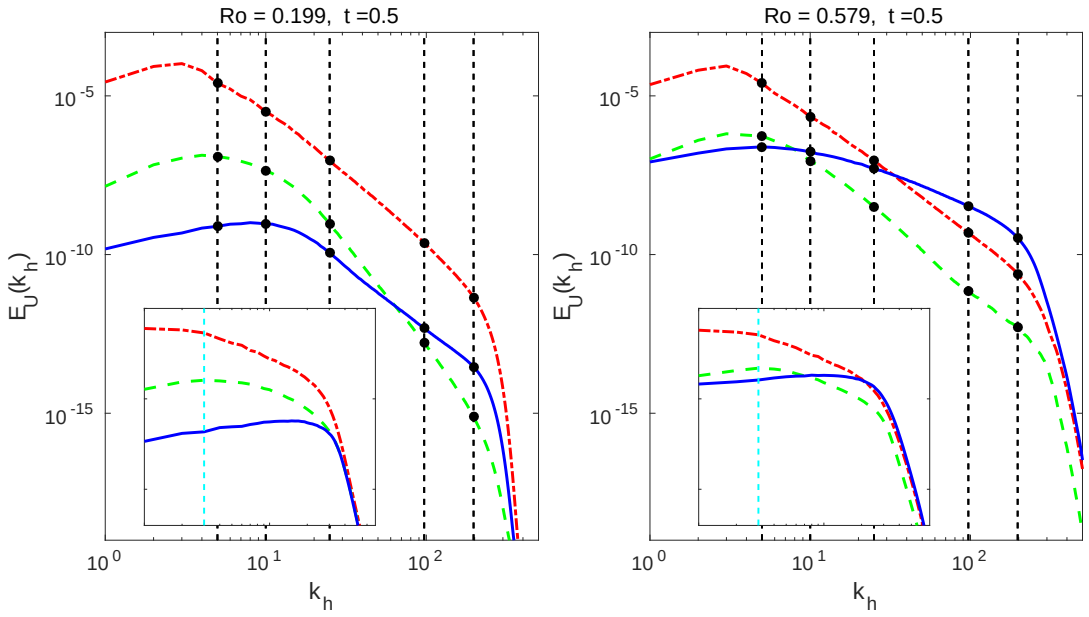


Figure 4–10: The geostrophic (red dash-dotted curve), balanced ageostrophic (dashed green curves) and unbalanced ageostrophic (sold blue curves) energy spectra of simulations with $Ro = 0.199$ (left plot) and $Ro = 0.579$ (right plot). The main plots are horizontal energy spectra and the insets are vertical energy spectra. The dashed vertical lines show the set of k_h and k_z for which the time series are saved and the frequency spectra are calculated.

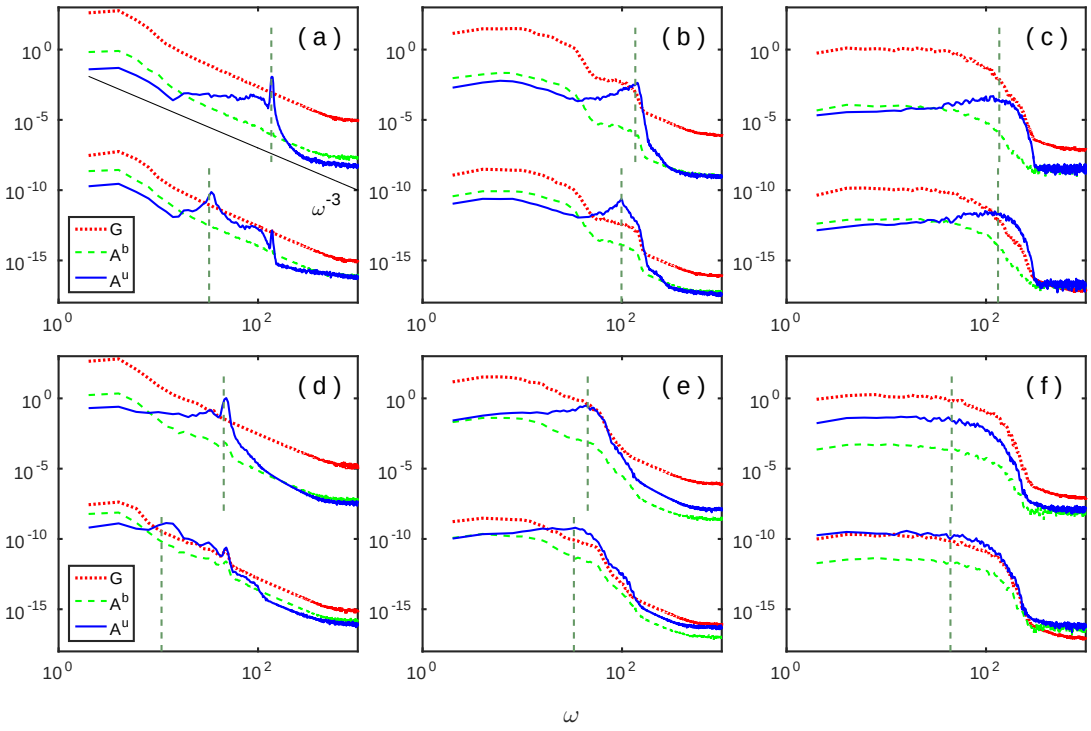


Figure 4–11: The geostrophic (red dotted), balanced ageostrophic (green dashed) and unbalanced ageostrophic (solid blue) frequency spectra of modes with fixed k_h and k_z. $k_h = 5$ in the left panels ((a) and (d)), $k_h = 25$ in the middle panels ((b) and (e)) and $k_h = 97$ in the right panels ((c) and (f)). The baroclinic modes ($k_z = 24$) are shifted down (multiplied by 10^{-10}) for better visualization; hence, they lie below the barotropic modes ($k_z = 0$). The top row ((a), (b) and (c)) corresponds to the simulation with $Ro = 0.199$ and the bottom row ((d), (e) and (f)) with $Ro = 0.579$. Dashed vertical lines are the linear frequency of each mode.

regime of low k_h and small Rossby number. At the low frequencies of this panel, the amplitude of balanced modes are higher than the unbalanced modes, which could be inferred from the energy spectra of figure 4–10 as well. However, close to the linear frequency, the amplitude of the unbalanced modes exceeds that of the balanced modes. Considering that the maximum amplitude of G (at low frequencies) is still much larger than the peak of A^u at $\sigma_{\mathbf{k}}$, we expect wave dynamics to be overshadowed by geostrophic motion in this regime. Excluding the balanced part, A^u displays a sharp peak at low Ro and low k_h, signalling that the unbalanced motion in this regime behaves like linear IGWs. This is quite expected as the scale-dependent Ro is very low at this small wavenumber. Hence, nonlinear interactions are weak. E_G and E_{A^b}, on the other hand, show similar behaviour as the balanced ageostrophic modes are approximately slaved to geostrophic modes according to the relationship (4.11). The other interesting feature of the low-Ro low-k_h regime is the slope of the geostrophic frequency spectrum, which is close to -3. This can be explained by Taylor's hypothesis for homogeneous statistically stationary turbulence, which demonstrates that the temporal correlation of a signal can be identified with its spatial correlation, if the flow is statistically stationary and homogeneous. Therefore, the Fourier transform of the temporal correlation should be scaled with frequency in a similar way as the spatial correlation is scaled with wavenumber.

The spectra of $k_h = 25$ delineate the role of IGWs in the breakdown of balance. At this wavenumber, the amplitude of unbalanced modes increases more significantly from $\omega = 36$ to $\omega = 136$. Over this range of frequencies, the geostrophic amplitude plateaus. Therefore, unlike the corresponding spectrum at $k_h = 5$, it is shallower and

no longer shows the -3 slope. This is due to the interaction between the balanced and unbalanced modes at these frequencies and indicates that the geostrophic time scale is being influenced by the wave time scale. This influence leads to further amplification of geostrophic modes close to the linear frequency at $k_h = 97$. At this wavenumber, the balanced modes have no distinct slow time scale since the geostrophic curve stays rather flat from the very low frequencies to the linear frequency. Knowing that at this Ro the total energy spectrum does not display a shallow tail, one can conclude that the separation of time scales breaks at wavenumbers smaller than the kink of the energy spectrum.

By comparing the corresponding results for the two Rossby numbers (top and bottom rows of figure 4–11), we find that the peaks of A^u at $Ro = 0.579$ are less sharp but have higher amplitudes. In other words, at higher Rossby numbers the wave modes are stronger but also interact more with the other modes, and hence behave less linearly. For example, the peak of the barotropic A^u has the amplitude of 0.012 at $Ro = 0.199$ and 1.053 at $Ro = 0.579$. The former peak is also much spikier than the latter.

The change of frequency spectra across different scales can better be seen in figure 4–12, where more horizontal wavenumbers are displayed together. The frequency spectra of geostrophic and ageostrophic modes have been previously studied by Deusebio *et al.* (2013) and Kafiabad & Bartello (2017). The former used forcing for geostrophic modes, and the latter performed decaying simulations with nonlinearly balanced ICs. Most features of figure 4–12 qualitatively agree with the results

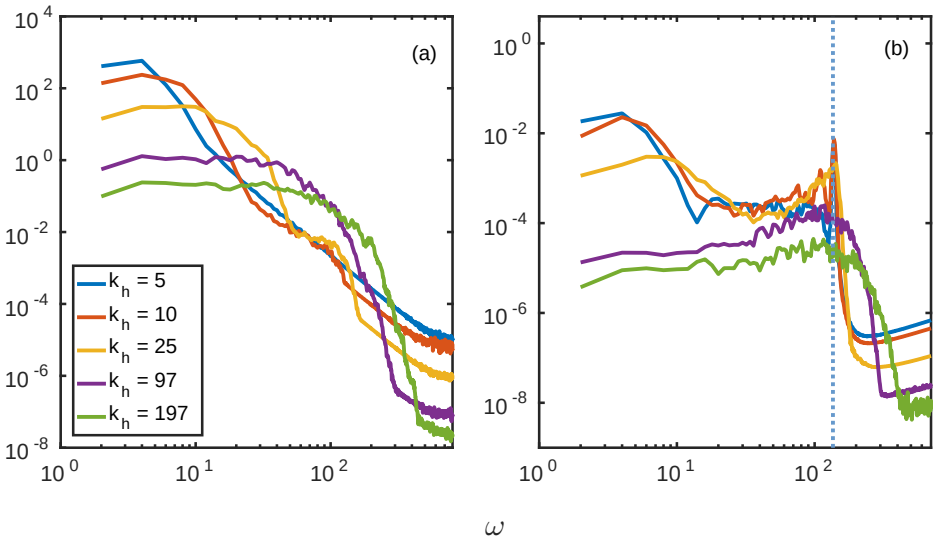


Figure 4–12: The frequency spectra of the geostrophic (left) and the unbalanced ageostrophic (right) modes for $k_h \in \{5, 10, 25, 97, 197\}$. $Ro = 0.199$.

of the aforementioned studies. By increasing k_h, the peak of A^u at the linear frequency spreads out to the extent that at $k_h = 97$ and $k_h = 197$ there is no distinct peak in the unbalanced frequency spectrum. Hence, it can be concluded that at the shallow range of the energy spectrum, even the unbalanced modes do not exhibit wave-like dynamics. At $k_h = 5$, there is a distinct peak in geostrophic spectra, which corresponds to the slow geostrophic time scale. However, by increasing k_h, the large amplitudes of the geostrophic spectra spread toward higher frequencies. Therefore, these curves become nearly flat over a wide range of frequencies, confirming that the time scale disparity loses its meaning at smaller scales. The other interesting features of the geostrophic curves are the secondary plateaus at $k_h = 10$ and $k_h = 25$, which appear after a drop from the largest amplitudes. As discussed, these ranges coincide

with the peaks of the unbalanced curves close to the linear frequency and reflect the interactions between balanced and unbalanced modes.

Parts of the trends observed across the different scales in figure 4–12 corroborate the results of Asselin *et al.* (2017). These authors also investigated the level of wave amplitude for the unbalanced motions at different wavenumbers. Due to the complications of their geometry and their stratification profile, they could not employ the geostrophic-ageostrophic decomposition or higher-order definitions of balance used in here. Instead they employed high-pass filtering of the superinertial frequencies ($\omega > f$) to derive their high-frequency modes. They examined the time-lagged correlation between the velocity components of these high-pass-filtered modes to investigate the level of wave activity. They concluded that even the unbalanced modes show less and less wave behaviour as k_h increases.

Considering the peaks of A^u at different k_h in figure 4–11, one can conclude that they are spikier at lower horizontal wavenumbers. There are also distributions of frequencies around these peaks where the amplitude of A^u rises and has higher values compared to the neighbouring frequencies. Although at lower k_h the peaks are spikier, the distributions around them are wider. For instance, for the barotropic and baroclinic modes at $k_h = 5$ and $Ro = 0.199$ in figure 4–11, the spectrum rises at $\omega = 14$ and falls right after the linear frequency ($\omega = N = 136$). The distribution is narrower at $k_h = 25$ and the same Ro, starting from $\omega = 36$ and ending at $\omega = 136$. One can also observe that barotropic curves drop sharply at the frequencies larger than the linear frequency, while the baroclinic curves spread out. This is likely due to the disparity between the linear frequencies of different modes with different k_z.

More specifically, changing k_z at low k_h affects the linear frequency more than a similar change at higher k_h. If we change k_z from zero to Δk_z, according to 4.24 the linear frequency changes from N to

$$\sigma^*_{\boldsymbol{k}} = f\sqrt{1 + \frac{N^2/f^2 - 1}{1 + \Delta {k_z}^2/k_h^2}}. \tag{4.37}$$

If $\Delta {k_z}^2/k_h^2$ is small, $\sigma^*_{\boldsymbol{k}}$ stays closer to N. Larger values of k_h keep $\Delta {k_z}^2/k_h^2$ small. Hence, the linear frequency gets less affected by changing k_z at larger k_h. As a result, in figure 4–11 the dashed vertical lines of the barotropic and baroclinic modes get closer to one another as k_h increases. According to the locality assumption, triad interactions are more effective when their wavenumbers are close to each other. If this assumption holds in our simulations, at low k_h, the local wavenumbers have a wider range of frequencies, and their nonlinear interactions can widen the distribution of high-amplitude A^u. The secondary peaks at the baroclinic modes of the left column in figure 4–11 may also corroborate the reasoning above, since they occur close to $\omega = N$.

From this analysis, we conclude that spontaneously generated IGWs are observed when: 1) *Ro* is large enough that the unbalanced ageostrophic modes are comparable to the balanced geostrophic modes, 2) balance still dominants imbalance such that there is a separation of time scales and nonlinear interactions are not too strong. These opposing conditions mark the regime in which wave dynamics are most sensibly observed. It occurs at atmospheric scales slightly larger than the kink of the energy spectrum but not too large such that unbalanced motion is overshadowed by balanced dynamics. This range of scales is similar to that at which the maximum

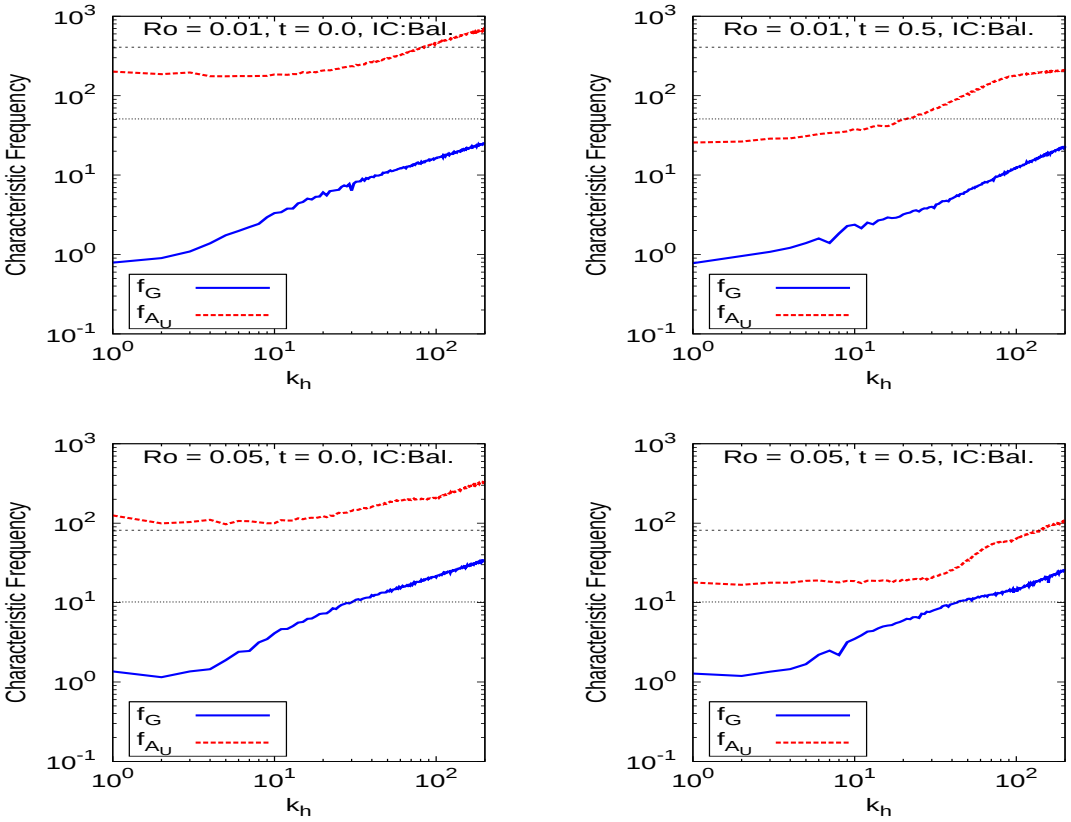


Figure 4–13: $f_G(k_h)$ and $f_U(k_h)$ of the Rossby numbers marked in the title of each panel. The left column corresponds to the beginning of integration time, and the right column to $t = 0.5\tau$

transfer between balance and imbalance occurs (discussed in §4.5.3). Although the scale range of maximum wave energy is relatively narrow, it plays an important role in the breakdown of balance. This breakdown of balance subsequently leads to development of shallow spectra at smaller scales. Another conclusion that can be made is wave dynamics are more pronounced in large-scale baroclinic modes than barotropic modes since their unbalanced contributions have higher amplitudes.

The time scale of the balanced and unbalanced modes changes in time and across spatial scales. To draw a picture of how this change takes place, we measure the following characteristic frequencies of our simulations

$$f_G(k_h) = \left[\sum_{|\mathbf{k}'-\mathbf{k}'\cdot\hat{z}|=k_h} \frac{dG_{\mathbf{k}'}/dt}{G_{\mathbf{k}'}} \right] / N_G(k_h) \tag{4.38a}$$

$$f_{A_U}(k_h) = \left[\sum_{|\mathbf{k}'-\mathbf{k}'\cdot\hat{z}|=k_h} \frac{dA^u_{\mathbf{k}'}/dt}{A^u_{\mathbf{k}'}} \right] / N_U(k_h), \tag{4.38b}$$

in which $N_G(k_h)$ and $N_U(k_h)$ are the number of modes which are summed for each k_h. Since the balanced flow is dominated by geostrophic modes, we plot the characteristic frequencies of these modes as representative of the balanced frequencies. This approach might not be as rigorous as the frequency spectrum analysis of figures 4–11 and 4–12 to quantify the time scales of each mode. Nevertheless, it gives a good qualitative picture of how the characteristic time scales change in time across many (spatial) scales.

$f_G(k_h)$ and $f_U(k_h)$ are depicted for two different Rossby numbers at $t = 0$ and $t = 0.5\tau$ in figure 4–13. Initially, there is a separation between the balanced and unbalanced time scales for both Rossby numbers due to NNMI. At the lower Rossby number the separation of time scales is maintained, whereas the balanced and unbalanced frequencies get closer to each other at the higher Rossby number. Figure 4–13 also demonstrates that the ratio of f_G to f_U decreases as k_h increases. The tapering of this ratio shows that NNMI is expectedly not very effective in the smaller scales. At the higher Ro at $t = 0.5\tau$, there is a substantial distance between

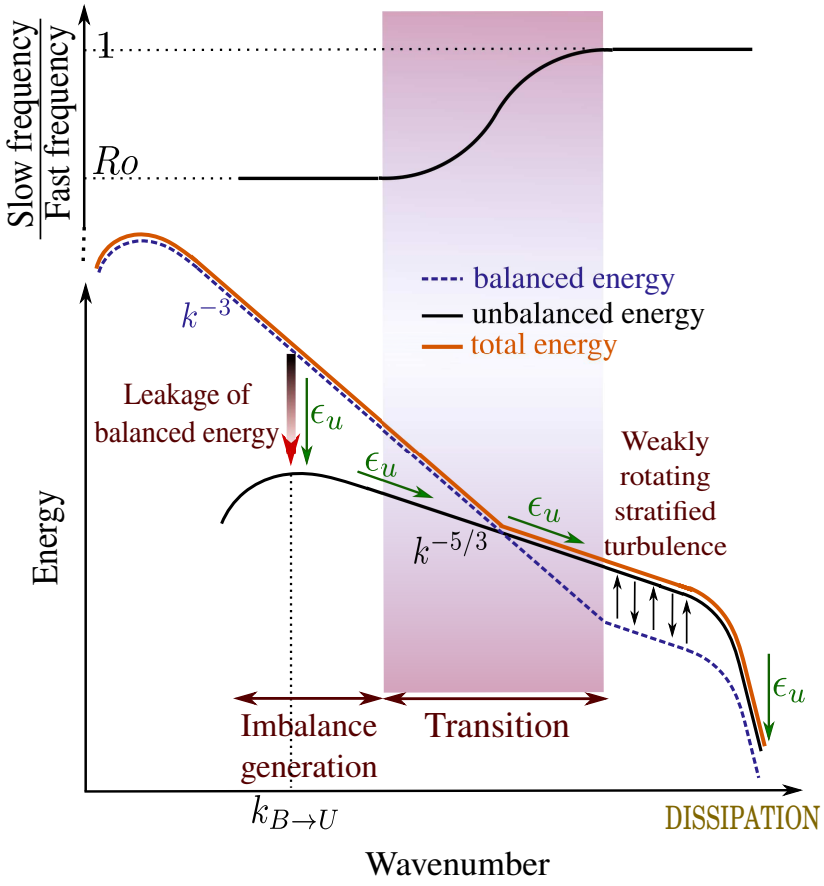


Figure 4–14: The schematic picture of the break in spectral scaling due to the breakdown of balance.

f_G and f_U at large scales despite the proximity of f_G and f_U at smaller scales, signalling that balance is maintained there.

4.6 Conclusion

Employing high-order balance, we elaborated on an internal mechanism that transfers energy from the balanced to the unbalanced part of a turbulent rotating stratified flow. This transfer of energy leads to a change of spectral scaling from k_h^{-3} to $k_h^{-5/3}$. This mechanism is summarised in figure 4–14, where the energy spectra

and the ratio of characteristic slow to fast frequencies are schematically portrayed. The total energy spectrum, which is the sum of the balanced and unbalanced energy spectra, has two power-law subranges. The large-scale subrange has a steep slope of -3 and is dominated by balanced energy. The small-scale subrange, on the other hand, is predominantly unbalanced and displays a shallow slope of -5/3. At the small-scale end of the steeper subrange, the balanced energy leaks into the unbalanced motion. The wavenumber at which the maximum transfer from balance to imbalance occurs is marked by $k_{B\to U}$ in figure 4–14. We showed that $k_{B\to U}$ is slightly larger than the peak of the balanced energy spectrum, indicating that balance breaks down at scales where the energy spectrum is still steep. This fact stems from the decreasing of the difference between the fast and slow time scales which starts at these scales according to our frequency analysis. As shown in figure 4–14, the ratio of characteristic slow to fast frequencies is equal to Ro at small wavenumbers and starts increasing at a wavenumber larger than $k_{B\to U}$. Therefore, there is a range of wavenumbers larger than $k_{B\to U}$ that the separation of time scales is somewhat maintained. This range of wavenumbers is marked as imbalance generation in figure 4–14 and has several characteristics. First, the fast frequency is approximately equal to the linear frequency of IGWs in this range, according to our frequency analysis in §4.5.6. Hence, one may conclude that the leaked balanced energy is mostly transferred to IGWs and they play an important role in the onset of balance breakdown. Second, once the IGWs are generated, resonant catalytic interactions likely enhance the forward cascade of energy in this range. In a resonant catalytic interaction, a

balanced mode facilitates the transfer of energy between two opposite-signed unbalanced modes and remains unchanged, as described by Bartello (1995). In the range of Rossby numbers investigated in this study, $k_{B\to U}$ is found to be dependent on the Rossby radius of deformation but invariant to a change of Rossby number. Unlike this wavenumber, the flux of energy from balance to imbalance (marked by ϵ_u in figure 4–14) depends sensitively on Ro (as also reported by Vallgren *et al.*, 2011). This flux is directed toward smaller scales and forms a power-law scaling of $k_h^{-5/3}$. At very small scales, in addition to the unbalanced spectrum, the tail of the balanced spectrum becomes shallow as well. At these scales, the balanced and unbalanced modes exchange energy, as the fast and slow time scales become the same and the flow becomes weakly rotating stratified turbulence. Therefore, the energy tends to equipartition among different modes (see Billant & Chomaz, 2001; Lindborg, 2006, for discussing equipartition of energy in stratified turbulence). There is a transition range of wavenumbers marked in figure 4–14 where the difference between the fast and the slow time scales decreases until they merge into a single turbulent time scale. In this range, the unbalanced modes show less and less wave-like behaviour and their nonlinear interactions with each other increase.

There have been previous studies relying on linear balance to decompose rotating stratified flows to a set of geostrophic and ageostrophic modes (Bartello, 1995; Ngan *et al.*, 2008; Bartello, 2010; Deusebio *et al.*, 2013). In this study, we extended the balance/imbalance decomposition to higher orders. We found that many results of low- and high-order balance decompositions are qualitatively similar. Nevertheless, they have quantitative differences that are listed below:

1. The most important difference between the generated ageostrophic energy and spontaneous imbalance is their scaling with Ro. Ageostrophic energy scales with Ro^2, whereas the scaling exponent is larger than 2 for spontaneous imbalance. At second-order balance, we found this exponent to be about 4.4. The scaling of ageostrophic energy can be explained by (4.16) which was discussed in §4.3.3. However, the reason behind the scalings of higher-order balance is considerably more subtle. The spontaneous generation of IGWs in low-order dynamical systems or special solutions of the fluid equations has been shown to scale inverse-exponentially with Ro (Vanneste & Yavneh, 2004; Nadiga, 2014). We, on the other hand, simulated turbulent balanced flows and considered imbalance generation in a broader sense. For these reasons, the exponential dependence of IGW on Ro was replaced by power-law dependence in our simulations, as speculated by Vanneste (2013). While we covered only a limited range of Rossby numbers (from $Ro = 0.3$ to $Ro = 0.8$), considering a broader range of Rossby numbers could make the scaling cleaner.
2. The energy transfer from geostrophic to ageostrophic modes reported by linear balance is higher than the energy transfer from balance to imbalance reported by high-order balance. This is because a part of geostrophic energy is transferred to balanced ageostrophic modes, which is rightly not accounted for in the transfer to imbalance at higher orders.
3. High-order balance displays a small kink in the ageostrophic energy spectrum, which is steep at low wavenumbers. Kafiabad & Bartello (2016) showed that the steep part of the ageostrophic energy spectrum is associated with balance

and the shallow part with imbalance. GCM studies such as Hamilton *et al.* (2008) and Evans *et al.* (2013) reported the divergent energy spectra display a kink as well, consistent with the likelihood that the large-scale dynamics of these models are nonlinearly balanced.

4. Imbalance generation takes place at early times. However, at later times as the Rossby number decays, the transfer between balance and imbalance reduces substantially. This transfer is less negligible for linear balance than high-order balance, which indicates that the flow adjusts to a new high-order state of balance.

It should be noted that the method employed for calculating spontaneous imbalance is justified if the flow is predominantly balanced. More particularly, the slave relation (4.11) is required to be a valid approximation. In hindsight, our results show that spontaneous imbalance occurs at scales where the scale-dependent Fr and Ro are low and the total energy spectrum is steep. Therefore, in this regime the flow is predominantly balanced. Moreover, it was observed that starting from an initially balanced condition, imbalance is mostly generated at early times, when the unbalanced shallow tail has not yet developed

The break of spectral scaling portrayed in this study is consistent with atmospheric observations (Nastrom & Gage, 1985; Lindborg, 1999). Nevertheless, several atmospheric phenomena are missing in our idealised simulations. Among these are the presence of external and internal boundaries (tropopause and topography) and convective processes. Hence, we cautiously avoid interpreting the results as the sole explanation of the mesoscale spectral scaling. Yet, the internal mechanism that was

delineated here may be a part of the explanation for the shallowing of the atmospheric (and the oceanic) spectrum. A different hypothesis was proposed by Gage (1979) and expanded by Lilly (1983) to explain the $k^{-5/3}$ scaling of the mesoscale. This hypothesis attributes the slope of the energy spectrum at mesoscales to the upscale (inverse) cascade of energy in quasi-2D stratified turbulence. The source of this energy was associated with small-scale convective processes by Lilly (1983). By contrast with this inverse-cascade scenario, we supported the hypothesis of a forward cascade of energy via unbalanced ageostrophic modes (see Lindborg, 2005, 2006; Brethouwer *et al.*, 2007; Vallgren *et al.*, 2011). Mesoscale meteorologists emphasise the necessity to consider convective sources, which may contribute to the scaling of the mesoscale energy spectrum concurrently (see e.g. Sun *et al.*, 2017). However, their comparative significance is not clear. As future work, properly-designed simulations can include the effects of both: the forward cascade of energy that is leaked from large-scale balance dynamics and the convective injection of unbalanced energy at smaller scales. GCM studies, which more realistically model large-scale balance dynamics, cannot resolve the convective scales. Hence, they rely on parametrization schemes to include these processes (e.g. Hamilton *et al.*, 2008). On the other hand, the mesoscale models that better resolve the convective scales cannot extend their large scales to synoptic scales to properly include balance dynamics. Therefore, idealised simulations might be more suitable to incorporate both balance dynamics and convective forcing in one simulation.

CHAPTER 5

Conclusion and future work

In this study, we have investigated how imbalance is generated spontaneously from a predominantly balanced flow. Starting from a high-order balanced initial condition, imbalance emerges at smaller scales if rotation is weak enough and Re is large enough. The energy of imbalance, however, comes from the large-scale balanced flow. We have shown that the signature of QG (and higher order) balance is an energy spectrum with a slope of -3 or steeper, while imbalance displays a slope of -5/3. Considering the simultaneous existence of balance and imbalance in our simulations, our energy spectrum has two subranges: a predominantly balanced subrange at large scales with a slope of -3 and an unbalanced subrange at smaller scales with a slope of -5/3. The transition from balance to imbalance is then marked by a kink in the energy spectrum. This picture is consistent with atmospheric observations (Nastrom & Gage, 1985; Lindborg, 1999) although much is lacking in our simulations.

We employ an initialisation scheme suggested by Baer & Tribbia (1977) in our turbulence simulations to produce a high-order balanced initial condition. This initialisation is implemented at second order in Ro, which is an order higher than QG.

With this high-order balanced initial condition, we calculated the slope of the balanced ageostrophic energy spectrum, which is shown to be steeper than that of the geostrophic spectrum. Hence, we conclude that balance occurs in conjunction with steep spectra. If Re is large enough, a shallow tail develops in the energy spectrum of the ageostrophic modes. At small Ro, this emerges at very small scales, but if Ro is increased, the transition moves upscale and starts at smaller wavenumbers. According to our results, the kink of the ageostrophic spectrum (i.e. the wavenumber at which the spectrum becomes -5/3) scales with Ro^{-2}. By implementing the Baer-Tribbia scheme at the end of the integration time, we can extract the balanced part of the ageostrophic modes. By subtracting the balanced part from the total, we calculate the unbalanced ageostrophic modes generated by the non-hydrostatic model. This decomposition of ageostrophic modes to balance and imbalance again reveals that the steep part of the ageostrophic spectrum is balanced and the shallow part is unbalanced. The change of slope in the ageostrophic energy spectrum had not been reported in other turbulence simulations, since at best they relied on linear balance in which the balanced modes are void of any ageostrophic dynamics (see e.g. Bartello, 2010; Deusebio *et al.*, 2013). However, recent high-resolution GCM models reported that the divergent energy has a spectrum with a steep part and a shallow tail (Hamilton *et al.*, 2008; Evans *et al.*, 2013). Knowing that in our simulations only the ageostrophic modes can contain divergent energy, these GCM results are consistent with ours, although we suspect them to suffer from insufficient vertical resolution.

If Ro and Re are small, the total energy spectrum stays predominantly geostrophic and, hence, maintains a steep slope. However, if they are large enough, the shallow tail of the ageostrophic energy spectrum may exceed the geostrophic spectrum. If this happens, the total energy spectrum, which is the sum of geostrophic and ageostrophic parts, will have a shallow tail as well. The fact that the ageostrophic energy spectrum consists of a steep and a shallow part implies that the kink of the total energy spectrum depends on the characteristics of balance at larger scales.

In chapter 4, we proposed a new method to derive the transfer between the balanced and unbalanced parts of the motion. The use of this method, unlike its predecessors, is not restricted to a particular order of balance and does not require balanced and unbalanced modes to be orthogonal. Applying it to our high-order balance model, we calculated the spectrum of energy transfer among balanced and unbalanced modes. By looking at the transfer spectra we found that a considerable amount of energy is initially transferred from balanced to unbalanced modes. The peak of this transfer occurs at a horizontal wavenumber larger than the peak of the energy spectrum but still within the subrange where the slope of the energy spectrum is steep. Later in time this energy is cascaded forward by the unbalanced modes at the large-scale end of the shallow subrange with no significant exchange of energy with the balanced modes. At the small-scale end of this subrange, however, all the modes (balanced and unbalanced) contribute to the forward cascade of energy as timescales become similar. By decomposing the transfer spectra to different triads, we find that, at scales slightly smaller than imbalance generation, triads with one balanced and two unbalanced modes most effectively interact. We then conclude that

the balanced modes facilitate the transfer of energy between two unbalanced modes and get slightly damped. This mechanism is similar to the catalytic interaction described in Bartello (1995). However, at much smaller scales where the shollow subrange emerges, the resonant catalytic interaction does not prevail since there is no separation of time scales. Over the range of low Rossby numbers examined, the scale at which maximum transfer occurs is invariant to a change of Ro. However, the magnitude of transfer from balance to imbalance increases at higher Rossby numbers. At second-order balance, this transfer scales approximately as $Ro^{4.4}$, and therefore it exceeds the balanced part at large Ro.

We have also analysed the frequency spectra of the balanced and unbalanced modes at different scales. At large scales, both the geostrophic and balanced ageostrophic modes peak at low frequencies, whereas the unbalanced modes peak at the linear frequency. This distinct separation of time scales confirms that balance can be maintained at these scales. As one considers smaller horizontal scales, these peaks become flatter and get closer to one another until they are identical. The scale at which they merge is slightly larger than the kink in the energy spectrum. This implies that the separation of time scales breaks down at scales larger than the kink of the energy spectrum.

Having the results of this research in mind, in the following section I describe two suggestions to expand the current work:

5.1 Dependence of the optimal order of balance on the Rossby number

For low-order systems, series-expansions are employed to derive balance models up to an arbitrary order of $O(\epsilon^N)$. However, it has been shown that these expansions

do not converge as $N \to \infty$ (Bender & Orszag, 1999; Vanneste, 2013). In other words, at a fixed ϵ there is an optimal N up to which the series converge and then start diverging. This optimal order is argued to depend on Ro. This result provokes the question of whether in turbulence setups there is an optimal order of balance which depends on Ro. Our unreported investigation of an iterative initialisation scheme raises the speculation of such a dependence. More specifically, we performed several simulations initialized with Machenhauer's (1977) iterative scheme. We found that at high and intermediate Rossby numbers imbalance generation decreases by performing more iterations up to an optimal number. More iterations, however, lead to the generation of more imbalance. It might be hard to draw a direct analogy between the orders of balance of the number of iterations in Machenhauer's scheme, but we speculate that the optimal order is Ro-dependent similar to the optimal number of iterations. To explore this question more rigorously, one needs to perform a hierarchy of initialisations at many different orders. In this study we managed to program and calculate balance only up to $O(\epsilon^2)$, as we found the study of higher orders computationally challenging. Higher computational power or hierarchical balance models that can be numerically calculated more efficiently may enable us to investigate how the optimal order of balance depends on the Rossby number.

5.2 The effect of convective forcing on the kink of the atmospheric spectrum and the mesoscale subrange

As mentioned in section 1.5, Gage (1979) proposed that the mesoscale energy spectrum is generated by the upscale cascade of energy in quasi-2D turbulence, similar to the inverse energy cascade subrange in the theory of 2D turbulence by Kraichnan (1967). Lilly (1983) expanded this hypothesis and associated the source of this

energy with atmospheric convection at very small scales (order of tens of kilometres). For the reasons discussed in section 1.5, simulations do not display 2D-like turbulence with an inverse cascade of energy as the sole explanation of $k^{-5/3}$ scaling in the mesoscale subrange.

It might not be the case that one of the hypotheses mentioned in section 1.5 exclusively accounts for the mesoscale scaling. Instead, different hypotheses can have a share in explaining this phenomenon. As Lilly (1983) expressed himself, "this is essentially a quantitative problem, perhaps both suitable for detailed numerical simulation and turbulence closure model integration." I believe Lilly's intuition and wisdom still hold after more than three decades. Well-designed turbulence simulations that encompass more realistic features can elucidate the contribution of different hypotheses behind the mesoscale spectrum.

We argued in this thesis that the slow leakage of balanced energy leads to the formation of a shallow tail in the energy spectrum. We explained that this shallow range is dominated by the unbalanced ageostrophic modes which cascade the leaked energy forward via a catalytic interaction with the geostrophic modes. Admittedly, we did not include moist processes in our simulations. This raises the question of the relative significance of the mechanism we described compared to the direct forcing of imbalance by small-scale convection. At first glance it would seem the injection of energy generated by convection would be rapidly dissipated, although not without affecting local stability.

It can be argued that even GCM/NWP models cannot fully resolve convective scales. At best these models employ moist convective parametrization schemes. For

instance, Hamilton *et al.* (2008) studied the effect of moist processes on the energy spectrum using the schemes suggested by Emanuel & Živković-Rothman (1999) and Arakawa & Schubert (1974). He reported that the use of one of these over the other can shift the energy spectrum in amplitude. The insufficiency of these models' vertical resolution may be underlined by knowing that their effective resolution is much lower than their nominal resolution since they are usually based on finite-difference schemes and suffer from nonlinear aliasing instability. For example, Skamarock (2004) illustrated that for his finest horizontal resolution of 4 km, the effective resolution is at least 28 km. As will be discussed shortly, the lack of vertical resolution is even more of an issue in these models. Moreover, these studies rely on the hydrostatic approximation. Excluding non-hydrostatic effects inhibits convective instabilities that can potentially change the results. Based on these shortcomings as well as several other disadvantages that will be demonstrated below, it might not be conclusive to rely on GCM/NWP models in the preliminary studies of the effect of the convective forcing on the energetics of the atmosphere.

Considering all of the above, I propose a high-resolution non-hydrostatic simulation in which large scales are either initially nonlinearly balanced or only linearly balanced modes (geostrophic modes) are forced. Concurrent with this large-scale balance, buoyancy modes of a limited band are also forced at very small scales. This small-scale forcing is unbalanced and can mimic convective processes. The scale of the unbalanced forcing should be larger than the dissipation range. By comparing two sets of simulations with and without small-scale forcing, the potential effects of convective processes on the energy spectrum can be investigated. These include

the wavenumber at which the spectrum shallows as well as the flux of energy in the shallow range. To the best of my knowledge, there is no numerical turbulence study that incorporates both large-scale balance dynamics as well as small-scale unbalanced forcing. Fortunately, today's computational resources allow for simultaneous inclusion of these two phenomena with a reasonable separation of scales between them. Such a study will certainly shed more light on the possible energy and PV fluxes of the atmosphere.

The other important missing phenomenon in this study is the effect of the boundary layer near the ground. Topography-induced waves may affect the breakdown of balance. These waves, unlike spontaneously generated IGWs in this study, have preferred directions and are inhomogeneous. As future work, incorporation of the ground effects will complete the picture presented here. In addition to topography effects, the tropopause (which can be considered an internal boundary) may affect the energy spectrum in some regimes.

Following the suggestion of idealised simulations above, it is appropriate to defend the choice of these simulations over GCM/NWP models. Despite lacking realistic features such as proper boundaries, the idealised simulations have some advantages over GCM/NWP models, which may allow them to answer certain questions better. Some of these advantages are listed below.[1]

[1] An added motivation for this rather detailed comparison between idealised turbulence studies and GCM/NWP models is the comments of my colleagues in conferences as well as the review of my papers. The sole fact that nowadays GCM/NWP models are capable of producing correct energy spectra unjustifiably convinces some

First, GCM models due to their realistic global features cannot selectively focus on different ranges of scales. They only focus on global scales. Therefore, the range of mesoscales that these models resolve is small. For example, Hamilton *et al.* (2008) resolved less than a decade, once dissipation and subgrid parametrisations are excluded. Idealised simulations, on the other hand, can have parameters such that the kink of the energy spectrum moves to smaller wavenumbers and provides a larger span for the shallow mesoscale spectrum. Knowing that the kink of the atmospheric energy spectrum occurs at around 500 km in figure 1–4, while convective scales are as small as tens of kilometres (or even smaller), the unbalanced forcing needs a reasonable range between it and the large-scale balance dynamics. Considering this aspect of the problem, idealised simulations may provide a more enlightening platform.

Second, GCM/NWP models with physics parametrisation schemes need to be "tuned," which is often a very tedious step. Therefore, one faces serious difficulties and limits in changing the flow parameters such as Ro and Re (or equivalently the resolution). The question posed above requires changing parameters such that the relative strength of unbalanced forcing and balanced energy, as well as an objective exploration of robustness to Rossby, Froude and dissipation can be explored. These requirements are nearly unattainable by current GCM/NWP models, while they can be implemented in idealised turbulence simulations.

---

to prefer them unconditionally over idealised simulations. Therefore, in my opinion, re-emphasising the advantages and disadvantages of both approaches is helpful.

Third, GCM/NWP models face challenges in unravelling feedback mechanisms between surface fluxes, topography, vertical inhomogeneity in the form of troposhere, stratosphere, boundary layer, thermocline, etc. These along with convective adjustment schemes, radiation and other physics parametrisations make the study of selected phenomena and their effects on the dynamics very difficult. Idealized simulations can exclude these complications to limit their focus on one particular problem regarding the dynamics in a spatially homogenous and perhaps statistically stationary context.

Fourth, GCM/NWP models employ a grid that can best capture the close-to-QG dynamics at large scales. QG structures at these scales have a large horizontal to vertical length scale ratio. Therefore, the grids that are best suited to resolving global scales may be incorrect for small-scale phenomena that might be close to isotropic.

Fifth, due to their simplified geometry, idealised simulations can employ very efficient and clean numerical methods (such as the pseudo-spectral method), which allow them to attain higher effective resolutions than GCM/NWP models. To have reliable statistics of stratified turbulence, it has recently emerged that one of the necessary conditions is resolving the buoyancy scale U/N, which is around 1 km near the tropopause (Waite & Bartello, 2004; Lindborg, 2006; Waite, 2011). If this scale is not resolved, the vertical characteristic length will be dictated by vertical diffusion of momentum instead of stratification, which is very unrealistic and leads to spuriously steep horizontal energy spectra (Waite, 2011). In fact, it has been argued that it is necessary to resolve the small-scale regime of 3D isotropic turbulence, in addition to

the stratified range, by resolving the Ozmidov scale, which is on the order of tens of metres (Brethouwer *et al.*, 2007; Bartello & Tobias, 2013). How these resolution requirements are affected by the addition of weak rotation is still an open question and is one of central importance to meteorology and oceanography. In any event, these scales are not resolved by GCM/NWP models today. For instance, a high-resolution GCM study by Hamilton *et al.* (2008) used 24 vertical levels from the ground to about 1 hPa, which led to about 1.5 km spacing in the upper troposphere. Knowing that vertical momentum and thermal diffusion restrict the fully nonlinear scales to much larger than the grid scale, it is clear these studies fall short of resolving the vertical outer scale U/N. That being said, with their sub-grid dissipation schemes, these simulations might still offer realistic large-scale dynamics, but their mesoscale transition must be interpreted with caution.

In short, this study has attempted to connect two important realms of geophysical fluid dynamics: turbulence and balance dynamics. In doing so we calculated the turbulence statistics of balanced versus unbalanced flows. Among these is the scaling of the energy spectrum which was shown to be different for balanced and unbalanced components of the flow. This hints that the steep subrange of the observed atmospheric and oceanic energy spectra is predominantly balanced and the shallow subrange unbalanced. By taking the definition of balance to higher orders, we have demonstrated that the results of low- and high-order balance are qualitatively similar. However, we elaborated upon the quantitative differences, which provide insight for the interpretation of the results in related studies.

CHAPTER 6

Appendix

6.1 Appendix A: normal mode decomposition

The method of normal mode decomposition was put forward by Leith (1980) for the hydrostatic Boussinesq approximation. Bartello (1995) expanded it to the non-hydrostatic Boussinesq system, which is reviewed briefly in this appendix. This decomposition is based on the non-dissipative linearised equations in Fourier space, i.e. the RHS of equation (1.8):

$$\frac{\partial u_{\mathbf{k}}}{\partial t} - f\ v_{\mathbf{k}} + ik_x p_{\mathbf{k}} = 0, \tag{6.1a}$$

$$\frac{\partial v_{\mathbf{k}}}{\partial t} + f\ u_{\mathbf{k}} + ik_y p_{\mathbf{k}} = 0, \tag{6.1b}$$

$$\frac{\partial w_{\mathbf{k}}}{\partial t} - b_{\mathbf{k}} + ik_z p_{\mathbf{k}} = 0, \tag{6.1c}$$

$$ik_x u_{\mathbf{k}} + ik_y v_{\mathbf{k}} + ik_z w_{\mathbf{k}} = 0, \tag{6.1d}$$

$$\frac{\partial b_{\mathbf{k}}}{\partial t} + N^2 w_{\mathbf{k}} = 0. \tag{6.1e}$$

We define horizontal divergence as $\delta_{\mathbf{k}} = ik_x u_{\mathbf{k}} + ik_y v_{\mathbf{k}}$ and vertical vorticity as $\zeta_{\mathbf{k}} = ik_x v_{\mathbf{k}} - ik_y u_{\mathbf{k}}$. Assuming $k_z \neq 0$ and $k_h \neq 0$, the set of linearised equations in (6.1) can be written in terms of $\zeta_{\mathbf{k}}$, $\delta_{\mathbf{k}}$ and $b_{\mathbf{k}}$ as below[1]

$$\frac{\partial \zeta_{\mathbf{k}}}{\partial t} = -f\delta_{\mathbf{k}}, \tag{6.2a}$$

$$\frac{\partial \delta_{\mathbf{k}}}{\partial t} = +f\frac{k_z^2}{k^2}\zeta_{\mathbf{k}} - i\frac{k_h^2}{k^2}k_z b_{\mathbf{k}}, \tag{6.2b}$$

$$\frac{\partial b_{\mathbf{k}}}{\partial t} = -i\frac{N^2}{k_z}\delta_{\mathbf{k}}. \tag{6.2c}$$

The following change of variables

$$\delta_{\mathbf{k}} \to \frac{k_z}{k}D_{\mathbf{k}}, \quad b_{\mathbf{k}} \to \frac{N}{k_h}T_{\mathbf{k}} \tag{6.3}$$

yields

$$\frac{\partial}{\partial t}\begin{pmatrix} \zeta_{\mathbf{k}} \\ D_{\mathbf{k}} \\ T_{\mathbf{k}} \end{pmatrix} = i\begin{pmatrix} 0 & ifk_z/k & 0 \\ fk_z/k & 0 & -Nk_h/k \\ 0 & -Nk_h/k & 0 \end{pmatrix}\begin{pmatrix} \zeta_{\mathbf{k}} \\ D_{\mathbf{k}} \\ T_{\mathbf{k}} \end{pmatrix} \tag{6.4}$$

where the matrix in the RHS of (6.4) is Hermitian. Therefore, its eigenvalues are real, and the eigenvectors are orthogonal. For wavenumber $\mathbf{k}$, these are given by

$$\omega_{\mathbf{k}}^{(0)} = 0, \quad \omega_{\mathbf{k}}^{(\pm)} = \pm\left(\frac{f^2 {k_z}^2 + N^2 {k_h}^2}{{k_z}^2 + {k_h}^2}\right)^{1/2}, \tag{6.5}$$

[1] The cases of $k_z \neq 0$ and $k_h \neq 0$ should be decomposed differently. For details see Bartello (1995)

with corresponding eigenvectors

$$\chi_{\mathbf{k}}^{(0)} = \frac{1}{k\ \omega_{\mathbf{k}}^{(+)}} \begin{pmatrix} Nk_h \\ 0 \\ -ifk_z \end{pmatrix}, \quad \chi_{\mathbf{k}}^{(\pm)} = \frac{1}{\sqrt{2}k\ \omega_{\mathbf{k}}^{(\pm)}} \begin{pmatrix} ifk_z \\ k\omega_{\mathbf{k}}^{(\pm)} \\ -Nk_h \end{pmatrix} \tag{6.6}$$

The projection of flow variables, $\zeta_{\mathbf{k}}$, $\delta_{\mathbf{k}}$ and $b_{\mathbf{k}}$, onto the above eigenvectors gives the following sets of modes

$$G_{\mathbf{k}} = \langle \begin{pmatrix} \zeta_{\mathbf{k}} \\ D_{\mathbf{k}} \\ T_{\mathbf{k}} \end{pmatrix}, \chi_{\mathbf{k}}^{(0)} \rangle = \frac{Nk_h\zeta_{\mathbf{k}} + ifk_zT_{\mathbf{k}}}{k\ \omega_{\mathbf{k}}^{(+)}}, \tag{6.7}$$

$$A_{\mathbf{k}}^{(\pm)} = \langle \begin{pmatrix} \zeta_{\mathbf{k}} \\ D_{\mathbf{k}} \\ T_{\mathbf{k}} \end{pmatrix}, \chi_{\mathbf{k}}^{(\pm)} \rangle = \frac{k\ \omega_{\mathbf{k}}^{(\pm)} D_{\mathbf{k}} - ifk_z\zeta_{\mathbf{k}} - Nk_hT_{\mathbf{k}}}{\sqrt{2}k\ \omega_{\mathbf{k}}^{(\pm)}} \tag{6.8}$$

where $\langle X, Y \rangle = X_i Y_i^*$ denotes the scaler product, in which repeated indices, i, are summed and * denotes the complex conjugate. $G_{\mathbf{k}}$ are referred to as geostrophic, vortical or rotational modes, while $A_{\mathbf{k}}^{(\pm)}$ are referred to as ageostrophic, wave or gravitational modes.

6.2 Appendix B: numerical implementation of the Baer-Tribbia scheme

To describe our algorithm succinctly, we use the notation/terminology of linear and nonlinear operators in phase space interchangeably with the vector form modal variables. The evolution of the state variable X can be expressed by

$$\frac{\partial X}{\partial t} = i\mathcal{L}X + \mathcal{N}(X) \tag{6.9}$$

This equation is the equivalent of (1.5) where all the linear terms are collected in the matrix, $\mathcal{L}$, and the quadratic terms are presented with $\mathcal{N}(X)$, a nonlinear vector function of the state vector X. Using the pseudo-spectral technique, $N(X)$ can be efficiently computed by transforming the modal variables to real space, carrying the multiplication for physical variables and then transforming the multiplied terms back to spectral space. The operator $\mathcal{N}(X)$ can take any other vector with the same dimension as X and output the convolution sum fast and efficiently. In addition to this operator, we consider the following linear projections

$$\zeta = \mathcal{G}X, \quad \eta = \mathcal{A}X, \quad X = \mathcal{P}^{-1}(\zeta, \eta) \tag{6.10}$$

where $\mathcal{G}$ and $\mathcal{A}$ are the projections on geostrophic and ageostrophic manifolds. $\mathcal{P}^{-1}$ performs the inverse transform by taking the projected geostrophic and ageostrophic components and outputting the original state variable. Note that all operators in (6.10) are linear, hence rather easy to compute.

Using these we derive the nonlinear terms in (2.5) when ageostrophic modes are set to zero

$$\widehat{\eta} = \Phi(G,0) = \mathcal{A}\,\mathcal{N}(\mathcal{P}^{-1}(\mathcal{G}X,0)) \quad \widehat{\zeta} = \Psi(G,0) = \mathcal{G}\,\mathcal{N}(\mathcal{P}^{-1}(\mathcal{G}X,0)) \tag{6.11}$$

In the next step, all coefficients in (2.16) are derived as follows

$$\Psi^{0,0}_{\boldsymbol{k}} = \widehat{\eta} \tag{6.12}$$

$$\Psi^{1,0}_{\boldsymbol{k}} = \mathcal{A}\,\mathcal{N}(\mathcal{P}^{-1}(\mathcal{G}X,\widehat{\eta})) - \mathcal{A}\,\mathcal{N}(\mathcal{P}^{-1}(0,\widehat{\eta})) - \mathcal{A}\,\mathcal{N}(\mathcal{P}^{-1}(\mathcal{G}X,0)) \tag{6.13}$$

$$\begin{aligned}\Psi^{0,1}_{\boldsymbol{k}} = &\sum_{\boldsymbol{kp+q}} \psi_{GG}\left(G_{\boldsymbol{p}} + \sum_{\boldsymbol{p=m+n}} \phi_{GG} G_{\boldsymbol{m}} G_{\boldsymbol{n}}\right)\left(G_{\boldsymbol{q}} + \sum_{\boldsymbol{q=r+s}} \phi_{GG} G_{\boldsymbol{r}} G_{\boldsymbol{s}}\right) \\ &- \sum_{\boldsymbol{k=p+q}} \psi_{GG} G_{\boldsymbol{p}} G_{\boldsymbol{q}} - \sum_{\boldsymbol{k=p+q}} \psi_{GG}\left(\sum_{\boldsymbol{p=m+n}} \phi_{GG} G_{\boldsymbol{m}} G_{\boldsymbol{n}}\right)\left(\sum_{\boldsymbol{q=r+s}} \phi_{GG} G_{\boldsymbol{r}} G_{\boldsymbol{s}}\right)\end{aligned}$$

$$= \mathcal{A}\,\mathcal{N}(\mathcal{P}^{-1}(\widehat{\zeta} + \mathcal{G}X,0)) - \mathcal{A}\,\mathcal{N}(\mathcal{P}^{-1}(\mathcal{G}X,0)) - \mathcal{A}\,\mathcal{N}(\mathcal{P}^{-1}(\widehat{\zeta},0)) \tag{6.14}$$

6.3 Appendix C: triad interactions at the breakdown of balance

If we write the state variables as the sum of their balanced and unbalanced components $\mathbf{X_k} = \mathbf{B_k} + \mathbf{U_k}$ and substitute them into (4.2), we get

$$\frac{d}{dt}(\mathbf{U_k} + \mathbf{B_k}) = -i\mathbf{L_k}(\mathbf{U_k} + \mathbf{B_k}) + \mathbf{N_k}(\mathbf{B},\mathbf{U}). \tag{6.15}$$

Unlike the geostrophic and ageostrophic modes (zeroth order of balance), the decomposition at higher orders is not orthogonal. Therefore, for the second moment, we introduce cross-terms (multiplication of balance and unbalance) in the time derivative. If we denote the complex of conjugate of () with $()^*$, the dot-product of (6.15) with $\mathbf{B}_\mathbf{k}^*$ and $\mathbf{U}_\mathbf{k}^*$ added to its complex conjugate yields

$$\begin{aligned}
&\frac{d}{dt}(\mathbf{U_k}+\mathbf{B_k})\cdot\mathbf{B}_\mathbf{k}^*+\frac{d}{dt}(\mathbf{U}_\mathbf{k}^*+\mathbf{B}_\mathbf{k}^*)\cdot\mathbf{B_k}=\\
&\frac{d}{dt}(\mathbf{B_k}\cdot\mathbf{B}_\mathbf{k}^*)+\mathbf{B}_\mathbf{k}^*\cdot\frac{d}{dt}\mathbf{U_k}+\mathbf{B_k}\cdot\frac{d}{dt}\mathbf{U}_\mathbf{k}^*=\\
&-i\mathbf{B}_\mathbf{k}^{*T}\ \mathbf{L}(\mathbf{B_k}+\mathbf{U_k})-i(\mathbf{B}_\mathbf{k}^*+\mathbf{U}_\mathbf{k}^*)^T\ \mathbf{L}^T\ \mathbf{B_k}+\mathbf{N_k}\cdot\mathbf{B}_\mathbf{k}^*+\mathbf{N}_\mathbf{k}^*\cdot\mathbf{B_k}
\end{aligned}\tag{6.16}$$

$$\begin{aligned}
&\frac{d}{dt}(\mathbf{U_k}+\mathbf{B_k})\cdot\mathbf{U}_\mathbf{k}^*+\frac{d}{dt}(\mathbf{U}_\mathbf{k}^*+\mathbf{B}_\mathbf{k}^*)\cdot\mathbf{U_k}=\\
&\frac{d}{dt}(\mathbf{U_k}\cdot\mathbf{U}_\mathbf{k}^*)+\mathbf{U}_\mathbf{k}^*\cdot\frac{d}{dt}\mathbf{B_k}+\mathbf{U_k}\cdot\frac{d}{dt}\mathbf{B}_\mathbf{k}^*=\\
&-i\mathbf{U}_\mathbf{k}^{*T}\ \mathbf{L}(\mathbf{B_k}+\mathbf{U_k})-i(\mathbf{B}_\mathbf{k}^*+\mathbf{U}_\mathbf{k}^*)^T\ \mathbf{L}^T\ \mathbf{U_k}+\mathbf{N_k}\cdot\mathbf{U}_\mathbf{k}^*+\mathbf{N}_\mathbf{k}^*\cdot\mathbf{U_k}
\end{aligned}\tag{6.17}$$

in which $()^T$ is the vector/matrix transpose and we employed the fact that $\mathbf{L}$ is Hermitian. Quadratic nonlinearities in (6.16) and (6.17) can be expanded in terms of $\mathbf{B}$ and $\mathbf{U}$ as below

$$\mathbf{N_k}(\mathbf{U_k},\mathbf{B_k})=\xi_{BB}(\mathbf{B},\mathbf{B})+\xi_{BU}(\mathbf{B},\mathbf{U})+\xi_{UU}(\mathbf{U},\mathbf{U}).\tag{6.18}$$

We are considering the case of balanced IC, where $\mathbf{U}=0$ initially. If we restrict our attention to small time, then $\mathbf{B}$ is order unity and $\mathbf{U}$ is order $\tilde{\epsilon}$. Hence, we can add (6.16) and (6.17), and keep the terms up to the first order in $\tilde{\epsilon}$

$$
\begin{aligned}
&\frac{d}{dt}(\mathbf{B_k}\cdot\mathbf{B_k^*})+\frac{d}{dt}(\mathbf{B_k^*}\cdot\mathbf{U_k}+\mathbf{B_k}\cdot\mathbf{U_k^*})=\\
&-i\mathbf{B_k^{*T}}(\mathbf{L}+\mathbf{L}^T)\mathbf{B_k}-i\mathbf{U_k^{*T}}(\mathbf{L}+\mathbf{L}^T)\mathbf{B_k}-i\mathbf{B_k^{*T}}(\mathbf{L}+\mathbf{L}^T)\mathbf{U_k}+\\
&(\xi_{BB}(\mathbf{B},\mathbf{B})\cdot\mathbf{B_k^*}+(\xi_{BB}(\mathbf{B},\mathbf{B})^*\cdot\mathbf{B_k}+(\xi_{BU}(\mathbf{B},\mathbf{U})\cdot\mathbf{B_k^*}+(\xi_{BU}(\mathbf{B},\mathbf{U})^*\cdot\mathbf{B_k}\\
&(\xi_{BB}(\mathbf{B},\mathbf{B})\cdot\mathbf{U_k^*}+(\xi_{BB}(\mathbf{B},\mathbf{B})^*\cdot\mathbf{U_k}.
\end{aligned}
\tag{6.19}
$$

We sum the above equation over cylindrical shells to derive the budget of $E_B(k_h)$ defined in (4.31). Note that the second term in the LHS as well as the first line in the RHS of (6.19) do not have any imaginary parts (they are equal to their complex conjugate). Hence, they sum to zero over cylindrical shells.

$$
\begin{aligned}
&\frac{d}{dt}E_{\mathbf{B}}(k_h)=\frac{d}{dt}\sum_{|\mathbf{k}'-\mathbf{k}'\cdot z|=k_h}\langle\mathbf{B_k},\mathbf{B_k^*}\rangle=\\
&\sum_{|\mathbf{k}'-\mathbf{k}'\cdot z|=k_h}\big(\langle\xi_{BB}(\mathbf{B},\mathbf{B}),\mathbf{B_k^*}\rangle+\langle\xi_{BU}(\mathbf{B},\mathbf{U}),\mathbf{B_k^*}\rangle+\langle\xi_{BB}(\mathbf{B},\mathbf{B}),\mathbf{U_k^*}\rangle+c.c.\big).
\end{aligned}
\tag{6.20}
$$

References

Arakawa, A. & Schubert, W. H. 1974 Interaction of a cumulus cloud ensemble with the large-scale environment, part I. *Journal of the Atmospheric Sciences* **31** (3), 674–701.

Asselin, O., Bartello, P. & Straub, D. N. 2016 On quasigeostrophic dynamics near the tropopause. *Physics of Fluids* **28** (2), 026601.

Asselin, O., Bartello, P. & Straub, D. N. 2017 On boussinesq dynamics near the tropopause. *Journal of the Atmospheric Sciences* **submitted**.

Babin, A., Mahalov, A. & Nicolaenko, B. 1996 Global splitting, integrability and regularity of 3d euler and navier-stokes equations for uniformly rotating fluids. *European Journal of Mechanics - B/Fluids* **15** (3), 291–300.

Baer, F. 1977 Adjustment of initial conditions required to suppress gravity oscillations in nonlinear flows. *Beitraege zur Physik der Atmosphaere* **50** (3), 350–366.

Baer, F. & Tribbia, J. J. 1977 On complete filtering of gravity modes through nonlinear initialization. *Monthly Weather Review* **105** (12), 1536–1539.

Bartello, P. 1995 Geostrophic adjustment and inverse cascades in rotating stratified turbulence. *Journal of the Atmospheric Sciences* **52** (24), 4410–4428.

Bartello, P. 2010 Quasigeostrophic and stratified turbulence in the atmosphere. In *IUTAM Symposium on Turbulence in the Atmosphere and Oceans*, pp. 117–130. Springer.

Bartello, P., Métais, O. & Lesieur, M. 1996 Geostrophic versus wave eddy viscosities in atmospheric models. *Journal of the Atmospheric Sciences* **53** (4), 564–571.

Bartello, P. & Tobias, S. M. 2013 Sensitivity of stratified turbulence to the buoyancy reynolds number. *Journal of Fluid Mechanics* **725**, 1.

Batchelor, G. K. 1953 *The theory of homogeneous turbulence*. Cambridge university press.

Batchelor, G. K. 1969 Computation of the energy spectrum in homogeneous two-dimensional turbulence. *The Physics of Fluids* **12** (12), II–233.

Batchelor, G. K. 2000 *An introduction to fluid dynamics*. Cambridge university press.

Bender, C. M. & Orszag, S. A. 1999 *Asymptotic Methods and Perturbation Theory*. Springer.

Billant, P. & Chomaz, J. M. 2001 Self-similarity of strongly stratified inviscid flows. *Physics of Fluids* **13** (6), 1645–1651.

Blumen, W. 1978 Uniform potential vorticity flow: Part I. theory of wave interactions and two-dimensional turbulence. *Journal of the Atmospheric Sciences* **35** (5), 774–783.

Bokhove, O. & Shepherd, T. G. 1996 On hamiltonian balanced dynamics and the slowest invariant manifold. *Journal of the Atmospheric Sciences* **53** (2), 276–297.

Brethouwer, G., Billant, P., Lindborg, E. & Chomaz, J. M. 2007 Scaling analysis and simulation of strongly stratified turbulent flows. *Journal of Fluid*

Mechanics **585**, 343–368.

Browning, G. & Kreiss, H. O. 1982 Initialization of the shallow water equations with open boundaries by the bounded derivative method. *Tellus* **34** (4), 334–351.

Bühler, O., Callies, J. & Ferrari, R. 2014 Wave-vortex decomposition of one-dimensional ship-track data. *Journal of Fluid Mechanics* **756**, 1007.

Callies, J., Bühler, O. & Ferrari, R. 2016 The dynamics of mesoscale winds in the upper troposphere and lower stratosphere. *Journal of the Atmospheric Sciences* **73** (12), 4853–4872.

Callies, J. & Ferrari, R. 2013 Interpreting energy and tracer spectra of upper-ocean turbulence in the submesoscale range (1–200 km). *Journal of Physical Oceanography* **43** (11), 2456–2474.

Callies, J., Ferrari, R. & Bühler, O. 2014 Transition from geostrophic turbulence to inertia–gravity waves in the atmospheric energy spectrum. *Proceedings of the National Academy of Sciences* **111** (48), 17033–17038.

Camassa, R. 1995 On the geometry of an atmospheric slow manifold. *Physica D: Nonlinear Phenomena* **84** (3-4), 357–397.

Charney, J. G. 1949 On a physical basis for the numerical prediction of large-scale motions in the atmosphere. *Journal of Meteorology* **6**, 371–385.

Charney, J. G. 1971 Geostrophic turbulence. *Journal of the Atmospheric Sciences* **28** (6), 1087–1095.

Charney, J. G., Fjörtoft, R. & von Neumann, J. 1950 Numerical integration of the barotropic vorticity equation. *Tellus* **2** (4), 237–254.

Chekroun, M. D., Liu, H. & McWilliams, J. C. 2017 The emergence of fast

oscillations in a reduced primitive equation model and its implications for closure theories. *Computers & Fluids* **151**, 3–22.

Cullen, M. J. P. 2017 The impact of high vertical resolution in the met office unified model. *Quarterly Journal of the Royal Meteorological Society* **143** (702), 278–287.

Deusebio, E., Vallgren, A. & Lindborg, E. 2013 The route to dissipation in strongly stratified and rotating flows. *Journal of Fluid Mechanics* **720**, 66–103.

Dewan, E. M. 1979 Stratospheric wave spectra resembling turbulence. *Science* **204** (4395), 832–835.

Dritschel, D. G. & McKiver, W. J. 2015 Effect of Prandtl's ratio on balance in geophysical turbulence. *Journal of Fluid Mechanics* **777**, 569–590.

Dritschel, D. G. & Viudez, A. 2003 A balanced approach to modelling rotating stably stratified geophysical flows. *Journal of Fluid Mechanics* **488**, 123–150.

Emanuel, K. A & Živković-Rothman, M. 1999 Development and evaluation of a convection scheme for use in climate models. *Journal of the Atmospheric Sciences* **56** (11), 1766–1782.

Evans, K. J., Lauritzen, P. H., Mishra, S. K., Neale, R. B., Taylor, M. A. & Tribbia, J. J. 2013 AMIP simulation with the CAM4 spectral element dynamical core. *Journal of Climate* **26** (3), 689–709.

Falkovich, G. 1992 Inverse cascade and wave condensate in mesoscale atmospheric turbulence. *Physical Review Letters* **69** (22), 3173.

Feynman, R. P., Leighton, R. B. & Sands, M. 1963 *The Feynman lectures on physics. Vol. 1*. Addison-Wesley.

Ford, R., McIntyre, M. E. & Norton, W. A. 2000 Balance and the slow quasimanifold: some explicit results. *Journal of the Atmospheric Sciences* **57** (9), 1236–1254.

Gage, K. S. 1979 Evidence for a $k^{-5/3}$ law inertial range in mesoscale two-dimensional turbulence. *Journal of the Atmospheric Sciences* **36** (10), 1950–1954.

Gage, K. S. & Nastrom, G. D. 1986 Theoretical interpretation of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft during gasp. *Journal of the Atmospheric Sciences* **43** (7), 729–740.

Grant, H. L., Stewart, R. W. & Moilliet, A. 1962 Turbulence spectra from a tidal channel. *Journal of Fluid Mechanics* **12** (02), 241–268.

Hakim, G. J. 2000 Climatology of coherent structures on the extratropical tropopause. *Monthly weather review* **128** (2), 385–406.

Hamilton, Kevin, Takahashi, Yoshiyuki O. & Ohfuchi, Wataru 2008 Mesoscale spectrum of atmospheric motions investigated in a very fine resolution global general circulation model. *Journal of Geophysical Research: Atmospheres* **113** (D18).

Herring, J. R. & Métais, O. 1989 Numerical experiments in forced stably stratified turbulence. *Journal of Fluid Mechanics* **202**, 97–115.

Horn, L. H. & Bryson, R. A. 1963 An analysis of the geostrophic kinetic energy spectrum of large-scale atmospheric turbulence. *Journal of Geophysical Research* **68** (4), 1059–1064.

Hoskins, B. J., Draghici, I. & Davies, H. C. 1978 A new look at the ω-equation. *Quarterly Journal of the Royal Meteorological Society* **104** (439), 31–38.

Hoskins, B. J. & Hodges, K. I. 2002 New perspectives on the northern hemisphere winter storm tracks. *Journal of the Atmospheric Sciences* **59** (6), 1041–1061.

Julian, P. R., Washington, W. M., Hembree, L. & Ridley, C. 1970 On the spectral distribution of large-scale atmospheric kinetic energy. *Journal of the Atmospheric Sciences* **27** (3), 376–387.

Kafiabad, H. A. & Bartello, P. 2016 Balance dynamics in rotating stratified turbulence. *Journal of Fluid Mechanics* **795**, 914–949.

Kafiabad, H. A. & Bartello, P. 2017 Rotating stratified turbulence and the slow manifold. *Computers & Fluids* **151**, 23–34.

Kaneda, Y., Ishihara, T., Yokokawa, M., Itakura, K. & Uno, A. 2003 Energy dissipation rate and energy spectrum in high resolution direct numerical simulations of turbulence in a periodic box. *Physics of Fluids* **15** (2), L21–L24.

Kevorkian, J. & Cole, J. D. 2013 *Perturbation methods in applied mathematics*, , vol. 34. Springer Science & Business Media.

Klymak, J. M. & Moum, J. N. 2007*a* Oceanic isopycnal slope spectra. part I: Internal waves. *Journal of Physical Oceanography* **37** (5), 1215–1231.

Klymak, J. M. & Moum, J. N. 2007*b* Oceanic isopycnal slope spectra. part II: Turbulence. *Journal of Physical Oceanography* **37** (5), 1232–1245.

Kolmogorov, A. N. 1941 The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. In *Dokl. Akad. Nauk SSSR*, , vol. 30, pp. 301–305. JSTOR.

Kraichnan, R. H. 1967 Inertial ranges in two-dimensional turbulence. *The Physics*

of Fluids **10** (7), 1417–1423.

Kraichnan, R. H. 1971 Inertial-range transfer in two-and three-dimensional turbulence. *Journal of Fluid Mechanics* **47** (3), 525–535.

Kreiss, H. O. 1979 Problems with different time scales for ordinary differential equations. *SIAM Journal on Numerical Analysis* **16** (6), 980–998.

Kreiss, H. O. 1980 Problems with different time scales for partial differential equations. *Communications on Pure and Applied Mathematics* **33** (3), 399–439.

Kreiss, H. O. & Lorenz, J. 1994 On the existence of slow manifolds for problems with different timescales. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **346** (1679), 159–171.

Kundu, P. K. & Cohen, I. M. 2007 *Fluid Mechanics (4)*. Academic Press.

Lapeyre, G. 2017 Surface quasi-geostrophy. *Fluids* **2** (1), 7.

Laval, J. P., McWilliams, J. C. & Dubrulle, B. 2003 Forced stratified turbulence: successive transitions with reynolds number. *Physical Review E* **68** (3), 036308.

Leith, C. E. 1971 Atmospheric predictability and two-dimensional turbulence. *Journal of the Atmospheric Sciences* **28** (2), 145–161.

Leith, C. E. 1980 Nonlinear normal mode initialization and quasi-geostrophic theory. *Journal of the Atmospheric Sciences* **37** (5), 958–968.

Leith, C. E. & Kraichnan, R. H. 1972 Predictability of turbulent flows. *Journal of the Atmospheric Sciences* **29** (6), 1041–1058.

Lelong, M. P. & Riley, J. J. 1991 Internal wave-vortical mode interactions in strongly stratified flows. *Journal of Fluid Mechanics* **232** (1).

Lesieur, M. 2012 *Turbulence in fluids*, , vol. 40. Springer Science & Business Media.

Lilly, D. K. 1983 Stratified turbulence and the mesoscale variability of the atmosphere. *Journal of the Atmospheric Sciences* **40** (3), 749–761.

Lilly, D. K., Bassett, G., Droegemeier, K. & Bartello, P. 1998 Stratified turbulence in the atmospheric mesoscales. *Theoretical and Computational Fluid Dynamics* **11** (3-4), 139–153.

Lindborg, E. 1999 Can the atmospheric kinetic energy spectrum be explained by two-dimensional turbulence? *Journal of Fluid Mechanics* **388**, 259–288.

Lindborg, E. 2005 The effect of rotation on the mesoscale energy cascade in the free atmosphere. *Geophysical Research Letters* **32** (1).

Lindborg, E. 2006 The energy cascade in a strongly stratified fluid. *Journal of Fluid Mechanics* **550**, 207–242.

Lindborg, E. 2015 A Helmholtz decomposition of structure functions and spectra calculated from aircraft data. *Journal of Fluid Mechanics* **762**, R4.

Lindborg, E. & Cho, J. Y. 2000 Determining the cascade of passive scalar variance in the lower stratosphere. *Physical Review Letters* **85** (26), 5663.

Lorenz, E. N. 1980 Attractor sets and quasi-geostrophic equilibrium. *Journal of the Atmospheric Sciences* **37** (8), 1685–1699.

Lorenz, E. N. 1986 On the existence of a slow manifold. *Journal of the Atmospheric Sciences* **43** (15), 1547–1558.

Lorenz, E. N. 1992 The slow manifold-what is it? *Journal of the Atmospheric Sciences* **49** (24), 2449–2451.

Lorenz, E. N. & Krishnamurthy, V. 1987 On the nonexistence of a slow manifold. *Journal of the Atmospheric Sciences* **44** (20), 2940–2950.

Machenhauer, B. 1977 On the dynamics of gravity oscillations in a shallow water model with applications to normal mode initialization. *Beitraege zur Physik der Atmosphaere* **50**, 253–271.

Maffioli, A. & Davidson, P. A. 2015 Dynamics of stratified turbulence decaying from a high buoyancy reynolds number. *Journal of Fluid Mechanics* **786**, 210–233.

Marino, R., Mininni, P. D., Rosenberg, D. & Pouquet, A. 2013 Inverse cascades in rotating stratified turbulence: Fast growth of large scales. *Europhysics Letters* **102** (4), 44006.

Marino, R., Rosenberg, D., Herbert, C. & Pouquet, A. 2015 Interplay of waves and eddies in rotating stratified turbulence and the link with kinetic-potential energy partition. *Europhysics Letters* **112** (4), 49001.

McIntyre, M. E. & Norton, W. A. 2000 Potential vorticity inversion on a hemisphere. *Journal of the Atmospheric Sciences* **57** (9), 1214–1235.

McKiver, W. J. & Dritschel, D. G. 2008 Balance in non-hydrostatic rotating stratified turbulence. *Journal of Fluid Mechanics* **596**, 201–219.

McWilliams, J. C. 1989 Statistical properties of decaying geostrophic turbulence. *Journal of Fluid Mechanics* **198**, 199–230.

McWilliams, J. C., Weiss, J. B. & Yavneh, I. 1994 Anisotropy and coherent vortex structures in planetary turbulence. *Science* **264** (5157), 410–413.

McWilliams, J. C., Weiss, J. B. & Yavneh, I. 1999 The vortices of homogeneous geostrophic turbulence. *Journal of Fluid Mechanics* **401**, 1–26.

Melander, M. V., Zabusky, N. J. & McWilliams, J. C. 1988 Symmetric vortex merger in two dimensions: causes and conditions. *Journal of Fluid Mechanics* **195**, 303–340.

Mohebalhojeh, A. R. & Dritschel, D. G. 2001 Hierarchies of balance conditions for the f-plane shallow-water equations. *Journal of the Atmospheric Sciences* **58** (16), 2411–2426.

Molemaker, M. J., McWilliams, J. C. & Capet, X. 2010 Balanced and unbalanced routes to dissipation in an equilibrated eady flow. *Journal of Fluid Mechanics* **654**, 35–63.

Mydlarski, L. & Warhaft, Z. 1996 On the onset of high-reynolds-number grid-generated wind tunnel turbulence. *Journal of Fluid Mechanics* **320** (1), 331.

Nadiga, B. T. 2014 Nonlinear evolution of a baroclinic wave and imbalanced dissipation. *Journal of Fluid Mechanics* **756**, 965–1006.

Nastrom, G. D. & Gage, K. S. 1985 A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *Journal of the Atmospheric Sciences* **42** (9), 950–960.

Ngan, K., Bartello, P. & Straub, D. N. 2008 Dissipation of synoptic-scale flow by small-scale turbulence. *Journal of the Atmospheric Sciences* **65** (3), 766–791.

Obukhov, A. 1949 Kvoprosu o geostroficheskom vetra. *Izv. Akad. Nauk SSSR, Ser. Geograf. Geofiz.* **13**, 281–306.

Oppenheim, A. V. 1999 *Discrete-time signal processing*. Pearson Education India.

Ozmidov, R. V. 1963 Turbulence and mixing in the sea. *Progress in Sciences,*

Geophysics pp. 242–258.

Phillips, N. A. 1982 Jule charney's influence on meteorology. *Bulletin of the American Meteorological Society* **63** (5), 492–492.

Pinus, N. Z., Reiter, E. R., Shur, G. N. & Vinnichenko, N. K. 1967 Power spectra of turbulence in the free atmosphere. *Tellus* **19** (2), 206–213.

Polvani, L. M. 1991 Two-layer geostrophic vortex dynamics. part 2. alignment and two-layer v-states. *Journal of Fluid Mechanics* **225**, 241–270.

Polvani, L. M., Zabusky, N. J. & Flierl, G. R. 1989 Two-layer geostrophic vortex dynamics. part 1. upper-layer v-states and merger. *Journal of Fluid Mechanics* **205**, 215–242.

Pouquet, A. & Marino, R. 2013 Geophysical turbulence and the duality of the energy flow across scales. *Physical Review Letters* **111** (23), 234501.

Reznik, G. M., Zeitlin, V. & Ben Jelloul, M. 2001 Nonlinear theory of geostrophic adjustment. part 1. rotating shallow-water model. *Journal of Fluid Mechanics* **445**, 93–120.

Richardson, L. F. 1922 *Weather prediction by numerical process*. Cambridge University Press.

Riley, J. J. & deBruynKops, S. M. 2003 Dynamics of turbulence strongly influenced by buoyancy. *Physics of Fluids* **15**, 2047–2059.

Salmon, R. 1998 *Lectures on geophysical fluid dynamics*. Oxford University Press.

Saujani, S. & Shepherd, T. G. 2002 Comments on'balance and the slow quasi-manifold: some explicit results'. *Journal of the Atmospheric Sciences* **59** (19), 2874–2877.

Skamarock, W. C. 2004 Evaluating mesoscale nwp models using kinetic energy spectra. *Monthly Weather Review* **132** (12), 3019–3032.

Snyder, C., Muraki, D. J., Plougonven, R. & Zhang, F. 2007 Inertia-gravity waves generated within a dipole vortex. *Journal of the Atmospheric Sciences* **64** (12), 4417–4431.

Snyder, C., Skamarock, W. C. & Rotunno, R. 1993 Frontal dynamics near and following frontal collapse. *Journal of the Atmospheric Sciences* **50** (18), 3194–3212.

Sun, Y. Q., Rotunno, R. & Zhang, F. 2017 Contributions of moist convection and internal gravity waves to building the atmospheric- 5/3 kinetic energy spectra. *Journal of the Atmospheric Sciences* **74** (1), 185–201.

Sutcliff, R. 1947 A contribution to the problem of development. *Quarterly Journal of Royal Meteorological Society* **73**, 370–373.

Tennekes, H. & Lumley, J. L. 1972 *A first course in turbulence*. MIT press.

Tulloch, R. & Smith, K. S. 2006 A theory for the atmospheric energy spectrum: Depth-limited temperature anomalies at the tropopause. *Proceedings of the National Academy of Sciences* **103** (40), 14690–14694.

Tulloch, R. & Smith, K. S. 2009 Quasigeostrophic turbulence with explicit surface dynamics: Application to the atmospheric energy spectrum. *Journal of the Atmospheric Sciences* **66** (2), 450–467.

Vallgren, A., Deusebio, E. & Lindborg, E. 2011 Possible explanation of the atmospheric kinetic and potential energy spectra. *Physical Review Letters* **107** (26), 268501.

Vallis, G. K. 2006 *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press.

Van Kampen, N. G. 1985 Elimination of fast variables. *Physics Reports* **124** (2), 69–160.

Vanneste, J. 2013 Balance and spontaneous wave generation in geophysical flows. *Annual Review of Fluid Mechanics* .

Vanneste, J. & Yavneh, I. 2004 Exponentially small inertia-gravity waves and the breakdown of quasigeostrophic balance. *Journal of the Atmospheric Sciences* **61** (2), 211–223.

VanZandt, T. E. 1982 A universal spectrum of buoyancy waves in the atmosphere. *Geophysical Research Letters* **9** (5), 575–578.

Vautard, R. & Legras, B. 1986 Invariant manifolds, quasi-geostrophy and initialization. *Journal of the Atmospheric Sciences* **43** (6), 565–584.

Vinnichenko, N. K. & Dutton, J. A. 1969 Empirical studies of atmospheric structure and spectra in the free atmosphere. *Radio Science* **4** (12), 1115–1126.

Viúdez, A. 2007 The origin of the stationary frontal wave packet spontaneously generated in rotating stratified vortex dipoles. *Journal of Fluid Mechanics* **593**, 359–383.

Viúdez, A. 2008 The stationary frontal wave packet spontaneously generated in mesoscale dipoles. *Journal of Physical Oceanography* **38** (1), 243–256.

Viudez, A. & Dritschel, D. G. 2004 Optimal potential vorticity balance of geophysical flows. *Journal of Fluid Mechanics* **521**, 343–352.

Waite, M. L. 2011 Stratified turbulence at the buoyancy scale. *Physics of Fluids*

23 (6), 066602.

Waite, M. L. 2016 Dependence of model energy spectra on vertical resolution. *Monthly Weather Review* **144** (4), 1407–1421.

Waite, M. L. & Bartello, P. 2004 Stratified turbulence dominated by vortical motion. *Journal of Fluid Mechanics* **517**, 281–308.

Waite, M. L. & Bartello, P. 2006*a* Stratified turbulence generated by internal gravity waves. *Journal of Fluid Mechanics* **546**, 313–339.

Waite, M. L. & Bartello, P. 2006*b* The transition from geostrophic to stratified turbulence. *Journal of Fluid Mechanics* **568**, 89–108.

Waite, M. L. & Snyder, C. 2009 The mesoscale kinetic energy spectrum of a baroclinic life cycle. *Journal of the Atmospheric Sciences* **66** (4), 883–901.

Waite, M. L. & Snyder, C. 2013 Mesoscale energy spectra of moist baroclinic waves. *Journal of the Atmospheric Sciences* **70** (4), 1242–1256.

Warn, T. 1986 Statistical mechanical equilibria of the shallow water equations. *Tellus A* **38** (1), 1–11.

Warn, T. 1997 Nonlinear balance and quasi-geostrophic sets. *Atmosphere-Ocean* **35** (2), 135–145.

Warn, T., Bokhove, O., Shepherd, T. G. & Vallis, G. K. 1995 Rossby number expansions, slaving principles, and balance dynamics. *Quarterly Journal of the Royal Meteorological Society* **121** (523), 723–739.

Warn, T. & Menard, R. 1986 Nonlinear balance and gravity-inertial wave saturation in a simple atmospheric model. *Tellus A* **38** (4), 285–294.

Werne, J. & Fritts, D. C. 1999 Stratified shear turbulence: Evolution and

statistics. *Geophysical Research Letters* **26** (4), 439–442.

Whitehead, J. P. & Wingate, B. A. 2014 The influence of fast waves and fluctuations on the evolution of the dynamics on the slow manifold. *Journal of Fluid Mechanics* **757**, 155–178.

Wiin-Nielsen, A. 1967 On the annual variation and spectral distribution of atmospheric energy. *Tellus* **19** (4), 540–559.

Zeitlin, V., Reznik, G. M. & Ben Jelloul, M. 2003 Nonlinear theory of geostrophic adjustment. part 2. two-layer and continuously stratified primitive equations. *Journal of Fluid Mechanics* **491**, 207–228.