Internal Gravity Wave Emission in Different Dynamical Regimes

MANITA CHOUKSEY

IMPRS-ESM, MPI für Meteorologie, and Institut für Meereskunde, Universität Hamburg, Hamburg, Germany

CARSTEN EDEN AND NILS BRÜGGEMANN

Institut für Meereskunde, Universität Hamburg, Hamburg, Germany

(Manuscript received 10 August 2017, in final form 9 March 2018)

ABSTRACT

We aim to diagnose internal gravity waves emitted from balanced flow and investigate their role in the downscale transfer of energy. We use an idealized numerical model to simulate a range of baroclinically unstable flows to mimic dynamical regimes ranging from ageostrophic to quasigeostrophic flows. Wavelike signals present in the simulated flows, seen for instance in the vertical velocity, can be related to gravity wave activity identified by frequency and frequency–wavenumber spectra. To explicitly assign the energy contributions to the balanced and unbalanced (gravity) modes, we perform linear and nonlinear modal decomposition to decompose the full state variable into its balanced and unbalanced counterparts. The linear decomposition shows a reasonable separation of the slow and fast modes but is no longer valid when applied to a nonlinear system. To account for the nonlinearity in our system, we apply the normal mode initialization technique proposed by Machenhauer in 1977. Further, we assess the strength of the gravity wave activity and dissipation related to the decomposed modes for different dynamical regimes. We find that gravity wave emission becomes increasingly stronger going from quasigeostrophic to ageostrophic regime. The kinetic energy tied to the unbalanced mode scales close to Ro^2 (or Ri^{-1}), with Ro and Ri being the Rossby and Richardson numbers, respectively. Furthermore, internal gravity waves dissipate predominantly through small-scale dissipation, which emphasizes their role in the downscale energy transfer.

1. Introduction

Internal gravity waves (IGWs) occur naturally in the atmospheric and oceanic flows and influence the atmosphere mainly by vertical momentum transport and the ocean by density mixing. Despite their ubiquity and importance in the geophysical flows and numerous observational and numerical studies, the emission and dissipation of IGWs are not well understood. Consequently, the parameterization of IGWs remains a challenging task. This difficulty is in part related to the short spatial and temporal scales of IGWs, which render them hard to observe and difficult to resolve, in part to the specification of their sources, and to a certain extent to the difficulty of separating IGW from other motions. The latter is the theme of this paper, in which we diagnose IGWs emitted from balanced flows for different dynamical regimes and investigate their role in the downscale transfer of energy.

IGWs are forced mainly by orography, convection, and jet/front systems in the atmosphere and winds and tides in the ocean. Of these, the emission of IGWs from geostrophically balanced flows, referred to as "spontaneous generation" (see Vanneste 2013, and references therein), is of marked interest as it provides an avenue to the internal mechanisms in the flow that lead to IGW generation. IGWs generated spontaneously have been discussed extensively in both oceanic and atmospheric literature: in observations (e.g., Plougonven and Zhang 2014, their section 2), laboratory experiments (e.g., Williams et al. 2008), and several numerical simulations [discussed below; also see reviews by Vanneste (2013) and Plougonven and Zhang (2014), and references therein]. Spontaneous generation (or emission) is also of special interest for it allows for understanding the fundamental nature of balanced flows, which dominate much of the atmosphere and ocean.

The slow geostrophically balanced motions evolve over long time scales, whereas the IGW correspond to the fast unbalanced motions. The slowly varying nature

Corresponding author: Manita Chouksey, manita.chouksey@uni-hamburg.de

of the balanced motion led to the concept of "slow manifold" (Leith 1980; Lorenz 1980, 1986), which is defined as a subspace of the phase space that is strictly "invariant" and completely devoid of any IGW activity. However, the routinely observed emission of unbalanced fast motion (IGW) in geophysical flows questions the validity of such a manifold. The existence of the slow manifold —or rather its nonexistence—has been discussed by several authors (e.g., Lorenz and Krishnamurthy 1987; Lorenz 1992; Ford et al. 2000; Vanneste and Yavneh 2004), which makes clear that an exactly invariant slow manifold does not exist but rather manifolds with different degrees of "invariance," where the slow balanced and fast unbalanced motions coexist but their degree of interaction differs. Accordingly, such a manifold more suitably came to be called a fuzzy manifold (Warn 1997) or a slow quasi manifold (Ford et al. 2000). The nonexistence of an exactly invariant slow manifold implies that immaterial of the initially balanced conditions the slowly evolving balanced motion will always co-occur with the fast motion, and hence "spontaneous generation of IGWs for geophysical flows is inevitable" (Vanneste 2013). An insight into this mechanism also facilitates our understanding of the IGW emission from balanced mesoscale flows (mesoscale eddies) in the ocean, which has strong implications for the energy budget of the ocean.

Mesoscale eddies, a consequence of baroclinic instability, are ubiquitous in the ocean and are one of the most energetic components of the ocean energy budget. The eddies act as a reservoir of energy that enters the ocean at large scales, but energy in the ocean is finally dissipated at the viscous molecular scales via a downscale energy transfer. Balanced mesoscale flows are known to lose their energy to large oceanic scales through an upscale energy transfer (e.g., Charney 1971). How the balanced mesoscale eddies lose their energy to dissipative molecular scales is, however, unclear; this is where the unbalanced motions such as IGWs come into the picture. Studies suggest that balanced flows could lose their energy to unbalanced motions, like IGWs, through processes such as geostrophic adjustment, spontaneous emission (or loss of balance),[1] stimulated loss of balance, topographic interaction, direct extraction, and gravity wave drag. Geostrophic adjustment (e.g., Rossby 1938; Blumen 1972; Bartello 1995) differs from previously discussed spontaneous emission in that for the former process the flow is forced away from its

balanced state by an arbitrary initial condition and the flow then adjusts to its balanced state (geostrophy) while emitting unbalanced IGWs. Stimulated loss of balance (e.g., Gertz and Straub 2009; Xie and Vanneste 2015) (or stimulated emission) is different from loss of balance; it refers to the process by which externally forced waves can further "stimulate" the emission of waves, and this wave-mean interaction extracts energy from the balanced flow, whereas (spontaneous) loss of balance (e.g., Molemaker et al. 2005) is the process where the flow itself can transfer its energy into unbalanced motions when its balance breaks down. Topographic interaction (e.g., Dewar and Hogg 2010) requires that the balanced flow interact with topography and in the process emit unbalanced motions, whereas the energy transfer in the presence of winds from mesoscales directly to IGWs has been referred to as direct extraction and was discussed in the context of a wind-driven channel flow by Barkan et al. (2017).[2] Yet another mechanism is the transfer of energy from balanced flow to gravity waves by wave–mean-flow interaction referred to as gravity wave drag (cf. Bühler 2014; Eden and Olbers 2017), which in general can also by directed from the waves to the mean flow (i.e., in both directions). In these ways the balanced motions could find an energy pathway via unbalanced motions en route to viscous dissipation, resulting in the downscale transfer of energy. The idea is further motivated by numerous atmospheric and oceanic studies, and there seems to be a general consensus on this notion; some of these studies are briefly mentioned in the following passage. These studies also indicate a transition from theoretical studies where the mechanisms can be clearly identified to numerical simulations with more complex and more realistic flows, where identifying unambiguously a mechanism becomes increasingly challenging.

The loss of balance in a baroclinically unstable flow results in a transfer of energy from balanced mean flow to unbalanced motions and eventually to dissipation by means of a downscale energy transfer as shown by Molemaker et al. (2005). A direct interior route to dissipation by means of unbalanced motions in a Boussinesq flow that can initiate downscale energy transfer has been discussed in an idealized flow configuration by

---

[1] Note that spontaneous emission and loss of balance are different terms to represent essentially the same process [cf. Vanneste (2013) for a discussion].

[2] Note that direct extraction is similar to spontaneous emission except that in the former IGWs are emitted from the balanced flow by the action of winds, whereas the emission of IGWs by the latter mechanism does not require any external stimulus. Also note that direct extraction is different from stimulated loss of balance (or stimulated emission) because an IGW field need not be present for direct extraction. For a detailed discussion, see Barkan et al. (2017).

Molemaker et al. (2010). For a range of dynamical regimes, downscale energy transfer from baroclinically unstable flows in an idealized setting has been discussed by Brüggemann and Eden (2015), and they emphasize that an ageostrophic direct route to dissipation might be of importance in the energy budget of the ocean. For a realistic flow configuration, Capet et al. (2008b) find an increase in the downscale energy flux related to the ageostrophic dissipation route with an increased horizontal resolution that can better resolve ageostrophic dynamics. Another unbalanced route is the stimulated loss of balance that can propel the downscale energy transfer as shown by Gertz and Straub (2009) for an unstratified thin-aspect-ratio fluid. Spontaneous generation of IGWs from idealized dipoles (e.g., Sugimoto and Plougonven 2016) also suggest that the balanced flows could dissipate via unbalanced IGWs during a downscale energy transfer. An alternative perspective that links IGWs and balanced mesoscale eddies are the recent model-based estimates of small-scale dissipation rates in the ocean, which fail to reproduce the observations without eddy forcing taken into account in a parameterization of IGW, implying that eddy dissipation is necessary for the IGW energy budget (Pollmann et al. 2017).

Despite numerous observational and numerical studies on IGW emission from balanced motions, the exact mechanism behind this process remains poorly understood, and a puzzling part is the identification of IGWs. The complication related to the identification of the IGW signals is in part associated with the coupling of the balanced motions and IGWs. The extent of this coupling can be estimated by Rossby number Ro (or an equivalent geostrophic Richardson number Ri), which is a measure of the time scale separation between the slow balanced and fast unbalanced IGW motions. This coupling is weak for a regime with Ro $\ll$ 1, equivalent of a Ri $\gg$ 1, such that the fast and slow motions are "well separated"; on the contrary, for a finite Ro, equivalent to Ri = $O(1)$ or smaller, the fast and slow motions interact more strongly and the separation of these processes is not well defined (Vanneste 2013; Zeitlin 2008). The coupling between these motions adds to the intricacy of separation and detection of IGW signals from the balanced flow field. The emphasis of this paper is on the diagnosis of gravity wave signals.

The traditional approach to identify IGWs, or more generally unbalanced motions, is to use the fast fields such as the horizontal velocity divergence, the vertical component of the vorticity to determine the IGWs (e.g., Plougonven et al. 2005; Plougonven and Snyder 2007), or a spatial filtering to obtain small-scale vertical velocity as the signature of the IGWs (e.g., Sugimoto and Plougonven 2016). As another way, the full field of interest could be separated into horizontally nondivergent and vertically irrotational components (e.g., Molemaker et al. 2005) or simply put into rotational and divergent parts (Molemaker et al. 2010; Brüggemann and Eden 2015), which give the balanced and unbalanced contributions, respectively. Another technique to obtain IGW signals is from the quasigeostrophic omega equation, which gives the quasigeostrophic vertical velocity whose difference with the full vertical velocity yields the unbalanced IGW contribution (e.g., Danioux et al. 2012; Nagai et al. 2015). Although these methods work well in identifying IGWs owing to the waves' distinct spatial characteristics, the methods have the restraint that the unbalanced part interpreted as the IGW could still contain a notable amount of the balanced part, which hampers a concise interpretation of the signals.

A conceptually different approach is the linear modal decomposition, similar to projection onto the balanced manifold, which separates the slow balanced and fast IGW modes by decomposing the full field. Such a decomposition has been implemented for a linear system, for example, by Molemaker et al. (2010) and Borchert et al. (2014). The method applied by Borchert et al. (2014) has been applied recently to spontaneous emission of IGWs in an atmospheric configuration by Hien et al. (2018). However, the linear decomposition is no longer valid when applied to a nonlinear system, and thus an extension of this method to a nonlinear framework is desirable. This issue is addressed by a nonlinear normal mode initialization technique (NLNMI) developed independently by Machenhauer (1977) and Baer and Tribbia (1977), which allows for adjustments to the initial conditions in order to minimize the tendency of the system to generate fast motions. Relating these initialization procedures to quasigeostrophic balance, Leith (1980) derived decomposed modes for the hydrostatic Boussinesq equations, which was later generalized to the nonhydrostatic set of equations by Bartello (1995). More recently, the nonlinear initialization scheme of Baer and Tribbia (1977) was applied by Kafiabad and Bartello (2016) for balanced rotating dynamics to identify the energy cascades for differently initialized balanced regimes and by Kafiabad and Bartello (2017) to identify the mechanism and scales of spontaneous imbalance in a rotating stratified turbulence system. A decomposition based on normal mode inversion (similar to NLNMI) is discussed by McIntyre and Norton (2000) (see section 5). Another set of decomposition tools is based on the concept of balanced models to diagnose IGWs using the potential

vorticity equation. Such examples include an iterative decomposition tool for stratified, rotating flows known as the optimal potential vorticity (OPV) approach of Dritschel and Viúdez (2003), which has been applied to diagnose IGW packets spontaneously emitted from a balanced dipole by Viúdez (2008), and the renormalization group method to study spontaneous emission theoretically by Yasuda et al. (2015a,b). In this paper, we apply the nonlinear initialization procedure of Machenhauer (1977) with an aim to diagnose IGWs by decomposing the full field into its balanced and unbalanced counterparts, for a range of dynamical regimes from ageostrophic to quasigeostrophic.

The present work is also motivated by the previous work of Brüggemann and Eden (2015), who discussed spectral energy fluxes for a range of dynamical regimes ranging from ageostrophic (small Ri) to quasigeostrophic (larger Ri). From global maps of Ri obtained from an eddy-permitting realistic global ocean model, Brüggemann and Eden (2015) show (in their Figs. 7a and 7b) that low values of Ri occur at the surface, in the Southern Ocean, in western boundary currents, and at the equator. Most other regions have large Ri values, indicating quasigeostrophic conditions. Also, the submesoscale flows, which are dominant in the upper ocean, are characterized by Ri = $O(1)$ and Ro = $O(1)$ [see, e.g., reviews by Thomas et al. (2008) and Capet et al. (2008a)]. The regimes with smaller Ri show strong energy flux toward large wavenumbers, that is, a larger downscale transfer of energy (although there is also still an inverse energy transfer for the smaller wavenumber range). We use a very similar model setup as Brüggemann and Eden (2015), and by diagnosing gravity waves we also aim to answer the question whether this ageostrophic route toward dissipation is generated by gravity wave emission during ageostrophic baroclinic instability.

To weave together the numerous threads sketched up above, we use a simple model of baroclinic instability to simulate flows representing low to high Ri regimes from ageostrophic to quasigeostrophic (section 2). This allows us to study the evolution of the flows individually and evaluate the characteristics of the unbalanced motions without interference from other processes as is the case in a more complex flow configurations. To characterize the waves we first explore them in the Fourier space (section 3). We then apply Machenhauer's nonlinear initialization technique to the model data to obtain balanced and unbalanced modes and analyze gravity wave activity for different flow regimes (section 4). Further, we estimate the energy dissipation related to balanced and unbalanced modes for different dynamical regimes under study (section 5). Thereafter, a discussion of the results is presented

(section 6). Finally, we summarize the results and the key conclusions (section 7).

2. Baroclinic instability in different dynamical regimes

a. Numerical model

Baroclinic instability in our setup is represented in an idealized channel configuration resembling the classical Eady model (Eady 1949) of baroclinic instability: flow on an f plane with a prescribed stratification and a vertically sheared background zonal flow, under Boussinesq and hydrostatic approximations. Our model configuration differs from the Eady model by using a free surface and a meridional buoyancy gradient that is expressed as a sine function with an amplitude of M_0^2. The latter allows us to apply double-periodic boundaries to exclude lateral boundary instabilities. The presence of boundaries itself can lead to IGW generation [e.g., Borchert et al. (2014) and laboratory experiments by Williams et al. (2008)], and we suppress it with double-periodic boundaries since we focus on studying IGW emission from balanced flows. Note that the surface condition for the momentum balance is that of no stress (i.e., $du/dz = 0$). In its initial state, the model has a background flow in thermal wind balance with a constant stratification N_0^2 and a meridional buoyancy gradient that makes the flow baroclinically unstable. An example of the initial state temperature of the setup is shown in Fig. 1a. Temperature is the only active tracer in our setup, and hence temperature and buoyancy are equivalent. The model is forced with a restoring of the zonal mean flow and zonal mean buoyancy toward the initial state; there is no additional surface forcing and no bottom friction. The numerical code for the model is identical to the one in Eden (2016).

We simulate baroclinic instability for a range of dynamical regimes characterized by the geostrophic Richardson number Ri, which is defined here as the ratio of the vertical density stratification and vertical shear of the horizontal velocity. The quantities Ri, N_0, and M_0 are related as follows:

$$\mathrm{Ri} = \frac{N_0^2 f^2}{M_0^4}, \tag{1}$$

where f is the Coriolis parameter. The Ri sets the initial buoyancy restoring in the model, and hence the flow dynamics ranging from weakly stratified ageostrophic regime [Ri = $O(1)$] to strongly stratified quasigeostrophic regime (Ri ≫ 1), as described in Brüggemann and Eden (2015).

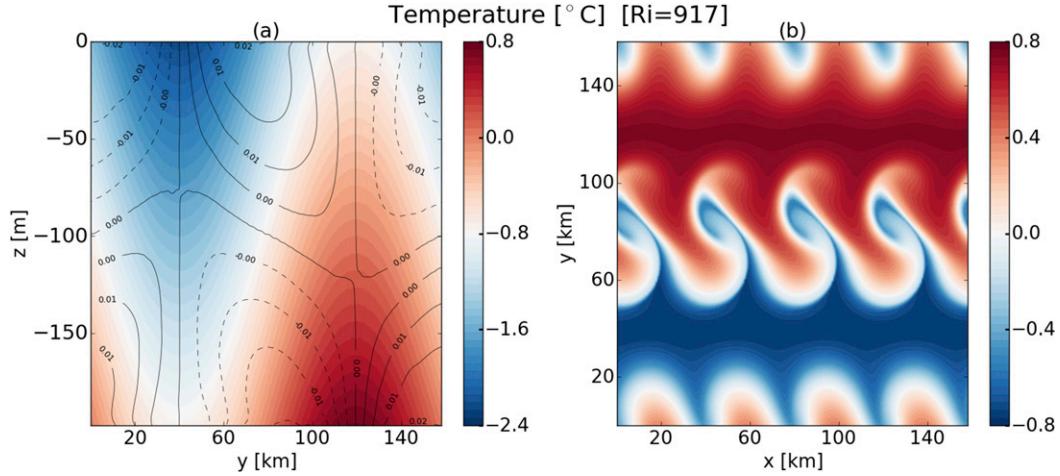The Ri can be related to the Rossby number Ro by assuming a thermal wind-sheared velocity and choosing

FIG. 1. Temperature field shown at different instances of the model run for Ri = 917 from the model setup. (a) Temperature at the initial state (time = 0) of the model with zonal velocity contours on top (8 days later) shown in black. (b) Temperature at the surface after about 100 days; notice the four wavelengths of the fastest-growing mode. Note that the vertical axis in (a) is depth and in (b) the meridional extent.

the Rossby radius L_r as the length scale. The Ri is then related to Ro in the following way:

$$\mathrm{Ri} = \frac{N_0^2 H^2}{U_0^2} = \frac{L_r^2 f^2}{U_0^2} = \mathrm{Ro}^{-2}, \qquad (2)$$

where $L_r = N_0 H / f$ is the Rossby radius of deformation, H is the basin depth, and U_0 is the mean velocity. In this way, Ri controls the degree of balance of the flow. The potential energy (PE) deposited into the model by the restoring is lost to the kinetic energy (KE) predominantly by baroclinic instability, as in the Eady model. At small scales, this energy is dissipated in the model by lateral biharmonic A_h and vertical harmonic friction A_v. Dissipation by biharmonic and harmonic friction is controlled by a grid Ekman number Ek, which is set to Ek = 0.1 for Ri = 3 and for Ri = 13, set to Ek = 0.05 for Ri = 327, and set to Ek = 0.03 for Ri = 917. At large scales, the energy is dissipated by linear relaxation (or drag) of the zonally averaged zonal flow to zero; that is, we add a term $\lambda_u(\overline{u} - u_*)$ to the momentum equation (as in Brüggemann and Eden 2015), which acts like a linear drag and where $\overline{u}$ denotes the zonally averaged zonal velocity, and $u_* = 0$ in our case. The time scale of the linear velocity drag λ_u (i.e., time scales on which this linear drag acts) and the relaxation time scale λ_T on which the buoyancy restoring acts are set proportional to the time scale of the fastest-growing mode $\sigma_{\max}$ (see Table 1). For the model simulations with Ri = 3, 13, and 327 we use $\lambda_u = 0.75\sigma_{\max}$ and $\lambda_T = 2\sigma_{\max}$. For Ri = 917, we use $\lambda_u = 0.5\sigma_{\max}$ and $\lambda_T = 1\sigma_{\max}$ in order to extract enough energy at the large scales to prevent the accumulation of energy and crashing of the

model. The range of values for λ_u and λ_T are listed in Table 1. Note that in our simulations we have used $\lambda_u \neq \lambda_T$, but we have also tested the case where $\lambda_u = \lambda_T$ for different values, and this is discussed in section 6.

A small random initial perturbation provided in temperature grows exponentially with time as the baroclinic instability sets in. For each stratification and shear, there exists a particular perturbation of a certain spatial scale that grows faster than perturbation of other scales. This particular spatial scale is on the order of the deformation radius (Eady 1949; Stone 1966) and is referred to as the fastest-growing mode, which becomes dominant over all the other perturbations and is therefore assumed to be the mode at which most of the PE is converted to KE. The linear growth rate $\sigma_{\max}$ and the corresponding wavenumber $k_{\max}$ of the fastest-growing mode can be expressed for the primitive equations as derived by Stone (1966):

$$k_{\max} \approx \sqrt{\frac{\mathrm{Ri}}{1 + \mathrm{Ri}}} \sqrt{\frac{5}{2}} L_r^{-1}, \quad \sigma_{\max} \approx \sqrt{\frac{\mathrm{Ri}}{1 + \mathrm{Ri}}} \sqrt{\frac{5}{54}} \frac{f}{\sqrt{\mathrm{Ri}}}. \qquad (3)$$

For quasigeostrophic approximation (large Ri), the length and time scales of this fastest-growing mode turn into expressions as derived by Eady (1949):

$$k_{\max} \approx 1.61 \, L_r^{-1}, \quad \sigma_{\max} \approx 0.31 \frac{f}{\sqrt{\mathrm{Ri}}}. \qquad (4)$$

Since our model is based on primitive equations, in our simulations we use Stone's formulation, and the model domain allows for four wavelengths of the fastest-growing

TABLE 1. An overview of the model parameters for our model setup. A range of values in the third column indicate the values of the respective parameter for Ri = 3 and Ri = 917 (left to right). A single value indicates the same value for all Ri.

Parameter	Definition	Value range
Number of grid points in (x, y, z) directions	nx, ny, nz	240, 240, 80
Channel depth	H	200 m
Aspect ratio	α	0.02
Coriolis frequency	f	$7 \times 10^{-5}\,\mathrm{s}^{-1}$
Brunt–Väisälä frequency	$N_0 = f/\alpha$	$3.5 \times 10^{-3}\,\mathrm{s}^{-1}$
Meridional stratification	$M_0 = \sqrt{fU_0/H}$	$4.9 \times 10^{-4} - 8.7 \times 10^{-5}\,\mathrm{s}^{-1}$
Mean flow	$U_0 = N_0 H/\sqrt{\mathrm{Ri}}$	$0.7 - 0.02\,\mathrm{m\,s}^{-1}$
Rossby radius of deformation	$L_r = N_0 H/f$	$10^4\,\mathrm{m}$
Wavelength of the fastest-growing mode	$k_{max} = \sqrt{5/2}\sqrt{\mathrm{Ri}/(1+\mathrm{Ri})}\,L_r^{-1}$	$(1.1 - 1.6) \times 10^{-4}\,\mathrm{m}^{-1}$
Growth rate of the fastest-growing mode	$\sigma_{max} = \sqrt{5/54}\sqrt{\mathrm{Ri}/(1+\mathrm{Ri})}\,f/\sqrt{\mathrm{Ri}}$	$1.5 \times 10^{-5} - 6.7 \times 10^{-7}\,\mathrm{s}^{-1}$
Length of the domain	$L_x = 4 \times 2\pi/k_{max}$	$22 - 15 \times 10^4\,\mathrm{m}$
Horizontal resolution	$dx = dy = L_x/nx$	936.6–662.6 m
Vertical resolution	$dz = H/nz$	2.5 m
Time step	$dt = dx(\mathrm{CFL}/U_0)$	1.33–89.80 s
Biharmonic horizontal friction	$A_h = \mathrm{Ek}fdx^4$	$(7.6 - 5.7) \times 10^5\,\mathrm{m}^4\,\mathrm{s}^{-1}$
Harmonic vertical friction	$A_v = \mathrm{Ek}fdz^2$	$1.1 \times 10^{-5} - 3.3 \times 10^{-7}\,\mathrm{m}^2\,\mathrm{s}^{-1}$
Linear drag coefficient	λ_u	$6.2 \times 10^{-5} - 1.8 \times 10^{-7}\,\mathrm{s}^{-1}$
Restoring time scale	λ_T	$3.0 \times 10^{-5} - 6.7 \times 10^{-7}\,\mathrm{s}^{-1}$

mode; that is, $L_x = 4 \times 2\pi/k_{max}$ (e.g., see Fig. 1b). The domain length is chosen to be equal in zonal and meridional directions, and so $L_x = L_y$. In the model setup the number of grid points in both horizontal directions is $nx = ny = 240$. Note, however, that the actual horizontal resolution, which determines the smallest resolved scales, depends on k_{max} and varies as we vary Ri for different simulations. The horizontal resolution is chosen as a fixed fraction of the deformation radius: $L_x/nx = 4 \times 2\pi/(120 \times 1.6 \times k_{max}) = 0.13L_r\sqrt{(1+\mathrm{Ri})/\mathrm{Ri}} \approx 0.13L_r$. Thus, the horizontal resolution and consequently the domain length depends on L_r, which changes with N (which in turn depends on Ri), while H and f are taken as constant. On the contrary, a fixed vertical depth of $H = 200$ m with $nz = 80$ layers provides a constant vertical resolution of 2.5 m for all simulations. The difference in domain sizes for different Ri can be seen in Fig. 2 for Ri = 3 and 917. Note, however, that even though the domain size for Ri = 3 is bigger than for Ri = 917, the dynamics at play in these regimes relate to the ageostrophic and quasigeostrophic oceanic motions, respectively. The time step of the model depends on the Courant–Friedrichs–Lewy (CFL) condition, mean flow, and the horizontal resolution. The CFL number is set to 0.001, 0.001, 0.001 and 0.003 for Ri = 3, 13, 327, and 917, respectively. An overview of the model parameters for the model setup is presented in Table 1. We next discuss the numerical simulations used in this work.

b. Numerical simulations

The numerical simulations can be used to investigate different dynamical regimes depending upon the choice of the parameter Ri: ageostrophic [Ri = $O(1)$] to quasigeostrophic (Ri $\gg$ 1). After about 45 days, all model simulations are in a statistically stationary equilibrium between the buoyancy forcing and the large- and small-scale dissipation. We disregard the spinup period here and consider the statistically stationary integrations only. Snapshots of KE and buoyancy for our setup's two extreme Ri (3 and 917) from the statistically stationary state are shown in Fig. 2. The differences between the two extreme regimes are evident from the spatial scales of the associated features in both buoyancy and KE, where the ageostrophic regime with Ri = 3 exhibits small-scale features with filament-like structures and has much higher KE than the quasigeostrophic regime with Ri = 917, which exhibits mesoscale eddy-like features with large spatial scales. A small Ri represents a weakly stratified but energetic flow referred to as an ageostrophic regime. Such a regime is characterized by a large Ro $(=1/\sqrt{\mathrm{Ri}})$ occurring in the ocean for instance in the mixed layer, near boundaries, or at the equator. Such ageostrophic motions are characterized by large velocity scales and spatial scales that are smaller than the Rossby radius of deformation with small-scale features such as filaments (e.g., Thomas et al. 2008; Capet et al. 2008a). On the other hand, a regime with a large Ri represents a strongly stratified regime in the ocean such as in the interior of the ocean. This regime exhibits a state of quasigeostrophic balance and is characteristic of the slowly varying mesoscale eddy field representing the dynamics resulting from baroclinic instability. This regime exhibits eddy-like features that have spatial scales on the order of the Rossby radius of deformation.
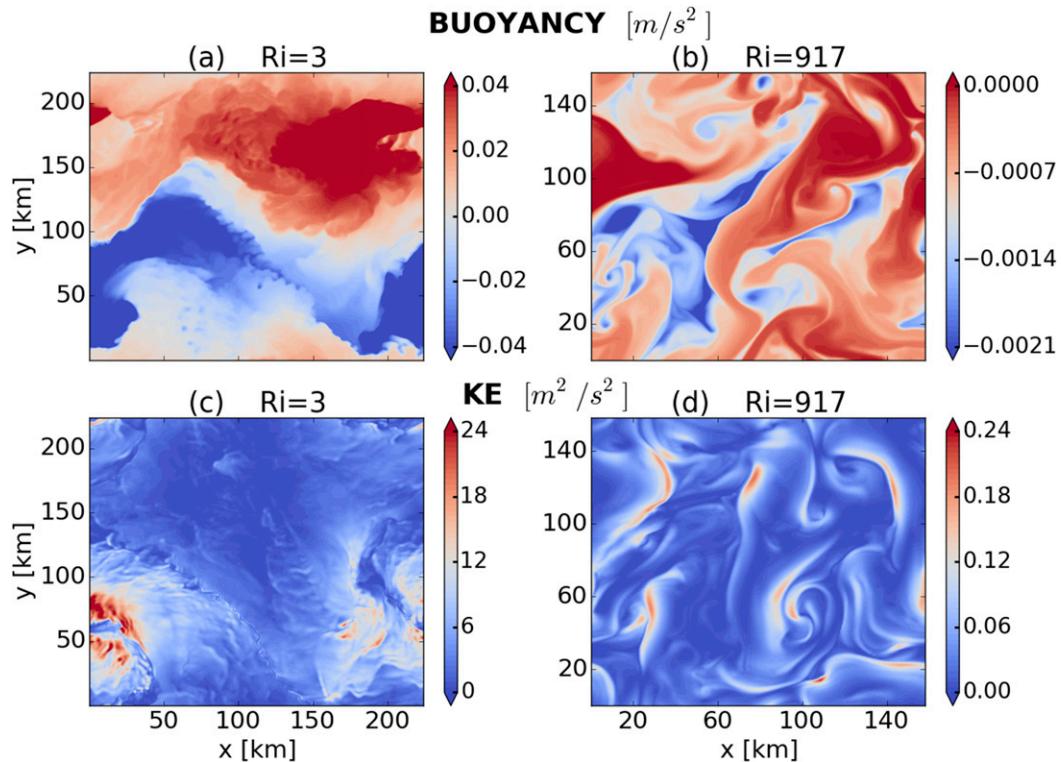
FIG. 2. Snapshots of (a),(b) buoyancy and (c),(d) KE at a depth of 1.25 m from the surface in the statistically stationary state of the high-resolution model simulations. The differences in the structure and magnitude are clearly evident between the dynamical regimes with (left) RI = 3 and (right) RI = 917, equivalent of an ageostrophic and quasigeostrophic regime, respectively.

The features seen in Fig. 2 hence result from different dynamics.

Snapshots of the vertical velocity corresponding to the Ri in Fig. 2 are shown in Fig. 3. These suggest the existence of wavelike features for both ageostrophic and quasigeostrophic regimes, similar to, for example, what is described in Plougonven and Snyder (2007) and similar to the features seen for horizontal maps of vertical velocity in numerical simulations [e.g., Wu and Eckermann (2008), Shutts and Vosper (2011) for global data, and Plougonven et al. (2015) for a mesoscale case study]. These wavelike features manifest themselves as wave trains, as can be seen for instance in the vertical velocity for Ri = 3 in Fig. 3. Since the crests and troughs seen in the figure

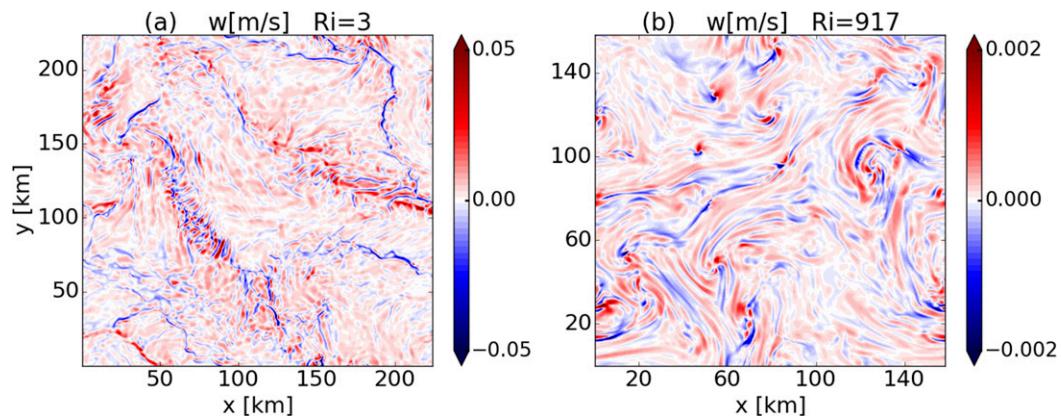

FIG. 3. Snapshots of vertical velocity at about 20-m depth, for Ri = (a) 3 and (b) 917 from the statistically stationary state of the model runs. Note that the color scales are different.

above are akin to gravity wave activity, those signals are accordingly interpreted by, for example, Plougonven and Snyder (2007) as gravity waves generated by the baroclinic instability process in the simulations. It is the aim of this study to investigate if those signatures are indeed gravity waves in a more qualitative manner or vertical velocities associated with the balanced mode. We proceed further to explore these wave signals in the Fourier space.

3. Analysis in Fourier space

a. Frequency spectrum

The presence of wavelike features in the physical space motivates us to delve further into Fourier space to identify the characteristic properties of these wave signals. A frequency (ω) spectrum of KE at 50-m depth for different Ri is shown in Fig. 4. The spectra shown are averages of eight[3] chunks, each of 45-day period, of the statistically stationary integrations of the model setup and averages over the model domain. We show results from only one depth, but frequency spectra calculated at other depths give similar results. Most of the energy is concentrated at the smallest frequencies, that is, the spectrum is red, but a certain amount of energy is also contained in superinertial frequencies with $\omega > f$ indicative of gravity waves, in particular for small Ri. The percentage of the KE content above the inertial frequency f is also indicated in the figure. It shows that the relative energy level contained in the superinertial frequencies is much higher for an ageostrophic regime than it is for other dynamical regimes. This energy in the superinertial frequencies could be associated with gravity waves, which have frequencies higher than f. On the other hand, the gravity waves can be Doppler shifted by the mean flow such that the frequency analysis alone does not provide a clear separation of the balanced and gravity mode, for which a frequency and wavenumber spectrum is better suited.

b. Energy in vertical modes

We begin with the vertical wavenumber and consider the energy distribution in vertical modes. All model variables are projected on vertical modes by transformations in the vertical, that is, discrete sine transformation of the vertical velocity w and buoyancy b, and a discrete cosine transformation of the horizontal velocity u, v and pressure p. The vertical eigenvalues m for $N = $ constant are given by $m = n\pi/H$ for the vertical
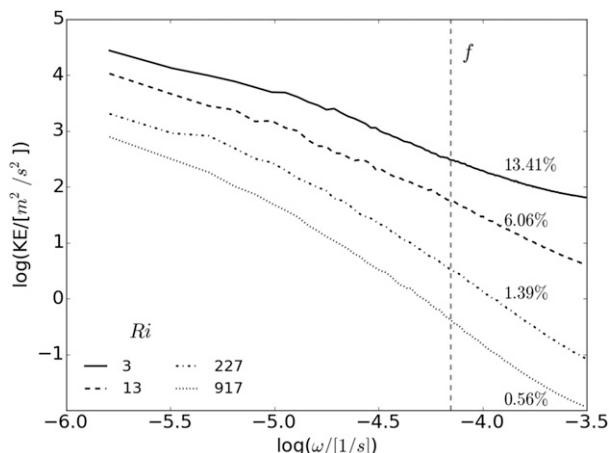

FIG. 4. Frequency spectrum for KE for different Ri. The spectrum shown is an eight-chunk average and is shown at 50-m depth. The vertical dashed gray line indicates the Coriolis frequency f, and the numbers to its right indicate the percentage of energy contained in the frequencies larger than f for each Ri. The values are similar also at other depths (not shown).

mode number $n = 0, 1, 2, 3, \ldots$. After decomposition into the vertical modes we calculate KE, available potential energy (APE), and total energy (TE) contained in each mode. APE is defined here as $P = b'^2/(2N_m^2)$, where $b' = b - N_m^2 z$ gives the difference between the local buoyancy b and the reference buoyancy $N_m^2 z$ of the time and global mean of N_m^2, which is the stratification of the equilibrated flow.

The distribution of TE and KE as a fraction of TE in the barotropic and first four baroclinic modes is shown in Table 2 for different Ri, again using eight chunks of statistically stationary integrations of the model setup. The values shown in Table 2 are averaged in time and horizontally. The breakdown of energy into vertical modes shows that both KE and APE (the remaining fraction of TE in Table 2) decrease in general with higher vertical modes, although APE appears to share a larger portion of TE for higher modes. For the first baroclinic mode KE dominates APE for small Ri while the reverse is true for large Ri, whereas for higher baroclinic modes APE dominates KE for all Ri.

Further, frequency and wavenumber spectra of KE in different vertical modes (not shown) show distinct differences between odd and even modes. TE is higher for even modes than it is for the odd modes, whereas KE is much higher for odd modes than even modes for all Ri. This disparity between odd and even modes might be related to the behavior of the fastest-growing mode in the simulations. To test this we project the fastest-growing mode ϕ on the vertical eigenfunctions Φ_n (see appendix A for a detailed derivation). The projection can be written as $\phi(z) = \sum_{n-1}^{\infty} A_n \Phi_n(z)$ and upon solving

---

[3] In Fig. 5 the spectra shown or Ri = 3 are averaged using 13 chunks.

TABLE 2. TE ($m^2 s^{-2}$) and KE as a fraction of TE in the first five vertical modes for different regimes indicated by Ri. TE and KE are averaged in time and horizontally after they are computed from vertically decomposed buoyancy and horizontal velocities.

Ri →	3		13		327		917	
Vertical modes ↓	TE	KE/TE	TE	KE/TE	TE	KE/TE	TE	KE/TE
0	819.53	1.00	159.02	1.00	29.56	1.00	13.24	1.00
1	18.52	0.76	2.46	0.68	0.18	0.22	0.06	0.11
2	64.10	0.03	10.23	0.02	4.71	0.0001	4.24	0.000 09
3	1.97	0.38	0.30	0.26	0.04	0.05	0.01	0.02
4	25.90	0.01	4.15	0.007	1.94	0.0005	1.74	0.000 03

the coefficient A_n takes the following form for baroclinic modes:[4]

$$A_n = \frac{2d}{H(1 + n^2\pi^2 d^2/H^2)}[(-1)^n \chi + 1]. \quad (5)$$

Here $d = f/Nk_h^2$, k_h is the horizontal wavenumber, and χ is a function of d (for the full expression of χ refer to appendix A). Since the energy in vertical modes depends on the coefficient A_n, and A_n in turn is proportional to $[(-1)^n x + 1]$, this disparity between modes might be inferred from Eq. (5) (f, N, and H being constant). The decisive factors in this expression are $(-1)^n$ and the $+1$ in the square brackets; the factor $(-1)^n$ acts as a switch and gives rise to the odd–even nonuniformity in the modes. Note, however, that this projection is performed for the Eady modes that might differ from the ones in the model since the model is based on primitive equations (section 2a). Although the model background state is not the same as the Eady state, we find a similar behavior in our simulations. That being so, the projection explains that the energy distribution not only tends to decrease with increasing vertical modes, it also shows a distinct distribution between odd and even modes. Further, we extend the analysis to the frequency and horizontal wavenumber space, to better identify gravity waves in the model simulations.

c. Frequency–wavenumber analysis

KE and APE in frequency–wavenumber (ω–k_h) space are obtained from a three-dimensional Fourier transform of the horizontal velocity in time and space (zonal and meridional direction), for different vertical modes. The zonal (k) and meridional (l) wavenumbers are collapsed together to give the horizontal wavenumber k_h. Variance preserving ω–k_h spectra of KE [shown as $\log_{10}(\omega k_h \times KE)$] for the first and second baroclinic modes ($n = 1, 2$) for Ri = 3 and Ri = 917 are shown in

Fig. 5. The figure also shows the shallow water gravity wave dispersion relation, which can be expressed as $\omega^2 = f^2 + gHk_h^2$ (barotropic mode) and $\omega^2 = f^2 + c_n^2 k_h^2$ (baroclinic modes), where g is acceleration due to gravity and $c_n = N/m$. The dispersion relation mentioned in the rest of the paper refers to this shallow water gravity wave dispersion relation. As the wave's frequency can be influenced by the Doppler shift, we show $\omega \pm U_0 k_h$ in the figure, where U_0 is the mean flow (see Table 1). We assume that the possible region for gravity wave lies approximately within the envelope of the Doppler-shifted extrema. We henceforth call this guideline region the gravity wave branch.

As is evident from Fig. 5, there is a substantial amount of energy in the gravity wave branch for Ri = 3 for both modes, while it is much smaller for Ri = 917 and outside of the gravity wave branch. Instead, most of the energy for Ri = 917 is located at the wavelength of the fastest-growing mode confirming that there is not much energy related to gravity waves for a quasigeostrophic regime. For an ageostrophic regime, on the contrary, the energy in the gravity wave branch suggests that ageostrophic dynamics resulting from baroclinic instability at small Ri could generate a significant amount of gravity wave energy. However, an ω–k_h spectrum is not enough to confirm this statement because the energy of the balanced mode could also be within the gravity wave branch. Especially for Ri = 3 in Fig. 5, the balanced mode lies mostly within the gravity wave branch. The coexistence of these processes makes it difficult to isolate the energy contributions from gravity waves or unbalanced modes and the balanced mode. To treat this difficulty and to clearly ascribe this energy to the gravity waves, we employ a modal decomposition method to decompose the full flow vector into these two modes, elaborated in the next section.

4. Modal decomposition: Balanced mode and unbalanced gravity modes

We use here a linear modal decomposition to diagnose the gravity wave oscillations in our simulations

---

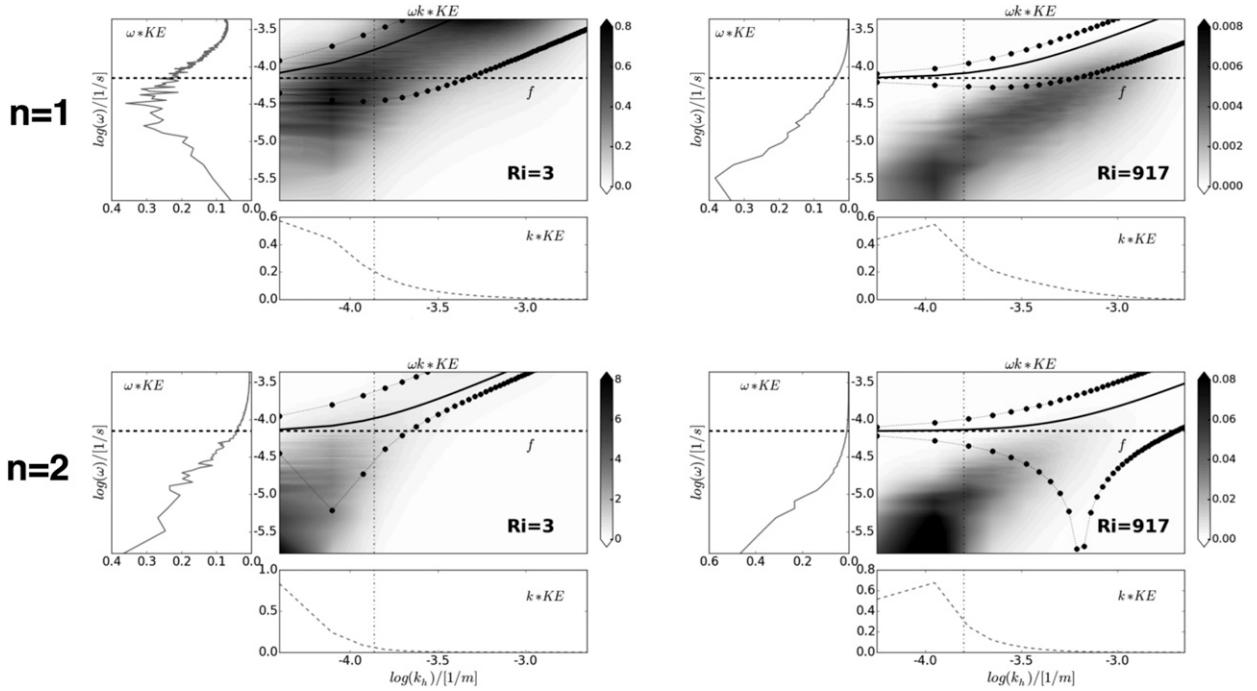[4] For barotropic mode, the factor 2 vanishes in the RHS expression of A_n. See appendix A.

FIG. 5. Variance-preserving frequency–wavenumber spectrum (log–log scale) for KE for the (top) first and (bottom) second baroclinic modes, compared for Ri = (left) 3 and (right) 917. Each subplot is as follows: in the left panel, a solid gray curve shows $\omega \times$ KE as a function of frequency ω averaged in wavenumber space; in the bottom panel, the dashed gray curve shows $k_h \times$ KE as a function of horizontal wavenumber k_h averaged in frequency space, and the dashed–dotted gray line indicates k_{max}, the wavenumber of the fastest-growing mode; in the center panel, the shading shows $\omega k_h \times$ KE as a function of ω and k_h, the solid black curve indicates the gravity wave dispersion relation, and lines with black circles indicate the Doppler shifted dispersion relation. The gravity wave branch is enclosed by the Doppler-shifted extrema. The dashed black line in the left and center panels represents the Coriolis frequency f. The spectra are calculated and averaged from eight chunks of 45-day length each from the model setup.

and extend this decomposition by an NLNMI by Machenhauer (1977), used in numerical weather prediction to generate an appropriate balanced initial state. This section presents an overview of the decomposition methods (details appear in appendix B).

The hydrostatic and Boussinesq system of equations,

$$\partial_t u + \mathbf{u} \cdot \nabla u = -\partial_x p + f v, \quad \partial_t v + \mathbf{u} \cdot \nabla v = -\partial_y p - f u,$$
$$\partial_t b + \mathbf{u} \cdot \nabla b = -w N^2, \tag{6}$$

complemented by the diagnostic relations $\partial_z p = b$ and $\nabla \cdot \mathbf{u} = 0$, can be written for the state vector $\mathbf{x}$ containing the relevant state variables as

$$\partial_t \mathbf{x} = i\mathbf{L} \cdot \mathbf{x} + \mathcal{N}(\mathbf{x}), \tag{7}$$

where $\mathbf{L} \cdot \mathbf{x}$ contains all the linear terms and the vector $\mathcal{N}$ contains the nonlinear and forcing terms.

After decomposition into vertical modes and then a Fourier transformation in x and y, the system in matrix notation yields

$$\tilde{\mathbf{x}} = \begin{pmatrix} \tilde{u} \\ \tilde{v} \\ \tilde{p} \end{pmatrix}, \quad \mathbf{L} = \begin{pmatrix} 0 & -if & -k \\ if & 0 & -l \\ -kc_n^2 & -lc_n^2 & 0 \end{pmatrix}, \tag{8}$$

where a Fourier-transformed quantity is represented by a tilde ($\sim$). The spectrum of the matrix $\mathbf{L}$, which is the set of eigenvalues of $\mathbf{L}$, describes the characteristic frequencies of the system and are given by $\omega = \omega^0 = 0$ and $\omega = \omega^\pm = \pm\sqrt{f^2 + c_n^2 k_h^2}$. Notice that $\omega^\pm$ resembles the shallow water dispersion relation for gravity waves with $c_{n>0} = N/m = NH/(n\pi)$ and $c_0 = \sqrt{gH}$.

The corresponding right ($q^{0,\pm}$) and left ($p^{0,\pm}$) eigenvectors[5] to the matrix $\mathbf{L}$ then give the state vector $\tilde{\mathbf{x}}$ in Fourier space, which can be expressed as follows:

$$\tilde{\mathbf{x}} = b \, q^0 + g^+ q^+ + g^- q^- \quad \text{with}$$
$$b = p^0 \cdot \tilde{\mathbf{x}}, \quad g^+ = p^+ \cdot \tilde{\mathbf{x}}, \quad g^- = p^- \cdot \tilde{\mathbf{x}}. \tag{9}$$

---

[5] For a given matrix, a right eigenvector is a column vector while a left eigenvector is a row vector. In the context of matrices, the commonly used "eigenvector" is the right eigenvector. Here we use the two eigenvectors separately. See appendix B for details.

This yields a projection of the state vector $\tilde{\mathbf{x}}$ on three eigenmodes or normal modes: one balanced slow mode and two unbalanced fast gravity wave modes corresponding to the characteristic frequencies ω^0 and $\omega^{\pm}$, respectively. Henceforth, we call these modes as balanced mode and unbalanced gravity modes (or simply unbalanced modes). The left and right eigenvectors are given explicitly in appendix B.

We define vectors $\tilde{\mathbf{x}}_B$ and $\tilde{\mathbf{x}}_G$ associated with the balanced manifold and unbalanced manifold, respectively:

$$\tilde{\mathbf{x}}_B = b\,q^0 = \mathbf{B} \cdot \tilde{\mathbf{x}}, \quad \text{and} \tag{10}$$

$$\tilde{\mathbf{x}}_G = g^+ q^+ + g^- q^- = \mathbf{G}^+ \cdot \tilde{\mathbf{x}} + \mathbf{G}^- \cdot \tilde{\mathbf{x}}. \tag{11}$$

Since the matrix $\mathbf{L}$ is Hermitian, the eigenvectors of $\mathbf{L}$ are mutually orthogonal, and so are the balanced and unbalanced modes and the associated linear manifolds. As the manifolds $\mathbf{B}$ and $\mathbf{G}$ are mutually orthogonal, they span the whole phase space, and therefore the vector $\tilde{\mathbf{x}}$ can be written as a linear combination of the two modes implying that $\tilde{\mathbf{x}} = \tilde{\mathbf{x}}_B + \tilde{\mathbf{x}}_G$ holds true (cf. Leith 1980; Theiss and Mohebalhojeh 2009). After reverse transformation from $\tilde{\mathbf{x}}$ to $\mathbf{x}$ (i.e., $\tilde{\mathbf{x}}_B$ to $\mathbf{x}_B$ and $\tilde{\mathbf{x}}_G$ to $\mathbf{x}_G$) the energy contained in the balanced and unbalanced gravity modes can be obtained. The $\mathbf{L}$ and $\mathbf{G}$ commute (i.e., $\mathbf{L} = \mathbf{G}$) because they share the same set of eigenvectors, and so any vector in the $\mathbf{G}$ manifold can be written as a linear combination of eigenvectors of $\mathbf{L}$.

a. Linear modal decomposition

In the linear case [$\mathcal{N} = 0$ in Eq. (7)] any $\tilde{\mathbf{x}}$ can be projected on the slow linear manifold. This part of $\tilde{\mathbf{x}}$ (i.e., $\tilde{\mathbf{x}}_B$) becomes stationary,

$$\partial_t \tilde{\mathbf{x}}_B = i\mathbf{L} \cdot \tilde{\mathbf{x}}_B = ib\,\omega^0 q^0 = 0, \tag{12}$$

and only the fast modes will evolve in time according to

$$\partial_t \tilde{\mathbf{x}}_G = i\mathbf{L} \cdot \tilde{\mathbf{x}}_G = \sum_{d=\pm} ig^d \omega^d q^d. \tag{13}$$

The balanced and the gravity modes are then contained in $\tilde{\mathbf{x}}_B$ and $\tilde{\mathbf{x}}_G$, respectively. We call the modes resulting from the linear normal mode decomposition of baroclinically unstable fully nonlinear model state the linear balanced mode (BAL_LIN) and linear unbalanced modes (UNB_LIN).

b. Nonlinear modal decomposition

The linear modes refer to the linear system and using them for the realistic nonlinear case does not provide a consistent decomposition. To handle this discrepancy we include in the decomposition the nonlinearity in the

system. For $\mathcal{N} \neq 0$, the time evolution of the slow mode is nonzero and its behavior becomes

$$\partial_t \tilde{\mathbf{x}}_B = \tilde{\mathcal{N}}(\mathbf{x}_B) \neq 0, \tag{14}$$

where $\mathbf{x}_B$ is the inverse Fourier transform of $\tilde{\mathbf{x}}_B$ and $\tilde{\mathcal{N}}$ the Fourier transform of $\mathcal{N}$. One approach to separate the modes in the nonlinear case is to choose a state that eliminates the fast modes such that only the slow mode remains, whose difference with the full vector then gives the isolated fast modes. According to Machenhauer (1977), the time changes in the fast modes (i.e., $\mathbf{G} = \sum_{d=\pm} \mathbf{G}^d$) can be eliminated by a nonlinear normal mode initialization technique requiring that the time derivative of the fast modes is zero; that is,

$$\mathbf{G} \cdot \partial_t \tilde{\mathbf{x}} = 0, \tag{15}$$

which for Eq. (7) becomes

$$\mathbf{G} \cdot \partial_t \tilde{\mathbf{x}} = i\mathbf{G} \cdot (\mathbf{L} \cdot \tilde{\mathbf{x}}) + \mathbf{G} \cdot \tilde{\mathcal{N}}(\mathbf{x}) = 0 \rightarrow (\mathbf{L} \cdot \mathbf{G}) \cdot (\mathbf{G} \cdot \tilde{\mathbf{x}})$$
$$= i\mathbf{G} \cdot \tilde{\mathcal{N}}(\mathbf{x}) \rightarrow \mathbf{G} \cdot \tilde{\mathbf{x}} = i(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G} \cdot \tilde{\mathcal{N}}(\mathbf{x}). \tag{16}$$

Here the linear operator $(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G} = \sum_{d=\pm}(\omega^d)^{-1}\mathbf{G}^s$ operates in the gravity mode space only where the eigenvalues are nonzero to avoid problems by singularities for the inversion $(\mathbf{L} \cdot \mathbf{G})^{-1}$ as claimed by Leith (1980).

Equation (16) is a nonlinear condition on $\mathbf{x}$, which is proposed by Machenhauer (1977) to be solved iteratively until convergence is reached, which is usually the case after a few steps. Starting with a linear slow mode $\mathbf{x}_B$ the iteration for the initialization technique is given by the following:

$$\tilde{\mathbf{x}}_1 = \tilde{\mathbf{x}}_B + i(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G} \cdot \tilde{\mathcal{N}}(\mathbf{x}_B), \tag{17}$$

$$\tilde{\mathbf{x}}_2 = \tilde{\mathbf{x}}_B + i(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G} \cdot \tilde{\mathcal{N}}(\mathbf{x}_1), \tag{18}$$

$$\dots, \quad \text{and} \tag{19}$$

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_B + i(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G} \cdot \tilde{\mathcal{N}}(\mathbf{x}_{k-1}). \tag{20}$$

It was shown by Leith (1980) that the first iteration step corresponds to the quasigeostrophic approximation. Hence, we use only the first step in our analysis; further iterations do not change the results much. The result of the iteration can now be used to calculate the nonlinear balanced mode; the difference of this balanced mode to the actual state vector can be interpreted as the nonlinear gravity mode. We call the modes resulting from the nonlinear decomposition as the nonlinear balanced
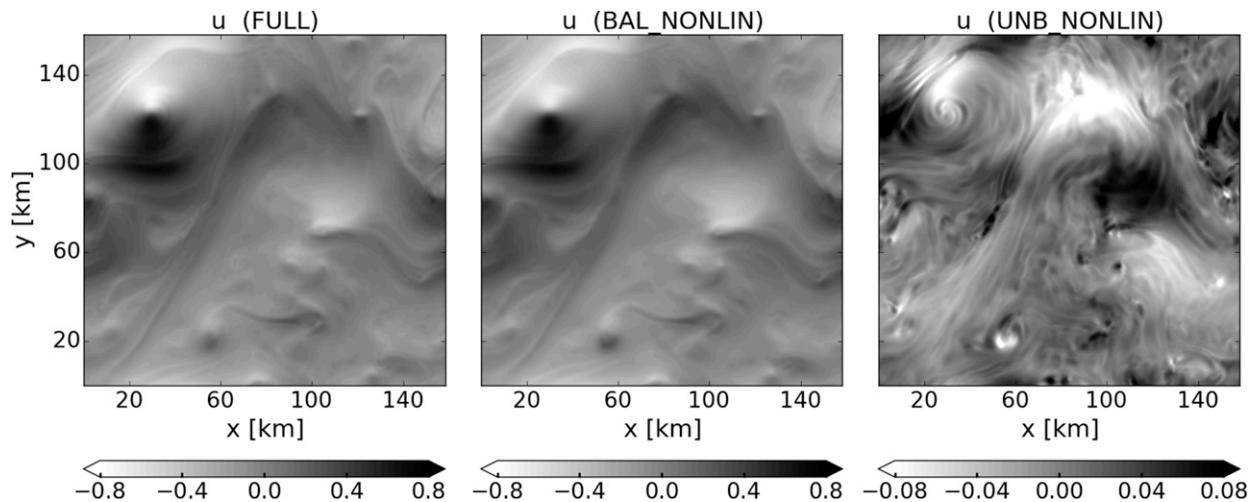
FIG. 6. Snapshots of zonal velocity for Ri = 327 with its (left) full component (FULL) and modally decomposed components: (center) balanced (BAL_NONLIN) and (right) unbalanced (UNB_NONLIN) modes using nonlinear decomposition. Note the difference in the magnitudes between the unbalanced component and the full and balanced components.

mode (BAL_NONLIN) and nonlinear unbalanced modes (UNB_NONLIN).

c. Decomposition results

As stated above, a nonlinear decomposition is more suitable for a fully nonlinear model state of a baroclinically unstable flow, such as our setup, so we first present an example of the nonlinear decomposition in physical space. Next we consider both the decompositions by means of ω–k_h spectra and elaborate on the differences. Note, however, that the nonlinear balanced and unbalanced modes do not show orthogonality, and thus the spectra related to the nonlinear decomposed components will not add up to the full component. Figure 6 shows snapshots of zonal velocity for Ri = 327 for the full velocity (FULL) and velocities from the nonlinear balanced mode (BAL_NONLIN) and nonlinear unbalanced modes (UNB_NONLIN) as an example of the nonlinear decomposition. For the balanced mode more large-scale features are present as in the full component, whereas the unbalanced mode shows indeed more small-scale features akin to gravity wave activity. Also, note that the magnitudes of the full and balanced components are one order higher than that of the unbalanced component. This separation is a first indication that the decomposition of the balanced and unbalanced modes using the modal decomposition is effective. To demonstrate this more quantitatively we apply next a frequency–wavenumber analysis to the decomposed fields.

The ω–k_h spectra (see Figs. 7, 8, and 9) are computed similarly to the method described in section 3c and then averaged for all depths for eight chunks of 45-day length

each. Recall from section 1 that the coupling between balanced and unbalanced modes tends to be weaker for the Ri $\gg$ 1 regime such that the temporal scales of the slow balanced and the fast unbalanced motions are well separated. The opposite is true for small Ri where this coupling is much stronger and the separation of the modes gets more difficult. We tackle the less complicated case first (for Ri $\gg$ 1) before expounding on a case with small Ri.

An ω–k_h spectra of KE for Ri = 327 is shown in Fig. 7 for modes obtained from the linear and nonlinear decomposition. As the balanced modes correspond to the motions with large temporal scales, in ω–k_h space the energy associated with the balanced modes is expected to lie toward low frequencies, away from the high frequencies. The energy associated with the unbalanced motions, on the other hand, is expected to lie in the region confined by the gravity wave branch, which is the superinertial frequency range that allows for gravity waves enveloped by the corresponding Doppler-shifted dispersion relation. The expectation is fulfilled in the case of balanced modes (BAL_LIN, BAL_NONLIN), as seen from Figs. 7a,b, where most of the KE lies outside the gravity wave branch and toward small frequencies and wavenumbers.

However, also for the unbalanced modes (UNB_LIN, UNB_NONLIN) (Figs. 7c,d) most of the KE lies outside the gravity wave branch against the expectation. However, KE in the gravity wave branch increases from the linear unbalanced modes to the nonlinear unbalanced modes (cf. Figs. 7c and 7d). The nonlinear decomposition in Fig. 7d indeed shows an increase of KE that sits inside the gravity wave branch, as compared to the linear
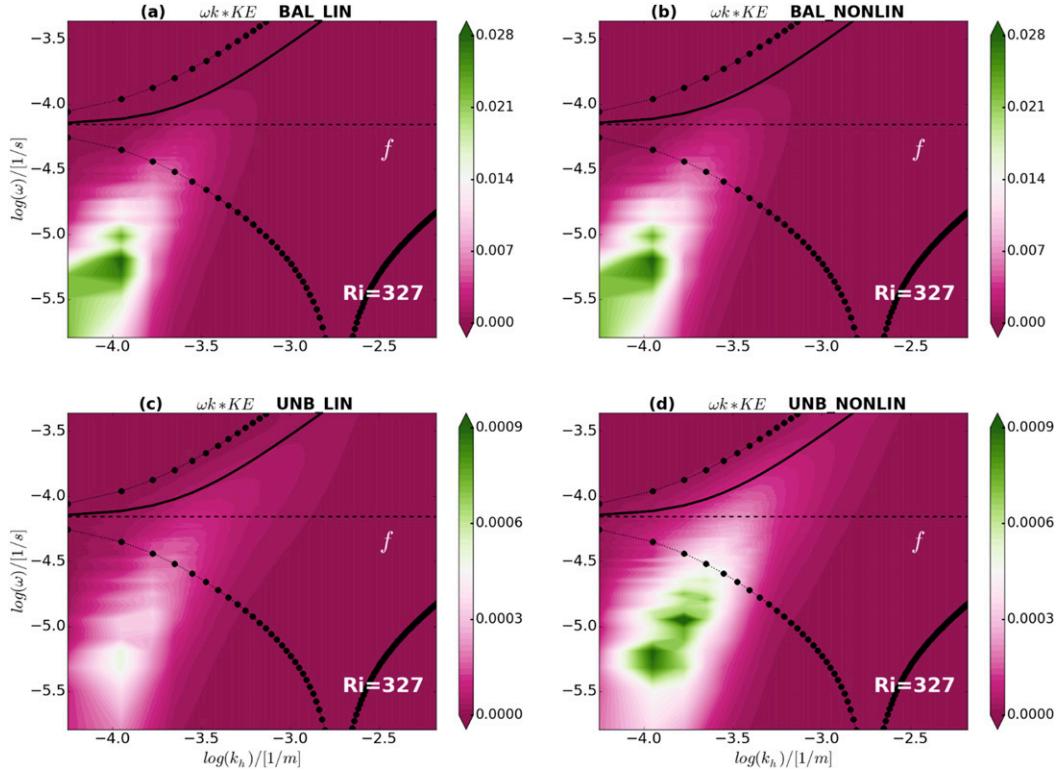
FIG. 7. Variance preserving frequency–wavenumber spectra of KE of the (a),(b) balanced and (c),(d) unbalanced modes for Ri = 327 averaged over all depths. The spectra are calculated and averaged from eight chunks of 45-day length each. The spectra in (a) and (c) are obtained from the linear decomposition (BAL_LIN, UNB_LIN) and the ones in (b) and (d) from the nonlinear decomposition (BAL_NONLIN, UNB_NONLIN). In each subplot, the color shading shows the respective KE as a function of ω and k_h, the dashed black line indicates the Coriolis frequency f, the solid black curve indicates the gravity wave dispersion relation, and lines with black circles show the Doppler-shifted dispersion relation. Note that the gravity wave branch is enclosed by the Doppler-shifted extrema. Notice a magnitude difference of two orders between the balanced and unbalanced modes.

decomposition in Fig. 7c. Thus, the nonlinear decomposition appears to be an improvement over the linear one and a better-suited tool than a linear decomposition to decompose the balanced and unbalanced flow components. We find similar improvement for the other model experiments (not shown, but listed in Table 3). However, there is still energy outside the gravity wave branch in Fig. 7d. This could be related to a misinterpretation of the gravity wave branch, since, as mentioned in section 3c, we use the mean flow prescribed initially U_0 (see Table 1) to calculate the Doppler shift. This mean flow might not be well suited for an "effective" Doppler shift of the waves, which means that the gravity wave branch would also change. On the other hand, the residual unbalanced energy outside the gravity wave branch could also be related to an (unknown) artifact of the nonlinear decomposition. A possibility for such an artifact is that we use the eigenvectors $\mathbf{q}^{0,\pm}$ and $\mathbf{p}^{0,\pm}$ of the analytical instead of the discrete system for the method. We will explore this issue further in later

studies, and in this paper, we consider the results only from the nonlinear decomposition since they show improvement with respect to our expectation.

Further, notice the difference of two orders in magnitude between the balanced and unbalanced modes in Fig. 7, stating that a significant amount of KE is contained in the balanced mode for large Ri. The negligible amount of KE in the gravity wave branch even for the unbalanced modes signifies that the gravity wave emission is weak in a Ri ≫ 1 regime. We now consider examples from other Ri.

Balanced (BAL_NONLIN) and unbalanced (UNB_NONLIN) modes from the nonlinear decomposition for the extreme Ri (=3) and an intermediate Ri (=13) in our simulations are shown in Fig. 8. For the balanced modes (see Figs. 8a,c) for Ri = 3 and Ri = 13, most of the KE lies outside the gravity wave branch, but some KE is also within this branch. This is associated with the strong coupling between the balanced and unbalanced motions for small Ri, for which the time-scale separation
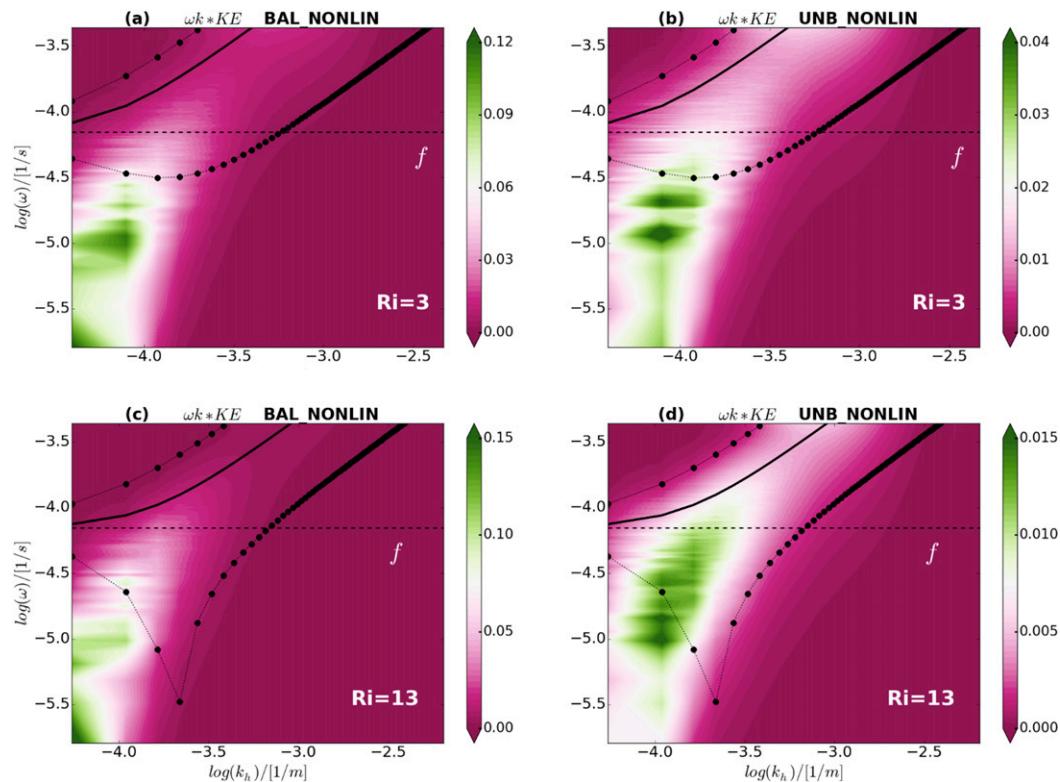
FIG. 8. Variance-preserving frequency–wavenumber spectra of KE similar to Fig. 7 but for different Ri: (top) 3 and (bottom) 13, from the nonlinear decomposition. (a),(c) The spectra for the nonlinear balanced modes (BAL_NONLIN) and (b),(d) the ones for the nonlinear unbalanced modes (UNB_NONLIN) are shown. The gravity wave branch is enclosed by the Doppler-shifted extrema. Note the difference in the color scales.

of these modes is not well defined. In contrast, most of the KE in the balanced mode for the aforementioned Ri = 327 clearly lies outside this branch. Also, the balanced mode has much higher KE for Ri = 327 than for smaller Ri.

This higher KE for the balanced component of Ri = 327 is not surprising, since as one moves from ageostrophy toward quasigeostrophy, one progresses toward a more "balanced" state, and not unexpectedly would one find balanced modes dominating the unbalanced modes. Put in other words, it implies that a regime in a quasigeostrophic balance will emit weak unbalanced gravity waves. In the context of modal decomposition, this suggests that for a regime in a quasigeostrophic balance the energy in unbalanced modes would be far less than in an ageostrophic regime. This conjecture is supported by Figs. 8b,d and Fig. 7d for the nonlinear unbalanced modes. The figures clearly illustrate the negligibly small KE in the gravity wave branch for the unbalanced modes of Ri = 327 in contrast to Ri = 3 and 13, where a significant amount of KE is concentrated within the gravity wave branch. The energy in the gravity wave branch for the unbalanced modes becomes even smaller

for the higher Ri = 917 (not shown). For the intermediate Ri = 13, KE in the unbalanced mode is aligned along the gravity wave dispersion relation.

Further we discuss an ω–k_h spectra for APE for the linear and nonlinear decompositions, and as an example we show the APE spectra for Ri = 13 in Fig. 9. The APE spectra exhibit a distribution between balanced and unbalanced modes similar to what is described before for KE (Figs. 8b,c). In the balanced modes (BAL_LIN, BAL_NONLIN) (Figs. 9a,b), most of the APE is present at lower frequencies, whereas for the unbalanced modes (UNB_LIN, UNB_NONLIN) APE tends to be present at higher frequencies. However, there appears to be more APE in the nonlinear modes (BAL_NONLIN, UNB_NONLIN) than in the linear modes (BAL_LIN, UNB_LIN), and unbalanced modes show more APE than the balanced modes.

The decomposition can now be used to quantify the energy associated with the balanced and the unbalanced modes. Table 3 shows the ratio of total KE (and APE) associated with the unbalanced modes to the total KE (and APE) associated with the balanced mode for the linear and nonlinear decompositions (i.e., the ratio
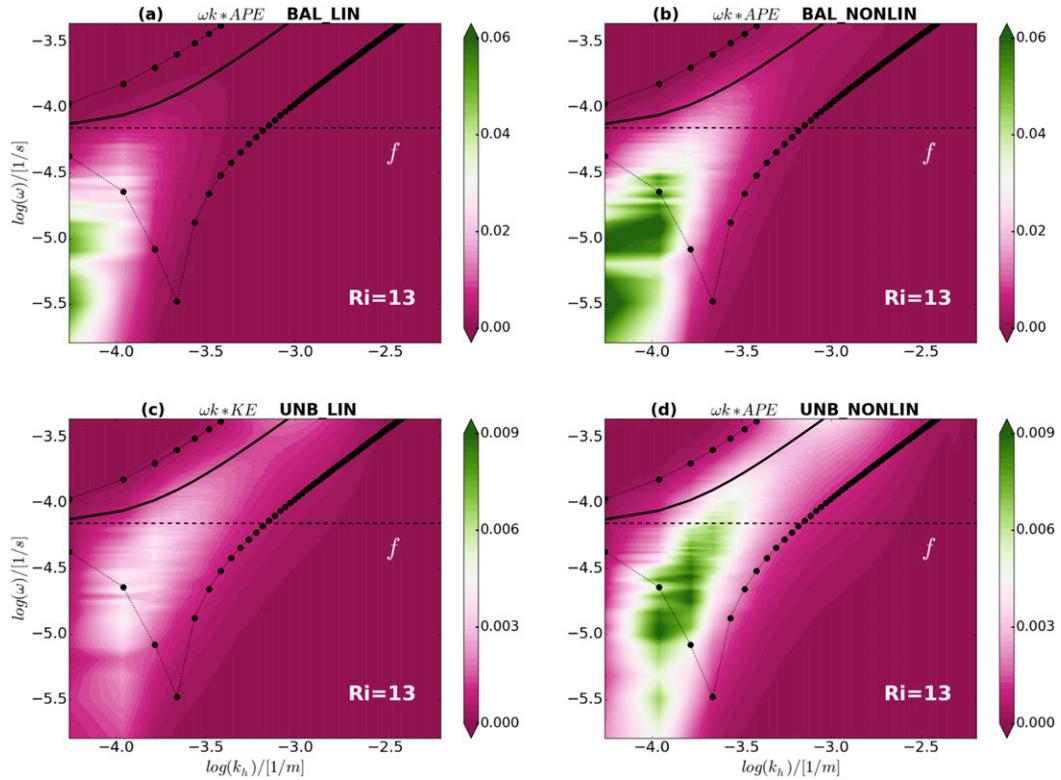
FIG. 9. Variance-preserving frequency-wavenumber spectra of APE similar to Fig. 7 but for Ri = 13. (a),(b) The balanced modes and (c),(d) the unbalanced modes. The spectra in (a) and (c) are obtained from the linear decomposition (BAL_LIN, UNB_LIN), and the ones in (b) and (d) from the nonlinear decomposition (BAL_NONLIN, UNB_NONLIN). The gravity wave branch is enclosed by the Doppler-shifted extrema. Notice that the color scales differ for unbalanced and balanced modes.

UNB/BAL for KE and APE). For all Ri, the fraction of KE (or APE) in the nonlinear unbalanced mode increases as compared to the linear one. This reinforces the point that the nonlinear decomposition is better suited as a decomposition tool than the linear decomposition. Not only on theoretical grounds but also in terms of the energy associated with the linear and nonlinear modes, the nonlinear decomposition is capable of better extracting wave signals related to the nonlinear gravity wave field. Further, in line with the results seen so far, the fraction of KE (and APE) related to the unbalanced modes increases as Ri decreases. In particular, for the nonlinear decomposition the regime with Ri = 3 shows a larger share of the unbalanced KE (about 37%), whereas this fraction associated with the nonlinear unbalanced modes reduces drastically to less than 2% for Ri = 917. The trend for APE is similar: Ri = 3 shows a larger share of the unbalanced APE (about 37%), whereas Ri = 917 shows less than 4% of this fraction, for the nonlinear decomposition.

The energy contributions tied to the balanced and unbalanced modes can be further utilized to understand the variation of energy with the dimensionless numbers that describe different flow regimes in our simulations. For this, a power law is obtained from the KE and APE of nonlinear unbalanced component that describes the variation of IGW emission with Ri and Ro. Based on the four simulations with different Ri, the decrease of the KE tied to the unbalanced component scales close to $Ri^{-1.05}$ (or $Ro^{2.1}$). A similar scaling is seen for APE for which the APE in the unbalanced component shows a scaling of $Ri^{-1.3}$ (or $Ro^{2.6}$). This is illustrated in Fig. 10 for the scaling of KE and APE in the nonlinear

TABLE 3. Ratios of KE and APE of the unbalanced component to that of the balanced component (UNB/BAL) obtained from linear and nonlinear decompositions, shown for different regimes indicated by Ri. The numbers indicate percentages.

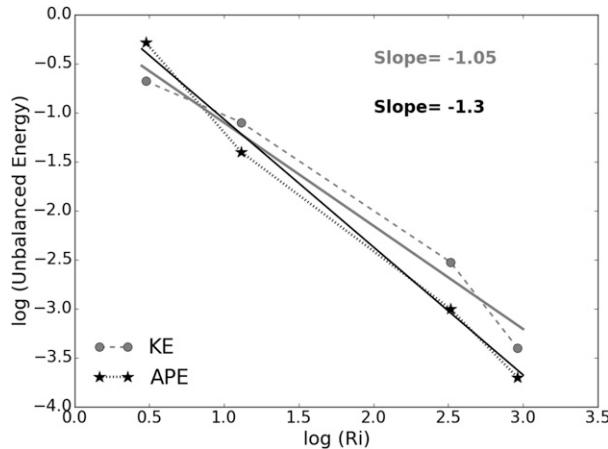| Ratio →	KE (UNB/BAL)		APE (UNB/BAL)	
Ri ↓	Linear	Nonlinear	Linear	Nonlinear
3	21.42	36.84	33.33	37.14
13	7.69	13.55	10.52	13.62
327	1.11	3.33	5.40	6.53
917	1.0	1.48	3.50	3.99

FIG. 10. Power laws for the KE and APE related to the nonlinear unbalanced component for different dynamical regimes indicated by Ri. The gray dashed line with circles represents KE and the black dotted line with stars indicates APE. The fitted slopes related to KE and APE are shown in gray and black solid lines, respectively.
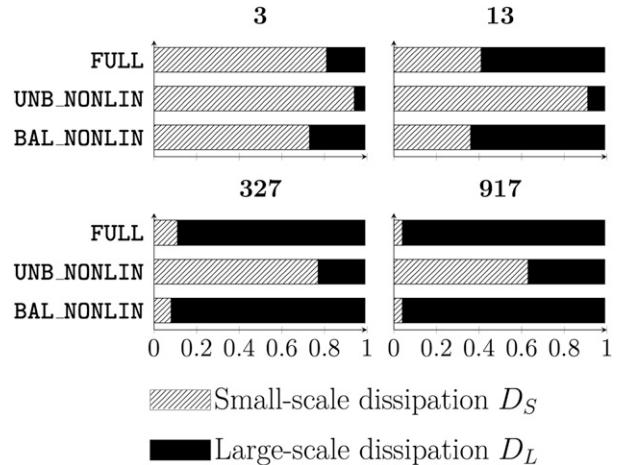

FIG. 11. Globally integrated values of KE dissipation for the full and the nonlinear balanced and unbalanced components (BAL_NONLIN, UNB_NONLIN) for all Ri. Dissipation contribution to the total KE dissipation from small scales (D_S) is shown by the hatched region and from large scales (D_L) by the black shaded region. Note that $D_S + D_L = 1$ for the respective mode.

unbalanced component with Ri. Also indicated are the respective slopes computed from the KE and APE in the unbalanced component. Although our results are based on an idealized representation of the oceanic flow field, such that these power laws may be considered unrealistic, the results do give a first indication of the relationship between IGW emission and Ri (or Ro) and provide estimates for the importance of this process in the ocean.

5. Energy dissipation under different dynamical regimes

The energy in the model setup is dissipated at small scales by horizontal biharmonic friction and harmonic vertical friction and at large scales by the zonal mean drag. We use the modal decomposition results to assess the differences in small- and large-scale dissipation associated with the velocities of the full model state and the balanced and unbalanced modes.

The equation for KE obtained from the horizontal momentum equation is written as

$$\partial_t K = -\nabla \cdot \mathbf{u} K - \nabla \cdot \mathbf{u} p + w b'$$
$$- \mathbf{u}_h \cdot \lambda_u(\overline{\mathbf{u}_h}) + \mathbf{u}_h \cdot F_u, \qquad (21)$$

where $K = u_h^2/2$ denotes the kinetic energy, $\mathbf{u} = (u, v, w)$ is the full, $\mathbf{u}_h$ is the horizontal, and $\overline{\mathbf{u}_h}$ is the zonal mean velocity. The dissipation terms, which are the last two terms in the RHS of Eq. (21), extract KE from the flow. Large-scale dissipation, which acts on the large scales, is denoted by the term $D_L = \mathbf{u}_h \cdot \lambda_u(\overline{\mathbf{u}_h})$, where λ_u is the

linear drag coefficient that extracts energy from the mean flow and is related to the maximum growth rate as $\lambda_u = 0.75\sigma_{max}$. Small-scale dissipation on the other hand damps the smallest scales and is denoted by $D_S = \mathbf{u}_h \cdot F_u$, where $F_u = A_h \nabla^4 \mathbf{u}_h + A_v \partial^2 \mathbf{u}_h/\partial z^2$ indicates the dissipation due to biharmonic and vertical friction, respectively (for A_h and A_v, see Table 1).

The global mean values of KE dissipation are illustrated in Fig. 11 for all Ri for their full velocity component (FULL) component and modally decomposed components: BAL_NONLIN and UNB_NONLIN. The figure illustrates the large-scale (D_L) and small-scale (D_S) dissipation values, shown as a fraction of the total dissipation ($D_S + D_L$) for the respective mode; note that $D_S + D_L = 1$. In addition, the dissipation values for the modes from linear decomposition and contributions to D_S by biharmonic friction D_b and vertical friction D_v are tabulated in Table C1 of appendix C.

For small Ri, D_S of the full component is larger than the corresponding D_L, while D_S becomes smaller and negligible for higher Ri, as seen from the figure. On the other hand, D_L dominates D_S at larger Ri and becomes smaller, but not negligible, for small Ri. The impact of vertical friction D_v is in general much smaller than the biharmonic friction D_b in all cases (Table C1, appendix C). Further, we weigh the variations in the dissipation due to the decomposed modes for different Ri. As mentioned earlier, the unscaled dissipation rates related to the balanced and unbalanced modes will not add up to the dissipation of the full velocities [i.e., D_S(UNB_NONLIN) + D_S(BAL_NONLIN) $\neq D_S$(FULL); also for D_L] because

we disregard covariances between the decomposed velocities here. Only for the total energy of the linear decomposition, the covariances of the state vector components **x** cancel.

The dissipation related to the nonlinear balanced modes, BAL_NONLIN, has higher values for D_L than D_S for all Ri except for Ri = 3. For Ri = 327 and Ri = 917, most of the dissipation in BAL_NONLIN occurs at large scales while it is almost negligible for small scales. For Ri = 3, however, a significant amount of dissipation still occurs at the small scales. For the nonlinear unbalanced modes, UNB_NONLIN, D_S is much larger than D_L for all Ri, with a maximum of D_S for Ri = 3, while D_L for UNB_NONLIN tends to become almost negligible for small Ri. Specifically, for the nonlinear unbalanced component UNB_NONLIN, more than 60% of the dissipation occurs at small scales for all Ri; for Ri = 3 and 13 the dissipation at small scales increases to more than 90%. This directs to the inference that in the gravity wave mode dissipation occurs predominantly at small scales for all Ri.

The dissipation results for the decomposed modes elucidate that regimes with larger Ri dissipate energy preferably at large scales contrary to the small Ri regimes, which dissipate energy predominantly at small scales. This result is consistent with previous results, for example, Brüggemann and Eden (2015), which predict that in wavenumber space regimes with large and small Ri show a dominant KE flux toward large scales and small scales, respectively. In their study, the velocity field was decomposed into its rotational and divergent components, analogous to the balanced and unbalanced gravity modes in our study. Brüggemann and Eden (2015) showed that a downscale energy transfer is associated with a divergent flow field. However, a specific connection to a process, such as IGWs, was not made. We suggest based on our results that gravity waves could be a potential participant in the downscale energy transfer via the ageostrophic route.

6. Discussion

The primary aim of this study is to objectively identify gravity wave activity in different dynamical regimes characterized by different Ri. Although IGW emission is studied in idealized stratified and rotating flow, the quantification of the energy related to IGWs obtained by the nonlinear decomposition of Machenhauer (1977) presents a first estimate of the importance and variation of gravity wave activity in different dynamical regimes in the ocean.

The KE and APE related to the unbalanced component scales as $\sim Ro^2$ (or Ri^{-1}). This scaling is different from the scaling suggested by the similar idealized study

of Brüggemann and Eden (2015), who found by a subjective fit a scaling of $\sim Ro^{0.8}$ for the energy flux toward small-scale dissipation. The results resonate with the laboratory experiments of Williams et al. (2008), who suggest a scaling of $\sim Ro^2$, but differs from more complicated scalings suggested by other studies (e.g., Vanneste and Yavneh 2004). The disagreement of our scaling to the one obtained by Brüggemann and Eden (2015) is most likely related to the fact that Brüggemann and Eden (2015) do not explicitly assign the contribution of the unbalanced component to the small-scale dissipation, and thus their scaling represents the overall small-scale dissipation. The $\sim Ro^2$ scaling found in our simulations is related explicitly to the unbalanced component of the flow and therefore to the IGW field.

The IGWs present in the flow may break by Kelvin–Helmholtz (KH) instabilities and generate small-scale turbulence. The onset of KH instabilities after frontogenetic collapse in a baroclinic unstable flow has been studied, for example, by Skyllingstad and Samelson (2012) and Skyllingstad et al. (2017), who resolve KH instability by using high resolution and a nonhydrostatic model with an LES closure. We do not explicitly resolve KH instability in our hydrostatic simulations but we can resolve gravity wave emission, and we argue that we can also resolve parts of the nonlinear downscale energy transfer of the gravity waves. In our simulations the energy is then dissipated at the higher end of the resolved spectrum by the small-scale dissipation (i.e., the biharmonic friction in our model). This process is thought to mimic a further downscale energy transfer of the waves up to that point where they finally break by KH instabilities, but it is clear that in a future study this process should also be included.

To shed some light on the role of the large-scale dissipation, we have tested the sensitivity of the model results to the dissipation parameter. We set the velocity drag λ_u and buoyancy restoring λ_T to equal values, that is, $\lambda_u = \lambda_T$, and test for three different values: $0.75\sigma_{max}$, $1\sigma_{max}$, and $2\sigma_{max}$. In all these cases, the spectra of KE (not shown) related to the balanced and unbalanced components are qualitatively similar to the spectra shown in this paper (Figs. 7, 8, and 9) where $\lambda_u \neq \lambda_T$ (cf. section 2a). There is also not much change in the ratio between the unbalanced and balanced energy and the dissipation values for the linear and nonlinear modes. An important difference that arises with changing λ_u and λ_T is that the choice of λ_u and λ_T determines to some extent the mean flow, which in turn determines the Doppler shift of the shallow water gravity wave dispersion relation. This affects the Doppler-shifted envelope that determines the gravity wave branch where IGW activity is most prominent. The sensitivity of λ_u and λ_T also has been studied by Brüggemann and Eden (2014) for

various combinations of λ_u and λ_T for a model configuration similar to the present one. Their experiments indicate that λ_T has a minor influence but λ_u influences the eddy activity, which is strongest for a regime with larger Ri.

In addition to this, a simulation is performed where the (zonally averaged) velocity is restored toward the mean flow given by the initial conditions, which are in a geostrophic balance. This was done since the restoring toward vanishing flow might generate gravity waves by forcing the mean flow out of balance, but this effect appears to be very small. Brüggemann and Eden (2014) also find no impact of the restoring toward vanishing flow compared to the restoring toward mean flow. This suggests that irrespective of the restoring to a background mean flow or vanishing flow, the internal gravity waves are always spontaneously emitted by the balanced component of the flow during baroclinic instability and are at least not directly related to the restoring itself.

Further, we have attributed the IGW signals in the simulations as emissions from balanced flows. However, the unbalanced motions generated from balanced motions could form triads and can further generate unbalanced motions. This process might also contribute to IGW activity. Triad interactions between slow balanced and fast unbalanced motions have been studied, for example, by Bartello (1995), which suggest that in a slow–fast–fast triad the slow mode can catalyze the flow of energy from one fast mode to another. This slow–fast–fast interaction also sweeps the balanced energy from the slow mode downscale to the scales of dissipation (Bartello 1995). This supports the notion of a downscale energy transfer via an ageostrophic route suggested by, for example, Brüggemann and Eden (2015).

7. Summary and conclusions

In this study, we diagnose internal gravity waves (IGWs) emitted from an initially balanced flow using the nonlinear initialization technique of Machenhauer (1977). This is a novel approach to the oceanographic problem of identifying IGW signals spontaneously emitted from an initially balanced flow.

We use an idealized numerical setup that is baroclinically unstable, and the choice of the Richardson number Ri allows us to emulate different dynamical regimes ranging from ageostrophic [Ri = $O(1)$] to quasigeostrophic (Ri $\gg$ 1) flows. We first diagnose IGWs in frequency and wavenumber space and then using linear and nonlinear modal decomposition. The modal decomposition yields a balanced mode and two unbalanced gravity modes, which we discuss in frequency–wavenumber space. Based on the energy distribution

between the unbalanced and balanced flow components, quantitative estimates are presented and a power law is suggested, which relates spontaneously emitted IGWs to Richardson (Ri) and Rossby (Ro) numbers. Further, an assessment of the small-scale and large-scale dissipation associated with the balanced and unbalanced modes sheds light on the downscale transfer of energy. The key results are as follows:

1) The nonlinear initialization technique of Machenhauer (1977) is efficient in decomposing the balanced mode and unbalanced IGW modes. Although this decomposition tends to get difficult for Ri = $O(1)$ or less, the nonlinear decomposition is promising and an improvement over the linear decomposition in the detection of IGWs.

2) An ageostrophic regime (small Ri) shows much more IGW activity than a quasigeostrophic regime (large Ri). Therefore, spontaneous emission of IGWs from the balanced flow occurs for all Ri but becomes increasingly weaker with increasing Ri (or decreasing Ro). Quantitative estimates based on our idealized simulations suggest that the decrease of the kinetic energy tied to the unbalanced component scales close to $Ri^{-1.05}$ (or $Ro^{2.1}$), whereas for available potential energy this scaling turns out to be $Ri^{-1.3}$ (or $Ro^{2.6}$).

3) IGWs dissipate predominantly through small-scale dissipation for all Ri regime, thus acting as a downscale route to dissipation for the balanced flow. The result emphasizes the role IGWs play in the downscale energy transfer leading to small-scale dissipation in the ocean.

It should be noted that the nonlinear decomposition might be less efficient for Ri < $O(1)$ since the interactions between the balanced and unbalanced motions are much stronger and the time-scale separation between these motions is minimal, which renders it hard to separate one mode from the other. For a regime Ri < $O(1)$, other kinds of instabilities such as symmetric instability and Kelvin–Helmhotlz instability become more prominent, such that the detection of IGWs becomes even more difficult. Nonetheless, for regimes that allow the separation of fast and slow motions, the procedure seems promising and can be applied to isolate gravity wave modes from balanced modes for future studies.

This study brings evidence to the role of IGWs in the downscale energy transfer in the ocean based on simulations of baroclinic instability that mimic dynamical regimes of the ocean in an idealized setting. This is important for the energy dissipation of the ocean as well as the dissipation of balanced mesoscale field in the ocean. At the same time, this process also sheds light on the generation of IGWs in the ocean, which is necessary for the IGW energy budget. An important result is the

power law of $\sim$Ro^2 for IGW emission in context of our idealized simulations, which can potentially be utilized to parameterize the spontaneous emission of IGWs.

APPENDIX A

Eady Mode Projection on Vertical Modes

The wave solution for the Eady case can be expressed as

$$\phi(z) = A \cosh(z/d) + B \sinh(z/d), \qquad (A1)$$

where $d = f/Nk_h^2$ (where $k_h^2 = k^2 + l^2$) and $A = BH(U_0 - c)/(U_0 d)$ and $B = 1$ are constants evaluated from the initial condition [for details of the Eady solution, see, e.g., Olbers et al. (2012)]. The phase speed c is given as

$$c = \frac{U_0}{2} \pm \frac{U_0 d}{H} \sqrt{1 + \frac{H^2}{4d^2} - \frac{H}{d} \coth\left(\frac{H}{d}\right)}. \qquad (A2)$$

Note that c becomes imaginary when the term under the square root becomes negative.

The set of vertical eigenfunction Φ_n is expressed as follows:

$$\Phi_n(z) = \cos\left(\frac{n\pi}{H}\right)z = \cos mz, \quad \text{and} \qquad (A3)$$

$$m = \frac{n\pi}{H}, \quad n = 0, 1, 2, 3, \ldots \qquad (A4)$$

The projection of the Eady mode on the vertical modes then can be written as

$$\phi(z) = \sum_{n=1}^{\infty} A_n \Phi_n(z), \qquad (A5)$$

and the coefficient A_n can be estimated from

$$\int_{-H}^{0} A_n \Phi_n^2(z)\, dz = \int_{-H}^{0} \phi(z)\Phi_n(z)\, dz, \qquad (A6)$$

$$\int_{-H}^{0} A_n \cos^2(mz)\, dz = \int_{-H}^{0} [\Re(A)\cosh(z/d) + \sinh(z/d)]\cos(mz)\, dz, \quad \text{and} \qquad (A7)$$

$$= \frac{d}{1 + m^2 d^2}\left\{\Re(A)\left[\sinh\left(\frac{H}{d}\right)(-1)^n\right] - \cosh\left(\frac{H}{d}\right)(-1)^n + 1\right\}; \qquad (A8)$$

for $n = 0$ (barotropic mode),

$$A_0 = \frac{d}{H(1 + m^2 d^2)}[\Re(A)\sinh(H/d) - \cosh(H/d) + 1], \qquad (A9)$$

and for $n = 1, 2, 3, \ldots$ (baroclinic modes)

$$A_n = \frac{2d}{H(1 + m^2 d^2)}\{(-1)^n[\Re(A)\sinh(H/d) - \cosh(H/d)] + 1\}. \qquad (A10)$$

APPENDIX B

Modal Decomposition

a. Eigenvectors

The three right (column) eigenvectors $\mathbf{q}^{\pm,0}$ to the matrix $\mathbf{L}$ with $\mathbf{L} \cdot \mathbf{q}^{\pm,0} = \omega^{\pm,0}\mathbf{q}^{\pm,0}$ are given by

$$\mathbf{q}^{\pm} = \begin{bmatrix} (ifl - k\omega)/(\omega^2 - f^2) \\ (-ifk - l\omega)/(\omega^2 - f^2) \\ 1 \end{bmatrix}, \quad \mathbf{q}^0 = \begin{pmatrix} -il/f \\ ik/f \\ 1 \end{pmatrix}, \quad (B1)$$

and the three left (row) eigenvectors $\mathbf{p}^{\pm,0}$ to the matrix $\mathbf{L}$ with $\mathbf{p}^{\pm,0} \cdot \mathbf{L} = \omega^{\pm,0}\mathbf{p}^{\pm,0}$ are given by

$$\mathbf{p}^{\pm} = \frac{\omega^2 - f^2}{2\omega^2}\left[\frac{(-ifl - k\omega)c_m^2}{\omega^2 - f^2}, \frac{(ikf - l\omega)c_m^2}{\omega^2 - f^2}, 1\right],$$

$$\mathbf{p}^0 = \frac{f^2}{(\omega^{\pm})^2}(ilc_m^2/f, -ikc_m^2/f, 1). \qquad (B2)$$

Note that it holds that

$$\mathbf{p}^{\pm} \cdot \mathbf{q}^{\pm} = \mathbf{p}^0 \cdot \mathbf{q}^0 = 1,$$

$$\mathbf{p}^{\pm} \cdot \mathbf{q}^{\mp} = \mathbf{p}^{\mp} \cdot \mathbf{q}^{\pm} = \mathbf{p}^0 \cdot \mathbf{q}^{\pm} = \mathbf{p}^{\pm} \cdot \mathbf{q}^0 = 0. \qquad (B3)$$

(For $\mathbf{q}^{\pm}$ and $\mathbf{p}^{\pm}$, read ω as $\omega^{\pm}$.)

b. Projection matrices

Mathematically, $\mathbf{B}$ and $\mathbf{G}$ are the projection matrices $\mathbf{B} = \mathbf{q}^0 \cdot \mathbf{p}^0$ and $\mathbf{G}^{\pm} = \mathbf{q}^{\pm} \cdot \mathbf{p}^{\pm}$:

TABLE C1. Globally integrated values for KE dissipation shown as a fraction of the total dissipation. The total dissipation is a sum of large-scale ($D_L = \mathbf{u}_h \cdot \lambda_u \overline{\mathbf{u}_h}$) and small-scale ($D_S = \mathbf{u}_h \cdot F_u$) dissipation, and D_S comprises the contributions from biharmonic friction D_b and vertical friction D_v (i.e., $D_b + D_v = D_S$). The dissipation values are shown for different regimes indicated by Ri for their full velocity component (FULL) and modally decomposed components: BAL_LIN, BAL_NONLIN, UNB_LIN, and UNB_NONLIN.

Ri →	3				13				327				917			
Component ↓	D_b	D_v	D_S	D_L	D_b	D_v	D_S	D_L	D_b	D_v	D_S	D_L	D_b	D_v	D_S	D_L
FULL	0.80	0.003	0.81	0.18	0.41	0.007	0.41	0.58	0.11	0.004	0.11	0.88	0.03	0.001	0.04	0.95
UNB_LIN	0.85	0.008	0.86	0.13	0.83	0.03	0.87	0.12	0.60	0.08	0.69	0.30	0.41	0.10	0.51	0.48
UNB_NONLIN	0.91	0.02	0.94	0.05	0.87	0.04	0.91	0.08	0.62	0.14	0.77	0.22	0.34	0.29	0.63	0.36
BAL_LIN	0.50	0.002	0.50	0.49	0.22	0.003	0.22	0.77	0.06	0.002	0.06	0.93	0.03	0.002	0.03	0.96
BAL_NONLIN	0.71	0.01	0.73	0.26	0.34	0.01	0.36	0.63	0.08	0.009	0.08	0.91	0.03	0.008	0.04	0.95

$$\mathbf{B} = \mathbf{q}^0 \cdot \mathbf{p}^0 = \frac{c_m^2}{(\omega^\pm)^2} \begin{pmatrix} l^2 & -kl & -ilf/c_m^2 \\ -kl & k^2 & ikf/c_m^2 \\ ilf & -ikf & f^2/c_m^2 \end{pmatrix}, \quad \text{and} \tag{B4}$$

$$\mathbf{G}^\pm = \mathbf{q}^\pm \cdot \mathbf{p}^\pm = \frac{1}{2(\omega^\pm)^2} \left\{ \begin{array}{ccc} \left[f^2l^2 + k^2(\omega^\pm)^2\right]\big/ k_h^2 & klc_m^2 - if\omega^\pm & ifl - k\omega^\pm \\ klc_m^2 + if\omega^\pm & \left[f^2k^2 + l^2(\omega^\pm)^2\right]\big/k_h^2 & -ifk - l\omega^\pm \\ (-ifl - k\omega^\pm)c_m^2 & (ifk - l\omega^\pm)c_m^2 & k_h^2 c_m^2 \end{array} \right\}, \tag{B5}$$

with the properties

$$\mathbf{L} = \omega^0 \mathbf{B} + \sum_{d=\pm} \omega^d \mathbf{G}^d = \begin{pmatrix} 0 & -if & -k \\ if & 0 & -l \\ -kc_m^2 & -lc_m^2 & 0 \end{pmatrix} \tag{B6}$$

and

$$\mathbf{G}^\pm \cdot \mathbf{G}^\pm = \mathbf{G}^\pm, \quad \mathbf{G}^\pm \cdot \mathbf{G}^\mp = 0$$

$$\rightarrow \mathbf{L}^n = \sum_{s=\pm} (\omega^s)^n \mathbf{G}^s$$

$$\rightarrow f(\mathbf{L}) = \sum_{s=\pm} f(\omega^s) \mathbf{G}^s \tag{B7}$$

for a general function f, but note that $\mathbf{L}$ is singular since $\omega^0 = 0$.

c. Operator $i(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}$

The matrix $i(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G}$ introduced in section 4b is given by

$$i(\mathbf{L} \cdot \mathbf{G})^{-1} \cdot \mathbf{G} = \frac{1}{(\omega^\pm)^2} \begin{pmatrix} 0 & f & -ik \\ -f & 0 & -il \\ -ikc_m^2 & -ilc_m^2 & 0 \end{pmatrix}. \tag{B8}$$

APPENDIX C

Dissipation

Globally integrated values for KE dissipation at large and small scales are shown as a fraction of the total dissipation in Table C1. The dissipation values are shown for different regimes indicated by Ri for their FULL and modally decomposed components obtained from the linear (BAL_LIN, UNB_LIN) and nonlinear (BAL_NONLIN, UNB_NONLIN) decompositions.

REFERENCES

Baer, F., and J. J. Tribbia, 1977: On complete filtering of gravity modes through nonlinear initialization. *Mon. Wea. Rev.*, **105**, 1536–1539, https://doi.org/10.1175/1520-0493(1977)105<1536:OCFOGM>2.0.CO;2.

Barkan, R., K. B. Winters, and J. C. McWilliams, 2017: Stimulated imbalance and the enhancement of eddy kinetic energy dissipation by internal waves. *J. Phys. Oceanogr.*, **47**, 181–198, https://doi.org/10.1175/JPO-D-16-0117.1.

Bartello, P., 1995: Geostrophic adjustment and inverse cascades in rotating stratified turbulence. *J. Atmos. Sci.*, **52**, 4410–4428, https://doi.org/10.1175/1520-0469(1995)052<4410:GAAICI>2.0.CO;2.

Blumen, W., 1972: Geostrophic adjustment. *Rev. Geophys.*, **10**, 485–528, https://doi.org/10.1029/RG010i002p00485.

Borchert, S., U. Achatz, and M. D. Fruman, 2014: Gravity wave emission in an atmosphere-like configuration of the differentially heated rotating annulus experiment. *J. Fluid Mech.*, **758**, 287–311, https://doi.org/10.1017/jfm.2014.528.

Brüggemann, N., and C. Eden, 2014: Evaluating different parameterizations for mixed layer eddy fluxes induced by baroclinic instability. *J. Phys. Oceanogr.*, **44**, 2524–2546, https://doi.org/10.1175/JPO-D-13-0235.1.

——, and ——, 2015: Routes to dissipation under different dynamical conditions. *J. Phys. Oceanogr.*, **45**, 2149–2168, https://doi.org/10.1175/JPO-D-14-0205.1.

Bühler, O., 2014: *Waves and Mean Flows.* Cambridge University Press, 370 pp.

Capet, X., J. C. McWilliams, M. J. Molemaker, and A. Shchepetkin, 2008a: Mesoscale to submesoscale transition in the California Current System. Part I: Flow structure, eddy flux, and observational tests. *J. Phys. Oceanogr.*, **38**, 29–43, https://doi.org/10.1175/2007JPO3671.1.

——, ——, ——, and ——, 2008b: Mesoscale to submesoscale transition in the California Current System. Part III: Energy balance and flux. *J. Phys. Oceanogr.*, **38**, 2256–2269, https://doi.org/10.1175/2008JPO3810.1.

Charney, J. G., 1971: Geostrophic turbulence. *J. Atmos. Sci.*, **28**, 1087–1095, https://doi.org/10.1175/1520-0469(1971)028<1087:GT>2.0.CO;2.

Danioux, E., J. Vanneste, P. Klein, and H. Sasaki, 2012: Spontaneous inertia-gravity-wave generation by surface-intensified turbulence. *J. Fluid Mech.*, **699**, 153–173, https://doi.org/10.1017/jfm.2012.90.

Dewar, W. K., and A. M. Hogg, 2010: Topographic inviscid dissipation of balanced flow. *Ocean Modell.*, **32**, 1–13, https://doi.org/10.1016/j.ocemod.2009.03.007.

Dritschel, D. G., and A. Viúdez, 2003: A balanced approach to modelling rotating stably stratified geophysical flows. *J. Fluid Mech.*, **488**, 123–150, https://doi.org/10.1017/S0022112003004920.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52, https://doi.org/10.3402/tellusa.v1i3.8507.

Eden, C., 2016: Closing the energy cycle in an ocean model. *Ocean Modell.*, **101**, 30–42, https://doi.org/10.1016/j.ocemod.2016.02.005.

——, and D. Olbers, 2017: A closure for internal wave–mean flow interaction. Part II: Wave drag. *J. Phys. Oceanogr.*, **47**, 1403–1412, https://doi.org/10.1175/JPO-D-16-0056.1.

Ford, R., M. E. McIntyre, and W. A. Norton, 2000: Balance and the slow quasimanifold: Some explicit results. *J. Atmos. Sci.*, **57**, 1236–1254, https://doi.org/10.1175/1520-0469(2000)057<1236:BATSQS>2.0.CO;2.

Gertz, A., and D. N. Straub, 2009: Near-inertial oscillations and the damping of midlatitude gyres: A modeling study. *J. Phys. Oceanogr.*, **39**, 2338–2350, https://doi.org/10.1175/2009JPO4058.1.

Hien, S., J. Rolland, S. Borchert, L. Schoon, C. Zülicke, and U. Achatz, 2018: Spontaneous inertia–gravity wave emission in the differentially heated rotating annulus experiment. *J. Fluid Mech.*, **838**, 5–41, https://doi.org/10.1017/jfm.2017.883.

Kafiabad, H. A., and P. Bartello, 2016: Balance dynamics in rotating stratified turbulence. *J. Fluid Mech.*, **795**, 914–949, https://doi.org/10.1017/jfm.2016.164.

——, and ——, 2017: Rotating stratified turbulence and the slow manifold. *Comput. Fluids*, **151**, 23–34, https://doi.org/10.1016/j.compfluid.2016.10.020.

Leith, C., 1980: Nonlinear normal mode initialization and quasigeostrophic theory. *J. Atmos. Sci.*, **37**, 958–968, https://doi.org/10.1175/1520-0469(1980)037<0958:NNMIAQ>2.0.CO;2.

Lorenz, E. N., 1980: Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.*, **37**, 1685–1699, https://doi.org/10.1175/1520-0469(1980)037<1685:ASAQGE>2.0.CO;2.

——, 1986: On the existence of a slow manifold. *J. Atmos. Sci.*, **43**, 1547–1558, https://doi.org/10.1175/1520-0469(1986)043<1547:OTEOAS>2.0.CO;2.

——, 1992: The slow manifold—What is it? *J. Atmos. Sci.*, **49**, 2449–2451, https://doi.org/10.1175/1520-0469(1992)049<2449:TSMII>2.0.CO;2.

——, and V. Krishnamurthy, 1987: On the nonexistence of a slow manifold. *J. Atmos. Sci.*, **44**, 2940–2950, https://doi.org/10.1175/1520-0469(1987)044<2940:OTNOAS>2.0.CO;2.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model, with applications to normal mode initialization. *Beitr. Phys. Atmos.*, **50**, 253–271.

McIntyre, M. E., and W. A. Norton, 2000: Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.*, **57**, 1214–1235, https://doi.org/10.1175/1520-0469(2000)057<1214:PVIOAH>2.0.CO;2.

Molemaker, M. J., J. C. McWilliams, and I. Yavneh, 2005: Baroclinic instability and loss of balance. *J. Phys. Oceanogr.*, **35**, 1505–1517, https://doi.org/10.1175/JPO2770.1.

——, ——, and X. Capet, 2010: Balanced and unbalanced routes to dissipation in an equilibrated Eady flow. *J. Fluid Mech.*, **654**, 35–63, https://doi.org/10.1017/S0022112009993272.

Nagai, T., A. Tandon, E. Kunze, and A. Mahadevan, 2015: Spontaneous generation of near-inertial waves by the Kuroshio Front. *J. Phys. Oceanogr.*, **45**, 2381–2406, https://doi.org/10.1175/JPO-D-14-0086.1.

Olbers, D., J. Willebrand, and C. Eden, 2012: *Ocean Dynamics.* Springer, 703 pp.

Plougonven, R., and C. Snyder, 2007: Inertia–gravity waves spontaneously generated by jets and fronts. Part I: Different baroclinic life cycles. *J. Atmos. Sci.*, **64**, 2502–2520, https://doi.org/10.1175/JAS3953.1.

——, and F. Zhang, 2014: Internal gravity waves from atmospheric jets and fronts. *Rev. Geophys.*, **52**, 33–76, https://doi.org/10.1002/2012RG000419.

——, D. J. Muraki, and C. Snyder, 2005: A baroclinic instability that couples balanced motions and gravity waves. *J. Atmos. Sci.*, **62**, 1545–1559, https://doi.org/10.1175/JAS3426.1.

——, A. Hertzog, and M. J. Alexander, 2015: Case studies of nonorographic gravity waves over the Southern Ocean emphasize the role of moisture. *J. Geophys. Res. Oceans*, **120**, 1278–1299, https://doi.org/10.1002/2014JD022332.

Pollmann, F., C. Eden, and D. Olbers, 2017: Evaluating the global internal wave model IDEMIX using finestructure methods. *J. Phys. Oceanogr.*, **47**, 2267–2289, https://doi.org/10.1175/JPO-D-16-0204.1.

Rossby, C.-G., 1938: On the mutual adjustment of pressure and velocity distributions in certain simple current systems, II. *J. Mar. Res.*, **1**, 239–263, https://doi.org/10.1357/002224038806440520.

Shutts, G., and S. Vosper, 2011: Stratospheric gravity waves revealed in NWP model forecasts. *Quart. J. Roy. Meteor. Soc.*, **137**, 303–317, https://doi.org/10.1002/qj.763.

Skyllingstad, E. D., and R. Samelson, 2012: Baroclinic frontal instabilities and turbulent mixing in the surface boundary layer. Part I: Unforced simulations. *J. Phys. Oceanogr.*, **42**, 1701–1716, https://doi.org/10.1175/JPO-D-10-05016.1.

——, J. Duncombe, and R. M. Samelson, 2017: Baroclinic frontal instabilities and turbulent mixing in the surface boundary layer. Part II: Forced simulations. *J. Phys. Oceanogr.*, **47**, 2429–2454, https://doi.org/10.1175/JPO-D-16-0179.1.

Stone, P. H., 1966: On non-geostrophic baroclinic stability. *J. Atmos. Sci.*, **23**, 390–400, https://doi.org/10.1175/1520-0469(1966)023<0390:ONGBS>2.0.CO;2.

Sugimoto, N., and R. Plougonven, 2016: Generation and backreaction of spontaneously emitted inertia-gravity waves. *Geophys. Res. Lett.*, **43**, 3519–3525, https://doi.org/10.1002/2016GL068219.

Theiss, J., and A. R. Mohebalhojeh, 2009: The equatorial counterpart of the quasi-geostrophic model. *J. Fluid Mech.*, **637**, 327–356, https://doi.org/10.1017/S0022112009008052.

Thomas, L. N., A. Tandon, and A. Mahadevan, 2008: Submesoscale processes and dynamics. *Ocean Modeling in an Eddying Regime*, *Geophys. Monogr.*, Vol. 177, Amer. Geophys. Union, 17–38.

Vanneste, J., 2013: Balance and spontaneous wave generation in geophysical flows. *Annu. Rev. Fluid Mech.*, **45**, 147–172, https://doi.org/10.1146/annurev-fluid-011212-140730.

——, and I. Yavneh, 2004: Exponentially small inertia–gravity waves and the breakdown of quasigeostrophic balance. *J. Atmos. Sci.*, **61**, 211–223, https://doi.org/10.1175/1520-0469(2004)061<0211:ESIWAT>2.0.CO;2.

Viúdez, Á., 2008: The stationary frontal wave packet spontaneously generated in mesoscale dipoles. *J. Phys. Oceanogr.*, **38**, 243–256, https://doi.org/10.1175/2006JPO3692.1.

Warn, T., 1997: Nonlinear balance and quasi-geostrophic sets. *Atmos.–Ocean*, **35**, 135–145, https://doi.org/10.1080/07055900.1997.9649588.

Williams, P. D., T. W. Haine, and P. L. Read, 2008: Inertia–gravity waves emitted from balanced flow: Observations, properties, and consequences. *J. Atmos. Sci.*, **65**, 3543–3556, https://doi.org/10.1175/2008JAS2480.1.

Wu, D. L., and S. D. Eckermann, 2008: Global gravity wave variances from *Aura* MLS: Characteristics and interpretation. *J. Atmos. Sci.*, **65**, 3695–3718, https://doi.org/10.1175/2008JAS2489.1.

Xie, J.-H., and J. Vanneste, 2015: A generalised-Lagrangian-mean model of the interactions between near-inertial waves and mean flow. *J. Fluid Mech.*, **774**, 143–169, https://doi.org/10.1017/jfm.2015.251.

Yasuda, Y., K. Sato, and N. Sugimoto, 2015a: A theoretical study on the spontaneous radiation of inertia–gravity waves using the renormalization group method. Part I: Derivation of the renormalization group equations. *J. Atmos. Sci.*, **72**, 957–983, https://doi.org/10.1175/JAS-D-13-0370.1.

——, ——, and ——, 2015b: A theoretical study on the spontaneous radiation of inertia–gravity waves using the renormalization group method. Part II: Verification of the theoretical equations by numerical simulation. *J. Atmos. Sci.*, **72**, 984–1009, https://doi.org/10.1175/JAS-D-13-0371.1.

Zeitlin, V., 2008: Decoupling of balanced and unbalanced motions and inertia–gravity wave emission: Small versus large Rossby numbers. *J. Atmos. Sci.*, **65**, 3528–3542, https://doi.org/10.1175/2008JAS2481.1.