The interior energy pathway: inertial gravity wave emission by oceanic flows

Jin-Song von Storch, Gualtiero Badin and Marcel Oliver

Abstract We review the possible role of spontaneous emission and subsequent capture of internal gravity waves (IGWs) for dissipation in oceanic flows under conditions characteristic for the ocean circulation. Dissipation is necessary for the transfer of energy from the essentially balanced large-scale ocean circulation and mesoscale eddy fields down to smaller scales where instabilities and subsequent small scale turbulence complete the route to dissipation.

Spontaneous wave emission by flows is a viable route to dissipation. For quasi-balanced flows, characterized by a small Rossby number, the amplitudes of emitted waves are expected to be small. However, once being emitted into a three-dimensional eddying flow field, waves can undergo refraction and may be "captured." During wave capture, the wavenumber grows exponentially, ultimately leading to breakup and dissipation. For flows with not too small Rossby number, e.g. for flows in the vicinity of strong fronts, dissipation occurs in a more complex manner. It can occur via spontaneous wave emission and subsequent wave capture, with the amplitudes of waves emitted in frontal systems being expected to be larger than amplitudes of waves emitted by quasi-balanced flows. It can also occur through turbulence and filamentation emerging from frontogenesis.

So far, quantitative importance of this energy pathway—crucial for determining correct eddy viscosities in general circulation models—is not known. Toward an answer to this question, we discuss IGWs diagnostics, review spontaneous emission of both quasi-balanced and less-balanced frontal flows, and discuss recent numerical results based on a high-resolution ocean general circulation model.

Jin-Song von Storch
Max-Planck-Institut für Meteorologie, Hamburg, Germany,
e-mail: jin-song.von.storch@mpimet.mpg.de

Gualtiero Badin
Institute of Oceanography and Center for Earth System Research and Sustainability, Universität Hamburg, Germany, e-mail: gualtiero.badin@uni-hamburg.de

Marcel Oliver
Jacobs University, Bremen, Germany, e-mail: oliver@member.ams.org

1 Introduction

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

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

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

The process of spontaneous imbalance has been researched extensively following the pioneering work of Lighthill (1952) and has received interest in the past decade. For details, we refer the reader to the excellent reviews by Vanneste (2013) for the

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

theory and by Plougonven and Zhang (2014) for observations in the atmosphere. A broad overview can also be gained from the special issue on 'Spontaneous Imbalance' of the *Journal of the Atmospheric Sciences* from May 2009.

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

The Chapter is organized as follows. In Section 2, we review the process of spontaneous imbalance. As discussed, the emitted waves have only small amplitudes, but we speculate they might still play a non-negligible role for energy dissipation through their steepening following their capture by the mean flow. In Section 3, we thus review the mechanism of wave capture which is crucial if wave emission by quasi-balanced flows with small Rossby number is to play a noticeable role for the total dissipation of energy. In Section 4, we review the interaction between the spontaneously emitted IGWs and fronts, not necessarily in the small Rossby number regime, focusing especially on the role of waves in frontal geostrophic adjustment. In Section 5, we switch to the diagnostics for the detection of the waves, with particular emphasis especially given to the method of "optimal potential vorticity balance." In Section 6, we discuss the role of high-resolution general circulation model for investigating spontaneous wave emission. We conclude with a summary and discussion in Section 7.

2 Rotating shallow water equations and spontaneous emission

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

2.1 Shallow water on the f-plane

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

$$\partial_t(h\boldsymbol{u}) + \boldsymbol{\nabla}\cdot(h\boldsymbol{u}\otimes\boldsymbol{u}) + f\,(h\boldsymbol{u})^{\perp} = -\tfrac{1}{2}\,g\,\boldsymbol{\nabla}h^2\,, \tag{1a}$$

$$\partial_t h + \boldsymbol{\nabla}\cdot(h\boldsymbol{u}) = 0\,, \tag{1b}$$

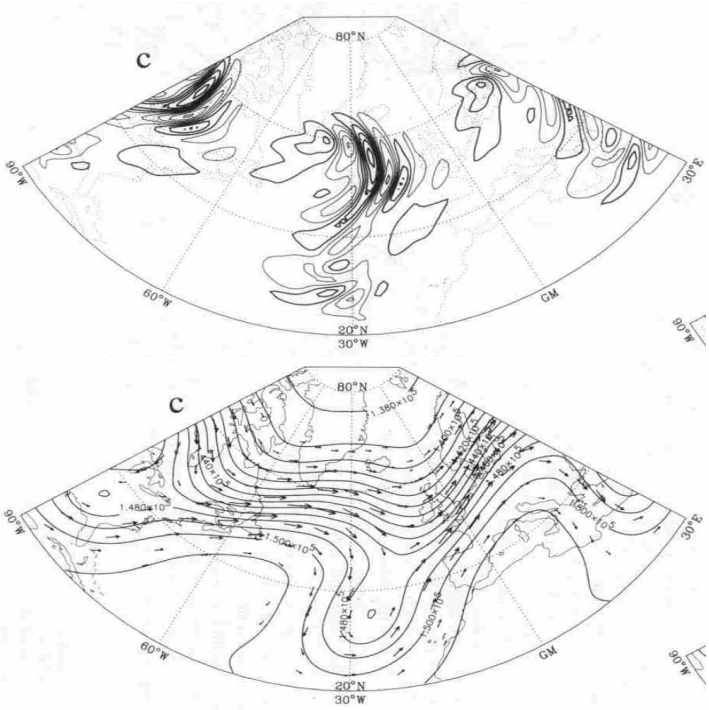

Fig. 1 Horizontal velocity divergence (top) and geopotential (bottom) at 130 mb on day 11 of the T126 model simulation reported by O'Sullivan and Dunkerton (1995). Vectors show the horizontal wind. The presence of IGWs is apparent in the velocity divergence field. The quasi-balanced background flow emitting the waves is described by the geopotential contours and wind vectors in the bottom figure. From O'Sullivan and Dunkerton (1995).

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

$$H = \frac{1}{2} \int (h\,|\boldsymbol{u}|^2 + g\,h^2)\,\mathrm{d}\boldsymbol{x}\,, \tag{2}$$

in time and the *potential vorticity*

$$q = \frac{\zeta + f}{h}\,, \tag{3}$$

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

$$Z = \frac{1}{2} \int h\, q^2 \, \mathrm{d}\boldsymbol{x} \,. \tag{4}$$

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

2.2 Spontaneous emission

Taking divergence and curl of the shallow water momentum equation (1a), we obtain, respectively,

$$\partial_t \boldsymbol{\nabla} \cdot (h\boldsymbol{u}) + \boldsymbol{\nabla}\boldsymbol{\nabla} : (h\boldsymbol{u} \otimes \boldsymbol{u}) - f\, \boldsymbol{\nabla}^\perp \cdot (h\boldsymbol{u}) = -\tfrac{1}{2}\, g\, \Delta h^2 \,, \tag{5a}$$

$$\partial_t \boldsymbol{\nabla}^\perp \cdot (h\boldsymbol{u}) + \boldsymbol{\nabla}^\perp\boldsymbol{\nabla} : (h\boldsymbol{u} \otimes \boldsymbol{u}) + f\, \boldsymbol{\nabla} \cdot (h\boldsymbol{u}) = 0 \,, \tag{5b}$$

where we use Δ to denote the Laplacian and $\boldsymbol{\nabla}\boldsymbol{\nabla}$ to denote the Hessian operator. Further, "$\otimes$" denotes the tensor product of two vectors and the colon denotes contraction over both indices of two 2-tensors. Taking the time derivative of (5a), eliminating the Coriolis term using (5b), and replacing $\boldsymbol{\nabla} \cdot (h\boldsymbol{u})$ by $-\partial_t h$ using the continuity equation (1b), we find

$$-\partial_t^3 h + \partial_t \boldsymbol{\nabla}^2 : (h\boldsymbol{u} \otimes \boldsymbol{u}) + f\, \boldsymbol{\nabla}^\perp\boldsymbol{\nabla} : (h\boldsymbol{u} \otimes \boldsymbol{u}) - f^2\, \partial_t h = -\tfrac{1}{2}\, g\, \partial_t \Delta h^2 \,. \tag{6}$$

Consider a vortical flow described by (6) confined to a bounded region of the spatial domain. We denote the constant layer depth far from the center of the vortical motions by h_0 and assume that h_0 is constant. Then the right hand side of (6) can be rewritten as

$$-\tfrac{1}{2}\, g\, \partial_t \Delta h^2 \;= -\tfrac{1}{2}\, g\, \partial_t \Delta (h - h_0)^2 - g\, h_0\, \partial_t \Delta h \,. \tag{7}$$

Writing $c_0 = (g h_0)^{1/2}$ to denote the gravity wave phase speed far from the center of vortical motions and rearranging terms, we obtain

$$\left(\frac{\partial^2}{\partial t^2} + f^2 - c_0^2\, \Delta \right) \partial_t h = \boldsymbol{\nabla}\boldsymbol{\nabla} : \mathsf{T} \tag{8}$$

with

$$\mathsf{T} = \partial_t (h\boldsymbol{u} \otimes \boldsymbol{u}) - f\, h\, \frac{\boldsymbol{u}^\perp \otimes \boldsymbol{u} + \boldsymbol{u} \otimes \boldsymbol{u}^\perp}{2} + \mathsf{I}\, \frac{g}{2}\, \partial_t (h - h_0)^2 \,, \tag{9}$$

where I denotes the 2×2-identity matrix.

Equation (8) can be seen as the shallow-water equivalent of the Lighthill equation (Ford, 1994; Ford et al., 2000). The left hand side of (8) is a linear inertia-gravity wave operator. When the Froude number $\mathrm{Fr} = U/c_0$, with U denoting the typical velocity in the region of the vortical flow, is small, the right hand side of (8) depends, to good approximation, on the vortical flow alone and can hence be considered as a given source of inertia-gravity waves. In other words, the source term is known as soon as the vortical flow is known. Ford et al. (2000) study the case when potential

vorticity anomalies are confined to a finite-size region surrounded by a region in which only linear waves propagate. Using matched asymptotic expansions with the Froude number as small parameter, they link an unsteady vortex pattern in the inner region to emitted propagating waves in the outer region, thereby giving an explicit description of the mechanism of spontaneous emission.

2.3 Beyond shallow water

Going beyond prototype studies, there is now an extensive body of work using high-resolution numerical simulations on more general flows including surface-intensified fronts (Gall et al., 1988; Danioux et al., 2012), wave emission by baroclinically unstable flows (O'Sullivan and Dunkerton, 1995; Zhang, 2004; Plougonven and Snyder, 2005, 2007; Viúdez, 2006; Viúdez and Dritschel, 2006; Nadiga, 2014), potential vorticity anomalies (Lott et al., 2010, 2012), and dipolar flows (Viúdez, 2007; Snyder et al., 2007, 2009; Wang et al., 2009; Wang and Zhang, 2010). Of particular interest for the present discussion is the study of O'Sullivan and Dunkerton (1995), illustrated in Fig. 1. The atmospheric flow considered there is comparable to the situation in the Southern Ocean, even though the oceanic baroclinic eddies have much smaller spatial scales than those in the atmosphere. Both the distorted jet considered by O'Sullivan and Dunkerton and the eddying circumpolar currents in the Southern Ocean represent essentially balanced flows.

Most of the studies cited above focus on spontaneous imbalance as a wave generation mechanism and do not directly consider the fate of waves after being emitted. According to Lighthill (1952) and Ford (1994), the waves resulting from spontaneous emission have wavelengths much larger than the spatial scales of the wave source. The waves identified in the references mentioned (e.g. in (O'Sullivan and Dunkerton, 1995)) have, on the contrary, extremely short wavelengths (i.e., wavelengths close to the grid size of the model). Thus, the working hypothesis is that these waves do not represent waves just emitted by the flows, but waves, after having been emitted, are refracted by the flows in which they are embedded, a possibility pointed out by Plougonven and Snyder (2005). This is further discussed in Section 3 below.

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

In this review, we focus on issues pertaining to the generic situation of extratropical flows and refer the reader for the specifics of equatorial dynamics to Le Sommer et al. (2004) and Medvedev and Zeitlin (2005). Regarding spontaneous generation of near-internal gravity waves in the Kuroshio front, Nagai et al. (2015) found that the energy emitted by the waves is larger than the energy produced by lee waves

created by the interaction of balanced flows with the bottom topography (Nikurashin and Ferrari, 2011). Most of the energy radiated from the frontal region, however, is being reabsorbed, suggesting that wave capture plays an important role in frontogenesis. The situation in frontal regions will be addressed in Section 4 further below.

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

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

3 Ray equations and wave capture

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

$$\frac{\mathrm{d}_{\mathrm{g}}\boldsymbol{x}}{\mathrm{d}t} \equiv \boldsymbol{c}_{\mathrm{g}} = \boldsymbol{c}_{\mathrm{g,i}} + \boldsymbol{u}\,, \tag{10a}$$

$$\frac{\mathrm{d}_{\mathrm{g}}\boldsymbol{k}}{\mathrm{d}t} = -(\boldsymbol{\nabla u})\cdot\boldsymbol{k}\,, \tag{10b}$$

where $\boldsymbol{c}_{\mathrm{g}}$ is the effective group velocity, $\mathrm{d}_g/\mathrm{d}t$ denotes the time derivative in a frame traveling with the wave packet, $\boldsymbol{\nabla u}$ is the gradient tensor of the flow $\boldsymbol{u}$.

To obtain equation (10b), the intrinsic dispersion relation is assumed to be independent of $\boldsymbol{x}$ and t and the WKB approximation is used. Under these assumptions, the path of the wave train is completely determined by equation (10a) and the temporal evolution of the wavenumber vector $\boldsymbol{k} = (k, l, m)$ along the path is described by (10b); the refraction process is, in particular, independent of stratification. The derivation of (10) in the spirit of a WKB approximation, and including the next

order terms is sketched in Vanneste (2013), who proceeds to explain the mechanism of spontaneous emission. Broutman et al. (2004) review ray tracing in geophysical fluid dynamics. Olbers et al. (2012) give an extensive treatment of wave propagation in a variable environment.[2]

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

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

$$\boldsymbol{\nabla u} = \begin{pmatrix} u_x & v_x & 0 \\ u_y & v_y & 0 \\ u_z & v_z & 0 \end{pmatrix}. \tag{11}$$

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

$$\boldsymbol{\nabla u} = \begin{pmatrix} -\alpha & 0 & 0 \\ 0 & \alpha & 0 \\ \beta & \gamma & 0 \end{pmatrix} \tag{12}$$

so that the flow is hyperbolic in the horizontal plane and the solution to (10b) reads

$$\boldsymbol{k}_\mathrm{h}(t) = (k_0\, \mathrm{e}^{\alpha t}, l_0\, \mathrm{e}^{-\alpha t})\,. \tag{13}$$

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

---

[2] By writing the dispersion relation in the form $\omega(\boldsymbol{k}, t) = \omega_\mathrm{i}(\boldsymbol{x}, t) + \boldsymbol{u}(\boldsymbol{x}, t) \cdot \boldsymbol{k} = \Omega(\boldsymbol{k}; \boldsymbol{x}, t)$ and assuming that the intrinsic dispersion relation and hence the intrinsic frequency ω_i is independent of $\boldsymbol{x}$, their equation (6.46) reduces to our equation (10b).

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

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

$$\boldsymbol{\nabla u} = \begin{pmatrix} u_x & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & w_z \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -a \end{pmatrix}, \tag{14}$$

so that the solution to (10b) takes the simple form

$$\boldsymbol{k}(t) = (k_0\, \mathrm{e}^{-at}, l_0, m_0\, \mathrm{e}^{at})\,. \tag{15}$$

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

The refraction through axial flows can be significant for two reasons. First, axial flows represent an essential ingredient of a realistic atmospheric and oceanic flow field. Second, the properties of waves predicted by (15) with a negative a so that the magnitude of the component of $\boldsymbol{k}$ along the flow is large (or the corresponding wavelength is small) and $\boldsymbol{k}$ is aligned with the flow (in other words, aligned with the wave crests are perpendicular to the flow) are qualitatively consistent with what is seen in the atmosphere (Fig. 1) and in the ocean (Fig. 9).

4 Interactions between IGWs and density fronts

4.1 Wave capture in frontal strain

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

$$\mathrm{Ri} = \frac{N_0^2}{(\Delta U/\Delta z)^2} = \frac{N_0^2\, f^2}{S_0^4}\,, \tag{16}$$

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

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

$$D_t \boldsymbol{u} + f\, \boldsymbol{e}_z \times \boldsymbol{u} = -\frac{1}{\rho_0} \boldsymbol{\nabla} p + b\, \boldsymbol{e}_z\,, \tag{17a}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0\,, \tag{17b}$$

$$D_t b = 0\,, \tag{17c}$$

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

$$\bar{\boldsymbol{u}}(\boldsymbol{x}, t) = \begin{pmatrix} \alpha\, x + \hat{u}(t)\, z \\ -\alpha\, y + \hat{v}(t)\, z \\ 0 \end{pmatrix}, \tag{18a}$$

$$\bar{b}(\boldsymbol{x}, t) = N^2(t)\, z - S^2(t)\, y\,, \tag{18b}$$

and

$$\bar{p}(\boldsymbol{x}, t) = -\rho_0 \left(\tfrac{1}{2}\, \alpha^2\, (x^2 + y^2) + f\, \alpha\, xy - \tfrac{1}{2}\, N^2(t)\, z^2 + S^2(t)\, yz \right). \tag{18c}$$

This mean flow yields the tensor

$$\boldsymbol{\nabla} \bar{\boldsymbol{u}} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & -\alpha & 0 \\ \hat{u}(t) & \hat{v}(t) & 0 \end{pmatrix}. \tag{19}$$

Inserting the mean flow ansatz (18) with vanishing perturbation quantities back into (17), we obtain a system of ordinary differential equations for the time dependent coefficients, namely

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\hat{u}\\ \hat{v}\\ S^2\\ N^2\end{pmatrix}=\begin{pmatrix}f\,\hat{v}-\alpha\,\hat{u}\\ -f\hat{u}+\alpha\hat{v}+S^2\\ \alpha S^2\\ \hat{v}\,S^2\end{pmatrix} \tag{20}$$

with initial conditions $u(0)=S_0^2/f$, $\hat{v}(0)=2\alpha\,S_0^2/f^2$, $N^2(0)=N_0^2$, and $S^2(0)=S_0^2$, which correspond to an initial vertical and horizontal buoyancy gradient and to corresponding geostrophic velocities in the presence of strain α. System (20) is easily solved for the vertical and horizontal buoyancy gradients,

$$S^2(t)=S_0^2\,\mathrm{e}^{\alpha t}\,, \tag{21a}$$

$$N^2(t)=N_0^2+\frac{S_0^4}{f^2}\,(\mathrm{e}^{2\alpha t}-1)\,, \tag{21b}$$

and the mean flow

$$\bar{\boldsymbol{u}}(\boldsymbol{x},t)=\begin{pmatrix}\alpha\,x+f^{-1}\,S^2(t)\,z\\ -\alpha\,y+2\alpha\,f^{-2}\,S^2(t)\,z\\ 0\end{pmatrix}. \tag{21c}$$

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

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

$$\bar{D}_t\boldsymbol{u}'+\boldsymbol{u}'\cdot\boldsymbol{\nabla}\bar{u}+\boldsymbol{u}'\cdot\boldsymbol{\nabla}\boldsymbol{u}'+f\,\boldsymbol{e}_z\times\boldsymbol{u}'=-\frac{1}{\rho_0}\boldsymbol{\nabla}p'+b'\,\boldsymbol{e}_z\,, \tag{22a}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{u}'=0\,, \tag{22b}$$

$$\bar{D}_t b'+\boldsymbol{u}'\cdot\boldsymbol{\nabla}\bar{b}+\boldsymbol{u}'\cdot\boldsymbol{\nabla}b'=0\,, \tag{22c}$$

where $\bar{D}_t=\partial_t+\bar{v}\,\partial_y$. Inserting the plane wave ansatz $\boldsymbol{u}'(\boldsymbol{x},t)=\mathrm{e}^{i\boldsymbol{k}\cdot\boldsymbol{x}}$ with $\boldsymbol{k}=(0,l,m)$ into (22), we find that the wavenumber vector changes in time according to

$$\boldsymbol{k}(t)=\left(0,\,l_0\,\mathrm{e}^{\alpha t},\,m_0-S_0^2\,f^{-2}\,(\mathrm{e}^{2\alpha t}-1)\right). \tag{23}$$

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

$$\theta=-\tan^{-1}\!\left(\frac{l}{m}\right) \tag{24}$$

changes with time in the presence of frontogenetic strain. According to (23), it asymptotes to

$$\theta_\rho = \tan^{-1}\left(\frac{S^2}{N^2}\right) \tag{25}$$

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

$$e = \tfrac{1}{2}\,|\boldsymbol{u}'|^2 \tag{26}$$

satisfies the evolution equation

$$\bar{D}_t e = -u'^* w'\,\partial_z \bar{u} - \frac{1}{\rho_0}\,\boldsymbol{\nabla}\cdot(\boldsymbol{u}'^* p') + w'^* b' + \alpha\,(|v'|^2 - |u'|^2) - v'^* w'\,\partial_z \bar{v}, \tag{27}$$

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

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

This suggests a route to dissipation of the mesoscale flow different from the scenario associated with wave capture. In wave capture, the wave amplitude grows exponentially in time, leading to wave breaking and dissipation. If frontogenesis is present, multi-scale analysis shows that the wave action is no longer conserved, but decreases in time (Thomas, 2012). In this case, the energetics of the waves are modified by the Reynolds stress introduced by the additional shear production term related to the strain-driven ageostrophic flow. Numerical simulations carried out by Thomas (2012) using the explicit expression for the deformation shear (19) show that the evolution of the kinetic energy density of the IGWs is largely controlled by the last two terms in (27). In particular, in the presence of frontogenetic strain, i.e., when $\alpha > 0$, a perturbation satisfying $|v'| > |u'|$ acts to mix the momentum of the deformation field and extracts its kinetic energy—acting thus as an effective viscosity. In contrast, the ageostrophic shear production term is negative for all θ, resulting in a damping of the IGWs, except when the initial angle is $90^\circ < \theta_0 < 180^\circ$, for which the streamlines are tilted against the ageostrophic shear and the IGWs exhibit a transient growth that is long enough to affect the energetics.

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

$$\frac{\mathrm{Ri}^{1/2} + 1}{\mathrm{Ri}^{1/2} - 1}, \tag{28}$$

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

4.2 Role of IGWs in frontal geostrophic adjustment

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

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

The classic theory of frontogenesis relies on the semi-geostrophic approximation (Hoskins and Bretherton, 1972), which in turn relies on balanced dynamics, thereby removing IGWs altogether. The turbulence emerging in the process of semi-geostrophic frontogenesis was studied numerically by Ragone and Badin (2016) who found that ageostrophic advection is responsible for the growth of filamentary structures with skewed vorticity. Filamentation itself can induce a forward cascade of potential temperature/buoyancy variance and can thus be a possible route to dissipation, even if this might still be inhibited by the fact that semi-geostrophic flows are balanced. In this sense, the energy pathway in frontal regions can be much more complex than the pathway through wave capture alone.

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

$$\partial_t u + u\,\partial_x u - f v = -g\,\partial_x h\,, \tag{29a}$$

$$\partial_t v + u\,\partial_x v + f u = 0\,, \tag{29b}$$

$$\partial_t h + \partial_x (uh) = 0\,. \tag{29c}$$

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

In addition to the potential vorticity $q = (\partial_x v + f)/h$, the system possesses a second Lagrangian invariant, the *geostrophic momentum* $M = v + fx$. Introducing a Lagrangian x-coordinate via $\dot{X} = u(X, t)$, we write the two momentum equations in system (29) as

$$\ddot{X} - fv = -g\,\partial_x h\,, \tag{30a}$$

$$\dot{v} + f\dot{X} = 0\,. \tag{30b}$$

The second equation can be integrated immediately, giving

$$v(a,t) + fX(a,t) = M(a)\,, \tag{31}$$

where we write a to denote a Lagrangian label coordinate. The geostrophic momentum $M(a)$ is determined by the initial conditions: given the initial velocity v_0, one has $M(a) = v_0(a) + fa$. Note that the adjusted state are no longer slowly evolving, but exactly stationary in Lagrangian coordinates. Further, in Lagrangian coordinates the problem reduces to a semi-geostrophic problem in which geostrophy is retained in (30a), but where (30b) is purely ageostrophic. The continuity equation (29c) in Lagrangian coordinates takes the form

$$h(X(a,t),t) = \frac{h_0(a)}{X'(a,t)}\,, \tag{32}$$

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

$$\partial_x h = \frac{h_0'}{(X')^2} - h_0\,\frac{X''}{(X')^3}\,. \tag{33}$$

Writing $\phi(a,t) = X(a,t) - a$ to denote the displacement of a particle from its initial position and inserting (31) and (33) into (30a), we obtain a closed equation for ϕ,

$$\ddot{\phi} + f^2\,\phi + \frac{g\,h_0'}{(1+\phi')^2} + \frac{g\,h_0}{2}\left[\frac{1}{(1+\phi')^2}\right]' = f\,v_0\,, \tag{34}$$

which can be solved with appropriate initial conditions. Assuming positive potential vorticity as a condition for solvability and taking appropriate frontal initial conditions, Zeitlin et al. (2003) solved (34), thereby showing that to leading order in the cross-frontal Rossby number, the fast and slow dynamics split so that there are no fast oscillations trapped in the frontal region. An example of a solution to (34) is shown in Fig. 2. The front is located at $x = 0$, situated in between a double jet configuration set by the initial conditions $h_0(x) = 1 + \exp(-x^2)$, $v_0(x) = -2(x + 0.2\sin(x))\exp(-x^2)$, and $u_0(x) = 0.1\exp(-x^2)$. The horizontal axes show the horizontal distance x and time. Surfaces show the time evolution of the particle displacement ϕ. An initially small departure from balance evolves emitting fast IGWs as well as slowly dispersing near-inertial oscillations. Evolving toward an adjusted state which is a minimizer of the energy, the system sheds energy also via the emergence of shocks. One of these shocks is visible on the left of the figure, where the relative vorticity takes negative values.

This scenario was extended to two-layer stratification by Le Sommer et al. (2014) and to stratified rectilinear fronts by Plougonven and Zeitlin (2005). The presence of stratification allows not only for the process of baroclinic instabilities, but also

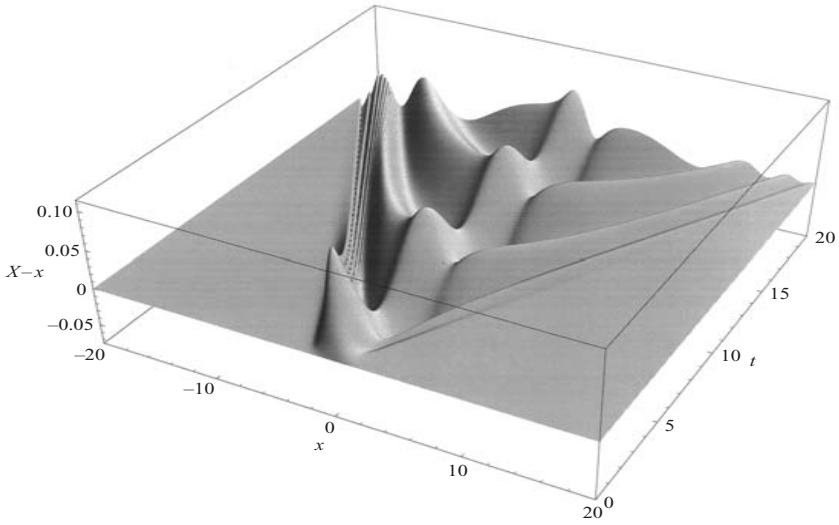

Fig. 2 Solution of (34) with initial conditions $h_0(x) = 1 + \exp(-x^2)$, $v_0(x) = -2(x + 0.2\sin(x))\exp(-x^2)$, $u_0(x) = 0.1\exp(-x^2)$. The horizontal axes show the horizontal distance x and time. Surfaces show the time evolution of the particle displacement $\phi = X - x$. The solution shows emission of fast gravity waves and slowly dispersing near-inertial oscillations. A shock is visible on the left side of the front (negative values of x). From Zeitlin et al. (2003).

for symmetric inertial instabilities. Plougonven and Zeitlin (2005) demonstrate that nonexistence of a smooth adjusted state in vertically bounded domains is generic, with the system evolving toward a singularity in finite time, even in the absence of a background confluent flow. In this setting, there are trapped modes with horizontal scales comparable to the width of the jet. The existence of shocks and the breaking trapped modes can represent a possible pathway toward dissipation.

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

Further complication might be introduced by the presence of a background confluent flow, which might be responsible for time dependent Rossby and Richardson numbers, as discussed in Section 4.1. In turn, this might affect the transition toward a possible adjusted state. A study of an initially smooth time evolving confluent flow would be also interesting for the comparison with the behavior of balance models in the same configuration, such as the surface quasi-geostrophic model, for which proof of the formation of a singularity in a finite time is linked to the formation of singularities in the 3D Euler equation and for which the problem is still open (Constantin et al., 1994).

5 Diagnostics

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

Flows characterized by a small Rossby number evolve on timescales much longer than the timescales of IGWs emitted by the flows. In this situation, the emitted waves can reasonably well be diagnosed using a linear time filter based on a Fourier decomposition. Linear filters are straightforward and easy to apply. Fig. 3 shows the geographical distribution of the super-inertial kinetic energy, obtained by applying such a filter. However, it is not clear how small the Rossby number should be for ensuring adequate wave detection and whether this procedure leads to a consistent separation over the global domain. A review on time filters can be found in von Storch and Zwiers (1999).

An alternative characterization is provided by balance relations which can be derived by systematic asymptotic expansion in various ways; see, e.g., the review by McIntyre (2015). One possibility is the elimination of the vertical velocity associated with balanced motion through the Omega equation to identify the residual signal of the waves (Danioux et al., 2012). There is a variety of more sophisticated balance relations such as the second order correction to quasi-geostrophic theory (Muraki et al., 1999), the semi-geostrophic (Hoskins and Bretherton, 1972; Hoskins, 1975), surface semi-geostrophic (Badin, 2013; Ragone and Badin, 2016), δ-γ balance (Mohebalhojeh and Dritschel, 2001), the L_1-model and variants (Salmon, 1985, 1996; McIntyre and Roulstone, 2002; Oliver and Vasylkevych, 2016), and many others; see, e.g., Barth et al. (1990) for a long list of intermediate models and a comparison of their performance in geostrophic turbulence and Allen et al. (2002) for a numerical comparison in a nonturbulent regime. The general advantage of balance relations is that they offer an *a priori* way of separating motion slaved to PV from other types of motion, at least to some order of approximation in the limit of small Rossby number. The drawback is that they require a careful scale analysis and that they are, by construction, restricted to a specific limited regime of validity. Nonetheless, they offer ways to "diagnose the diagnostics" as the only first-principles analysis available.

A third alternative is the equation-free "optimal potential vorticity balance" scheme pioneered by Viúdez and Dritschel (2004). Their work is based on the notion that geophysical flows are determined to a large extend by potential vorticity, so that the balanced component of the fields are those which are slaved to potential vorticity. Their procedure can be implemented on top of a purely numerical model by slowly increasing the influence of nonlinear interactions from a trivial initial state where the splitting into balanced and imbalanced modes is explicit to the fully nonlinear final state. So long as the energy in the imbalanced degrees of freedom remains adiabatically invariant, the final state will be approximately balanced whenever the initial state is. Optimal balance has some highly desirable features, in particular, carrying no explicit assumptions on scales, only weak dependence on numerical parameters such as the time horizon of integration, and the possibility to define balance having nothing but an actual numerical code. Moreover, in simple settings it

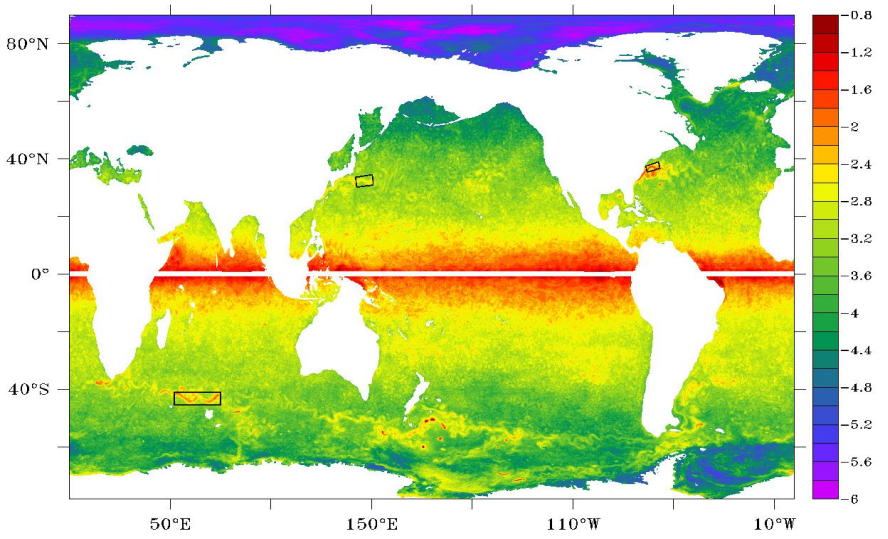

Fig. 3 Square root of super-inertial kinetic energy, $\sqrt{k_{\mathrm{sup}}}$, at 291 m in a 0.1° simulation performed with the MPIOM OGCM. The super-inertial kinetic energy is obtained by Fourier decomposing the hourly zonal and meridional velocity in the frequency domain and calculating the mean variance of the zonal and meridional velocity fluctuations at frequencies larger than the local inertial frequency. The color scale is in units of m/s. Numbers given refer to $\log_{10}\sqrt{k_{\mathrm{sup}}}$.

can be proved that the procedure yields near-optimal results in the asymptotic limit of large scale separation between balanced and imbalanced motion (Cotter, 2013; Gottwald et al., 2017). However, implementation and pre-asymptotic behavior in realistic configurations remains largely unexplored.

No matter whether a linear or a dynamical filter is used, we recommend to first assess the flow field by using e.g. the Rossby number (Section 5.1). Some pitfalls encountered when using linear filters is briefly discussed in Section 5.2. The equation-free "optimal potential vorticity balance" scheme, which we hope to be applicable for a large range of dynamical regimes, is discussed in Sections 5.3 and 5.4.

5.1 Characterization of flow regimes via the Rossby number

Any splitting of motions into balanced and imbalanced components, whether by linear filtering or more sophisticated techniques, relies on at least some degree of scale separation. For geophysical flow, the most important scale separation parameter is the Rossby number, informally described as the ratio of the inertial force to the Coriolis force. Small Rossby number flows are dominated by planetary rotation and can be thought of as well-balanced. When the Rossby number is well below one,

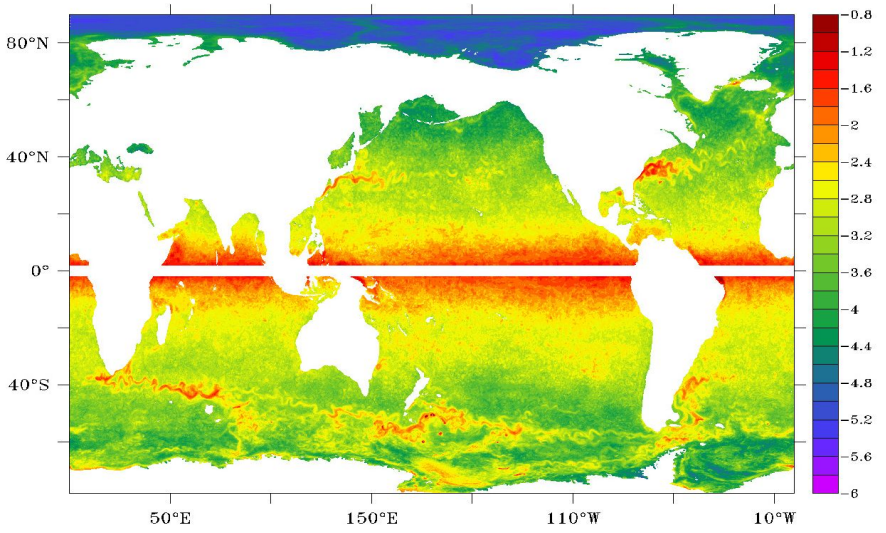

Fig. 4 Square root of super-inertial kinetic energy, $\sqrt{k_{\mathrm{sup}}}$, at 291 m in the same 0.1° simulation considered in Fig. 3. The super-inertial kinetic energy is obtained by filtering out the sub-inertial variance using a running average filter with the filter window identical to the local inertial period. The color scale is in units of m/s. Numbers given refer to $\log_{10}\sqrt{k_{\mathrm{sup}}}$.

the time scales of the mean flow are clearly separated from IGWs. In this case, IGWs can be identified using simple Fourier filters. For larger Rossby numbers, balance relations or dynamical filters are required to separate IGWs from the full flow field. We emphasize, however, that for Rossby numbers close to or larger than unity, the concept of balance breaks down altogether and no method will lead to an unambiguous and physically meaningful splitting; see, e.g., the discussion in McIntyre (2009).

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

$$\mathrm{Ro}_{\mathrm{Lagr}} = \frac{\pi}{\tau f} \tag{35}$$

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

$$R = \frac{|\nabla_{\mathrm{h}}^{\perp} \cdot \boldsymbol{u}_{\mathrm{h}}|}{|f|} \,. \tag{36}$$

Averaging can be used to eliminate possible influence of short-term fluctuations on R. Alternatively, R can be defined using the geostrophic velocity in (36).

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

5.2 Linear filters

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

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

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

$$\bar{x}_t = \sum_{k=-K}^{k=K} a_k \, x_{t+k} \tag{37}$$

with equal weights

$$a_k = \begin{cases} \frac{1}{2K+1} & \text{for } |k| \le K \,, \\ 0 & \text{for } |k| > K \,. \end{cases} \tag{38}$$

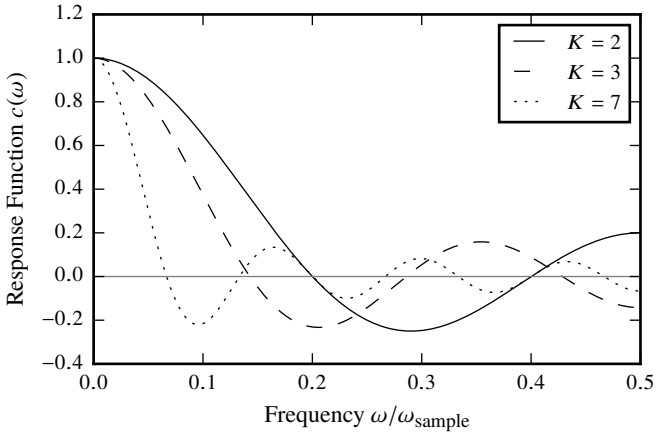

Fig. 5 Response functions $c(\omega)$ of three running average low-pass filters with different filter windows. The frequency axis is labeled in units of the sample frequency ω_{sample} and shows frequencies up to the Nyquist frequency $\omega_{\text{sample}}/2$.

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

$$\bar{S}(\omega) = |c(\omega)|^2 \, S(\omega) \tag{39}$$

with frequency response function

$$c(\omega) = \frac{\sin((2K+1)\,\pi\omega)}{(2K+1)\,\sin(\pi\omega)} \,. \tag{40}$$

Fig. 5 shows $c(\omega)$ for a running average filter with $K = 2$, 3, and 7. The magnitude of $c(\omega)$ is close to one at low frequencies and becomes small and oscillatory at high frequencies, i.e., the filter acts as a low-pass filter. Further, $c(0) = 1$ and $c(\omega) = 0$ at frequencies $\omega = j/(2K+1)$ for $j = 1, \ldots, K$, i.e., the filter leaves the zero frequency component unchanged and completely filters out frequencies that are multiples of $1/(2K+1)$. Thus, one would typically take the local inertial period as the window length, so that variations at frequencies equal to an integer multiple of the local inertial frequency are filtered out. In the residual time series

$$x'_t = x_t - \bar{x}_t \,, \tag{41}$$

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

retained in x'_t. When the sub-inertial variations are strong, this contamination can be significant.

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

The difference between the two filters can be seen by comparing Fig. 3 with Fig. 4. In each case, the filter is tuned to the local inertial period τ_i. Fig. 3 uses a spectral filter with response function

$$c(\omega) = \begin{cases} 0 & \text{for } \omega < 1/\tau_\mathrm{i}\,, \\ 1 & \text{for } \omega \geq 1/\tau_\mathrm{i}\,. \end{cases} \tag{42}$$

In addition, the time series was de-trended and taped using the split-cosine-bell taper (von Storch and Zwiers, 1999) where the non-unit weights make up about 12% of an hourly time series over a period of one month. Fig. 4 is based on the same data, but with a running average filter of filter length τ_i and subsequent computation of the residual (41). The two different filters lead to different results. In particular, the super-inertial kinetic energy in the Southern Ocean is much weaker in Fig. 3 than in Fig. 4. The stronger super-inertial kinetic energy seen with the running average filter is a result of the contamination of the residual by the imperfect filtering of strong low-frequency components. Thus, one should not use a running average filter for the purpose of separating IGWs from eddying flows.

5.3 Optimal potential vorticity balance

In this section, we review the concept of optimal potential vorticity balance as introduced by Viúdez and Dritschel (2004). In the following Section 5.4, we then present a mathematical framework in which optimal balance can be understood systematically. It is our belief that the framework applies in much more general circumstances and can be used for high-accuracy diagnostics and balanced initialization.

The dynamics of large-scale geophysical flow is, to a surprising degree, determined by the evolution of a single scalar field, the potential vorticity q, alone (McIntyre and Norton, 2000). Indeed, in many cases the full equations can be written, equivalently, as a material transport equation for q augmented by additional equations for the evolution of the "ageostrophic components" $\boldsymbol{P}$. For example, for

the rotating Boussinesq model, the components of $\boldsymbol{P}$ are the horizontal divergence δ and the ageostrophic vorticity γ, see Section 2.2 in Franzke et al. (2019). For the rotating shallow water equations, Section 2.1 shows how to split off a second-order in time equation using $\partial_t h$ as the imbalance variable.

In all of these cases, leading order balance is obtained by simply setting $\boldsymbol{P} = 0$, but an optimally balanced flow will have small, but nonzero values for these "ageostrophic" variables. Balance is then viewed as a mapping $q \mapsto \boldsymbol{P} = \boldsymbol{\Phi}(q)$. The function $\boldsymbol{\Phi}$ can be seen as defining a manifold in phase space parameterized by potential vorticity. It is generally understood that a unique invariant slow manifold does not exist except in trivial cases, but that it is possible to construct balance manifolds which are almost invariant over very long intervals of time (Vanneste, 2013). The numerical task can now be described as follows. Given any potential vorticity field q^*, numerically compute one image $\boldsymbol{P}^*$ of the balance map in an optimal way, where possible optimality considerations are discussed in Section 5.4 below.

We further require that for a state where $\boldsymbol{P} = 0$ there exists a stationary "rest state potential vorticity" q_r. On the sphere or on the beta-plane, q_r will only depend on latitude. With $\boldsymbol{\eta}$ denoting the flow map, let us write $\boldsymbol{x} = \boldsymbol{\eta}(\boldsymbol{a}, t)$ to denote the Eulerian position of the fluid particle initially at Lagrangian label coordinate $\boldsymbol{a}$. Material conservation of potential vorticity is then expressed as

$$q(\boldsymbol{\eta}(\boldsymbol{a}, t), t) = q_0(\boldsymbol{a})\,. \tag{43}$$

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

$$q_0(\boldsymbol{a}, t) = q_\mathrm{r}(\boldsymbol{a}) + \rho(t/T)\, q_\mathrm{a}(\boldsymbol{a}) \tag{44}$$

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

This leaves us with the following task: solve the modified equations of motion with the temporal boundary conditions $\boldsymbol{P}(0) = 0$ and $q(T) = q^*$, then define $p^* = \boldsymbol{P}(T)$. Figure 6 shows a sketch of the situation. For every fixed t, the balance manifold is illustrated as a one-dimensional curve; the manifold itself is time dependent. At the start of the ramp when $t = 0$, the balance manifold is exact and trivial; we know how to explicitly balance the system. At the final time $t = T$, the balance manifold is only approximate and nontrivial. A trajectory from $t = 0$ to $t = T$ which reaches the imposed potential vorticity q^* at the end of the ramp, shown in red in Figure 6, remains approximately balanced so long as the change of the balance manifold in time

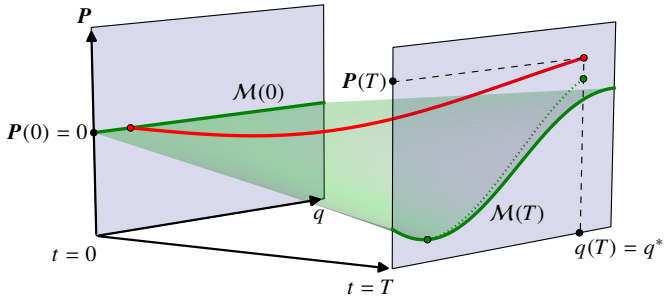

Fig. 6 Sketch of the ramping procedure in extended phase space.

is adiabatic. The quality of balance for the final state will be affected by the choice of ramp function, the choice of the ramp time T, and the scale separation between balanced and unbalanced motion. We will comment on the conjectured interplay of these parameters in the context of a simple model for balance in Section 5.4 below.

To use optimal potential vorticity balance in practice, two further problems need to be overcome. First, what is ramped in (44) is the initial, i.e. Lagrangian, potential vorticity q_0. The dependence of the Eulerian potential vorticity on q_0, however, is only straightforward in a semi-Lagrangian setting such as the CASL method by Dritschel and co-workers (Dritschel et al., 1999) or one of its variants. Second, the formulation of the problem as stated here is a boundary value problem in time. As such, a direct implementation requires specialized code. Viúdez and Dritschel (2004), however, have proposed an iterative scheme that can be performed with only an evolutionary code: Start at $t = T$ with $q(T) = q^*$ and $\boldsymbol{P}(T) = 0$ and integrate *backward* to $t = 0$. At $t = 0$, the solution will be generally imbalanced, so force the solution back to balance by resetting $\boldsymbol{P}(0) = 0$, keeping $q(0)$ unchanged. Now integrate *forward* to $t = T$. At $t = T$, the solution will still be approximately balanced, but in general $q(T) \neq q^*$. Now reset $q(T) = q^*$, keeping $\boldsymbol{P}(T)$ unchanged. Repeat until the iteration converges. In that case, the fixed point of the iteration is, by construction, a solution to the temporal boundary value problem as outlined earlier. Viúdez and Dritschel (2004) report fast convergence in their numerical experiments.

5.4 A simple model for optimal balance

A simple finite dimensional model for balance is the Hamilton system

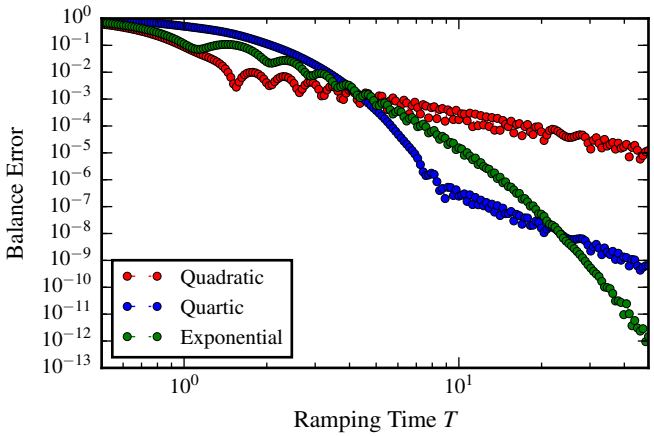

Fig. 7 Numerical test for a linear model problem. Shown is the balancing error as a function of the ramp time for ramp functions with different orders of vanishing at the end points of the ramp.

$$\dot{q} = p\,, \tag{45a}$$
$$\varepsilon\,\dot{p} = Jp - \nabla V(q)\,, \tag{45b}$$

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

$$J = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} \tag{46}$$

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

$$\dot{q} = p\,, \tag{47a}$$
$$\varepsilon\,\dot{p} = Jp - \rho(t/T)\,\nabla V(q)\,. \tag{47b}$$

We solve (47) as a boundary value problem where we impose that at $t = 0$, the system is on the trivial slow manifold and at $t = T$ the position variables take some prescribed value q^*. Thus, (47) is augmented with boundary conditions

$$p(0) = 0 \qquad \text{and} \qquad q(T) = q^*\,. \tag{48}$$

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

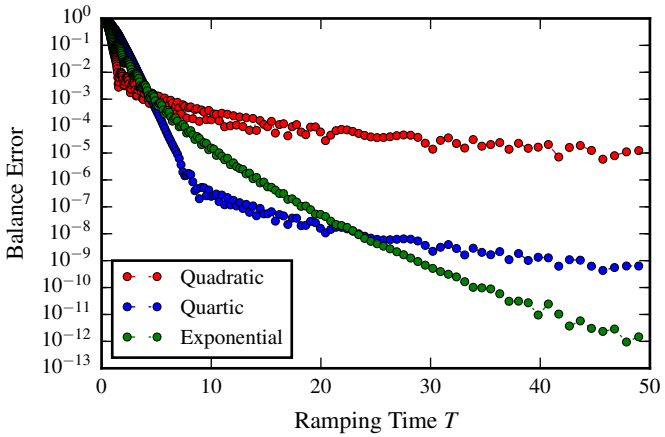

Fig. 8 The same data as shown in Figure 7 on a semi-log scale. The exponential regime of the error behavior can be seen as a straight line.

However, there is a second source of error besides the drift. The asymptotic series defining the family of manifolds $\mathcal{M}(t)$ for the ramped system differs from the asymptotic series for the trivial balance manifold at $t = 0$ and, similarly, it differs from the asymptotic series for the balance relation of the original system (45) by terms containing derivatives of ρ. More precisely, $O(\varepsilon^{n+1})$-accuracy of the procedure requires the first n derivatives of ρ to vanish at the temporal end points. This excludes uniformly analytic ramp functions. Yet, it is still possible to get exponential bounds, albeit with a smaller power of ε in the exponent; see Gottwald et al. (2017).

The behavior of the optimal balance procedure can be demonstrated most easily in the linear case. In this case, for a frozen value of the ramping parameter, the slow manifold is an exact invariant and can be computed by diagonalization of the full system matrix. Figure 7 shows the quality of balance achieved by the method for the simple potential $V(q) = \frac{1}{2}|q|^2$ and different ramp functions of the type

$$\rho(\theta) = \frac{f(\theta)}{f(\theta) + f(1-\theta)} \tag{49}$$

for $f(\theta) = \theta^2$ ("quadratic"), $f(\theta) = \theta^4$ ("quartic"), and $f(\theta) = \exp(-1/\theta)$ ("exponential"). The balance error is defined as the distance of $p(T)$ as computed via solving the boundary value problem of the ramped system to the slow eigenspace of the full system (45) computed by exact diagonalization.

The behavior for large ramp times (corresponding to good adiabatic invariance) is dominated by the error committed at the start and end point of the ramp. It shows the expected power law scaling in T, given by the order of vanishing of the ramp function at the end points. The behavior for small ramp times is dominated by the deviation from adiabatic invariance and shows exponential behavior. This can be

clearly seen from Figure 8 where the same data is plotted semi-logarithmically. The cut-over point moves to larger values of T as the order of vanishing of the ramp at the end points is increased.

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

6 High-resolution ocean general circulation models as a novel tool for studying spontaneous emission

So far, modeling studies on spontaneous imbalance have been concentrated on flows with idealized or simplified configurations. Even though great progress has been made in understanding spontaneous imbalance as a phenomenon of geophysical fluids (see discussion and references Section 2) and attempts have been made to make the considered configurations to resemble the real atmospheric or oceanic conditions, modeling studies using idealized configurations are always challenged by the question of whether and to what extent spontaneous imbalance in the considered flow configuration matters in reality. These questions can be answered by considering observations with high spatial and temporal resolution. High resolution is essential for detecting spontaneous imbalance in the real ocean. As high-resolution observations are often not available, it is natural to use global general circulation models (GCMs) capable of realistically simulating the global ocean.

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

An essential feature of the eddying ocean simulated by the 0.1° MPIOM is horizontally axial flows. These flows follow a bundle of isobars in a horizontal

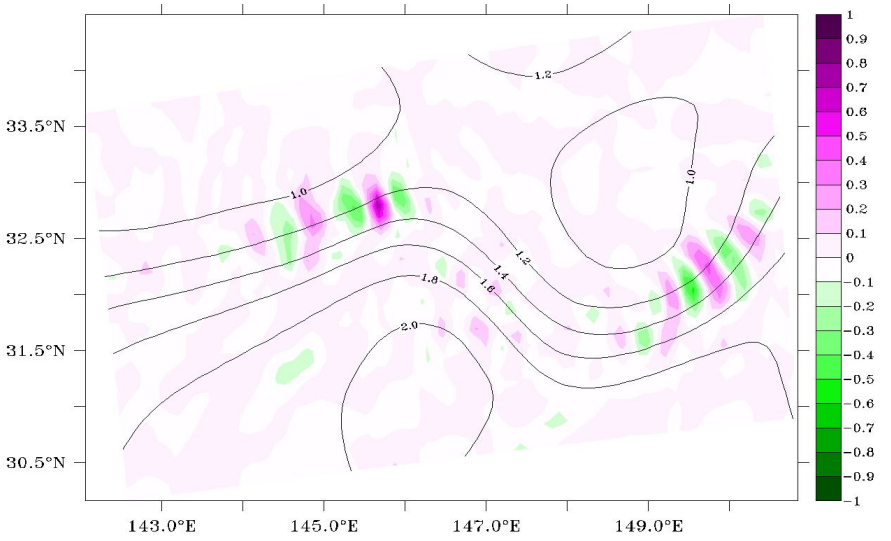

Fig. 9 Snapshots of high-pass filtered vertical velocities (color shadings) and dynamically relevant pressure (contours) identified from a 10° simulation performed with MPIOM. Units are 10^{-4} m/s for vertical velocity and 10^{-1} bar for pressure.

plane, such as those indicated by the black lines in Fig. 9. The isobars are curved, though not closed, and run parallel to each other. Flow following isobars is consistent with the idea that the eddying flows are near geostrophic balance. The shadings in Fig. 9 show a snapshot of vertical velocity containing only super-inertial fluctuations, indicating gravity wave trains. The wave trains are concentrated along axial flows with the wave crests being perpendicular to the flow. Their vertical structures are comparable to those of low modes internal waves. They propagate along the flow as if they were captured.

These wave trains can result from the refraction of waves by the axial flow. Following equation 10, the effect of the refraction depends on the flow gradient tensor $\boldsymbol{\nabla u}$ along the path of the wave train. For a non-axial flow, $\boldsymbol{\nabla u}$ can be quite complicated so that $\boldsymbol{k}_\mathrm{h}$ is generally not in the direction of $\boldsymbol{u}_\mathrm{h}$. For gravity waves, $\boldsymbol{k}_\mathrm{h}$ is in the direction of the horizontal component of the intrinsic group velocity. For these waves, the wave trains tend to move away from the given flow, making refraction by a particular gradient tensor $\boldsymbol{\nabla u}$ difficult. An axial flow has the property that the shear of the flow vanishes along the core. This makes $\boldsymbol{\nabla u}$ to take an extremely simple form for non-divergent and horizontally axial flows. With such a $\boldsymbol{\nabla u}$, waves whose wavenumber vector does not have a transverse-flow component initially will stay in the core and eventually becomes refracted by the negative stretching rate such that their along-flow wavenumber increases exponentially.

Fig. 9 provides a first indication that waves can be emitted and refracted by eddying flows in a realistic setting. Many questions are still unanswered. For instance, the 0.1° MPIOM-simulation discussed produces super-inertial kinetic energy not only

in mid-latitude oceans but also, with much greater strength, in the sub-tropical and tropical ocean (see Fig. 3). It is not clear whether these super-inertial fluctuations represent waves with properties similar to those found in the mid-latitude oceans shown in Fig. 9, what is the relation of these waves to the background flows that are characterized by increasingly larger Rossby number towards the equator, and what is the role of these super-inertial fluctuations for large-scale subtropical and tropical flows.

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

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

7 Discussion

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

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

The second set of questions concerns the use of optimal balance for diagnostic purposes. While the analysis of optimal balance is well understood in the context

of the finite dimensional model presented in Section 5.4, no rigorous results are yet available to cover even the RSW equations or any of the more complex models. To increase the applicability of optimal balance to a wide range of models, it is important to implement the ramp on top of a standard Eulerian formulation of the equations, e.g., by multiplying all nonlinear terms with the ramp function. To our knowledge, this has not been tested, but seems promising and within the framework outlined for the finite dimensional model problem in Section 5.4. Once this is achieved, we believe that the method has considerable potential for applications to the real ocean as a detailed understanding of the precise scalings is not necessary to apply the method so long as it is possible to adiabatically ramp the system into a simpler model for which balance can be unambiguously defined.

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

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

Apart from diagnostics, theoretical considerations, such as those based on ray equations, are indispensable. They are needed to interpret the filtered output and to ensure that the output has proper physical meaning and does not result from numerical artifacts. In general, a number of questions are still open and two of these questions are particularly important and should be tackled first. One concerns the spontaneous imbalance in the subtropical and tropical ocean, which seems to produce much stronger super-inertial kinetic energy than that in mid-latitude oceans according to Fig. 3. The other question concerns the difference in waves spontaneously emitted in different flow regimes, for instance those in quasi-balanced axial flows, compared with those in a frontal system. Even though one difference, namely the difference in wave intensity, has been identified by theoretical considerations (Vanneste, 2008, 2013), the long time behavior of waves emitted in different flow regimes and the resulting implications for the flows in different regimes has not been studied systematically. Both questions are crucial for understanding the role of spontaneous imbalance for the real ocean.

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

Acknowledgements We thank Tuba Masur for careful proof-reading of the manuscript. This paper contributes to the project "The interior energy pathway: internal wave emission by quasi-balanced flows" of the Collaborative Research Center TRR 181 "Energy Transfers in Atmosphere and

Ocean" funded by the German Research Foundation. Funding through the TRR 181 is gratefully acknowledged. GB was partially supported by DFG grants FOR1740, BA-5068/8-1, and BA-5068/9-1. MO was partially supported by DFG grant OL-155/6-1.

References

Allen, J. S., Holm, D. D., and Newberger, P. A. (2002). Toward an extended-geostrophic Euler–Poincaré model for mesoscale oceanographic flow. In Norbury, J. and Roulstone, I., editors, *Large-scale atmosphere–ocean dynamics*, volume 1, pages 101–125. Cambridge University Press.

Badin, G. (2013). Surface semi-geostrophic dynamics in the ocean. *Geophys. Astrophys. Fluid Dyn.*, 107(5):526–540.

Barth, J., Allen, J., and Newberger, P. (1990). On intermediate models for barotropic continental shelf and slope flow fields. Part II: Comparison of numerical model solutions in doubly periodic domains. *J. Phys. Oceanogr.*, 20(7):1044–1076.

Blumen, W. (1972). Geostrophic adjustment. *Rev. Geophys.*, 10(2):485–528.

Bouchut, F., Sommer, J. L., and Zeitlin, V. (2004). Frontal geostrophic adjustment and nonlinear wave phenomena in one-dimensional rotating shallow water. Part 2. High-resolution numerical simulations. *J. Fluid Mech.*, 514:35–63.

Broutman, D., Rottman, J. W., and Eckermann, S. D. (2004). Ray methods for internal waves in the atmosphere and ocean. *Annu. Rev. Fluid Mech.*, 36:233–253.

Bühler, O. and McIntyre, M. E. (2005). Wave capture and wave-vortex duality. *J. Fluid Mech.*, 534:67–95.

Charney, J. G. (1971). Geostrophic turbulence. *J. Atmos. Sci.*, 28(6):1087–1095.

Cho, H. R., Shepherd, T. G., and Vladimirov, V. A. (1993). Application of the direct liapunov method to the problem of symmetric stability in the atmosphere. *J. Atmos. Sci.*, 50(6):822–836.

Constantin, P., Majda, A., and Tabak, E. (1994). Formation of strong fronts in the 2-D quasigeostrophic thermal active scalar. *Nonlinearity*, 7:1495–1533.

Cotter, C. (2013). Data assimilation on the exponentially accurate slow manifold. *Phil. Trans. R. Soc. A*, 371(1991):20120300.

Cotter, C. J. and Reich, S. (2006). Semigeostrophic particle motion and exponentially accurate normal forms. *Multiscale Model. Sim.*, 5(2):476–496.

Danioux, E., Vanneste, J., Klein, P., and Sasaki, H. (2012). Spontaneous inertia-gravity-wave generation by surface-intensified turbulence. *J. Fluid Mech.*, 699:153–173.

Dritschel, D. G., Polvani, L. M., and Mohebalhojeh, A. R. (1999). The contour-advective semi-lagrangian algorithm for the shallow water equations. *Mon. Weather Rev.*, 127(7):1551–1565.

Ford, R. (1994). Gravity wave radiation from vortex trains in rotating shallow water. *J. Fluid Mech.*, 281:81–118.

Ford, R., McIntyre, M. E., and Norton, W. A. (2000). Balance and the slow quasi-manifold: Some explicit results. *J. Atmos. Sci.*, 57(9):1236–1254.

Ford, R., McIntyre, M. E., and Norton, W. A. (2002). Reply. *J. Atmos. Sci.*, 59(19):2878–2882.

Franzke, C. L. E., Oliver, M., Rademacher, J. D. M., and Badin, G. (2019). Multiscale methods for geophysical flows. This Volume.

Gall, R. L., Williams, R. T., and Clark, T. L. (1988). Gravity waves generated during frontogenesis. *J. Atmos. Sci.*, 45(15):2204–2219.

Gottwald, G. A., Mohamad, H., and Oliver, M. (2017). Optimal balance via adiabatic invariance of approximate slow manifolds. *Multiscale Model. Simul.*, 15(4):1404–1422.

Griffies, S. M., Winton, M., Anderson, W. G., Benson, R., Delworth, T. L., Dufour, C. O., Dunne, J. P., Goddard, P., Morrison, A. K., Rosati, A., Wittenberg, A. T., Yin, J., and Zhang, R. (2015). Impacts on ocean heat from transient mesoscale eddies in a hierarchy of climate models. *J. Climate*, 28(3):952–977.

Hoskins, B. J. (1975). The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, 32(2):233–242.

Hoskins, B. J. and Bretherton, F. P. (1972). Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.*, 29(1):11–37.

Jones, W. L. (1969). Ray tracing for internal gravity waves. *J. Geophys. Res.*, 74(8):2028–2033.

Klein, P. and Treguier, A. M. (1995). Dispersion of wind-induced inertial waves by a barotropic jet. *J. Mar. Res.*, 53(1):1–22.

Kunze, E. (1985). Near-inertial wave propagation in geostrophic shear. *J. Phys. Oceanogr.*, 15(5):544–565.

Kunze, E. and Boss, E. (1998). A model for vortex-trapped internal waves. *J. Phys. Oceanogr.*, 28(10):2104–2115.

Le Sommer, J., Medvedev, S., Plougonven, R., and Zeitlin, V. (2014). Singularity formation during relaxation of jets and fronts toward the state of geostrophic equilibrium. *J. Fluid Mech.*, 758:287–311.

Le Sommer, J., Reznik, G. M., and Zeitlin, V. (2004). Nonlinear geostrophic adjustment of long-wave disturbances in the shallow-water model on the equatorial beta-plane. *J. Fluid Mech.*, 515:135–170.

Lee, M.-M., Nurser, A. J. G., Coward, A. C., and de Cuevas, B. A. (2007). Eddy advective and diffusive transports of heat and salt in the southern ocean. *J. Phys. Oceanogr.*, 37(5):1376–1393.

Li, H. and von Storch, J.-S. (2013). On the fluctuating buoyancy fluxes simulated in a OGCM. *J. Phys. Oceanogr.*, 43(7):1270–1287.

Li, Z., von Storch, J.-S., and Müller, M. (2015). The M_2 internal tide simulated by a $1/10^\circ$ OGCM. *J. Phys. Oceanogr.*, 45(12):3119–3135.

Lighthill, M. J. (1952). On sound generated aerodynamically. I. General theory. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 211(1107):564–587.

Llewellyn Smith, S. G. (1999). Near-inertial oscillations of a barotropic vortex: Trapped modes and time evolution. *J. Phys. Oceanogr.*, 29(4):747–761.

Lott, F., Plougonven, R., and Vanneste, J. (2010). Gravity waves generated by sheared potential vorticity anomalies. *J. Atmos. Sci.*, 67(1):157–170.

Lott, F., Plougonven, R., and Vanneste, J. (2012). Gravity waves generated by sheared three-dimensional potential vorticity anomalies. *J. Atmos. Sci.*, 69(7):2134–2151.

Maltrud, M., Bryan, F., and Peacock, S. (2010). Boundary impulse response functions in a century-long eddying global ocean simulation. *Environ. Fluid Mech.*, 10(1–2):275–295.

Maltrud, M. E. and McClean, J. L. (2005). An eddy resolving global 1/10 ° ocean simulation. *Ocean Model.*, 8(1–2):31–54.

Masumoto, Y., Sasaki, H., Kagimoto, T., Komori, N., Ishida, A., Sasai, Y., Miyama, T., Motoi, T., Mitsudera, H., Takahashi, K., Sakuma, H., and Yamagata, T. (2004). A fifty-year eddy-resolving simulation of the world ocean – preliminary outcomes of OFES (OGCM for the Earth Simulator). *J. Earth Simulator*, 1:35–56.

McIntyre, M. (2015). Dynamical meteorology – balanced flow. In Pyle, J. and Zhang, F., editors, *Encyclopedia of Atmospheric Sciences*, pages 298–303. Academic Press, Oxford, second edition.

McIntyre, M. E. (2009). Spontaneous imbalance and hybrid vortex-gravity structures. *J. Atmos. Sci.*, 66:1315–1326.

McIntyre, M. E. and Norton, W. A. (2000). Potential vorticity inversion on a hemisphere. *J. Atmos. Sci.*, 57(9):1214–1235.

McIntyre, M. E. and Roulstone, I. (2002). Are there higher-accuracy analogues of semi-geostrophic theory? In Norbury, J. and Roulstone, I., editors, *Large-scale atmosphere–ocean dynamics*, volume 2, pages 301–364. Cambridge University Press.

Medvedev, S. B. and Zeitlin, V. (2005). Weak turbulence of short equatorial waves. *Phys. Lett. A*, 342(3):217–227.

Mohebalhojeh, A. R. and Dritschel, D. G. (2001). Hierarchies of balance conditions for the f-plane shallow-water equations. *J. Atmos. Sci.*, 58(16):2411–2426.

Müller, P., McWilliams, J. C., and Molemaker, M. J. (2005). Routes to dissipation in the ocean: The 2D/3D turbulence conundrum. In Baumert, H. Z., Simpson, J., and Sündermann, J., editors, *Marine Turbulence*, pages 397–405. Cambridge University Press.

Muraki, D. J., Snyder, C., and Rotunno, R. (1999). The next-order corrections to quasigeostrophic theory. *J. Atmos. Sci.*, 56(11):1547–1560.

Nadiga, B. T. (2014). Nonlinear evolution of a baroclinic wave and imbalanced dissipation. *J. Fluid Mech.*, 756:965–1006.

Nagai, T., Tandon, A., Kunze, E., and Mahadevan, A. (2015). Spontaneous generation of near-inertial waves by the kuroshio front. *J. Phys. Oceanogr.*, 45(9):2381–2406.

Nikurashin, M. and Ferrari, R. (2011). Global energy conversion rate from geostrophic flows into internal lee waves in the deep ocean. *Geophys. Res. Lett.*, 38(8):L08610.

Nikurashin, M., Vallis, G. K., and Adcroft, A. (2013). Routes to energy dissipation for geostrophic flows in the southern ocean. *Nat. Geosci.*, 6(1):48–51.

Obukhov, A. (1949). On the problem of geostrophic wind. *Izv. AN SSSR, Geograph. and Geophys. Ser*, 13:281–286.

Olbers, D., Willebrand, J., and Eden, C. (2012). *Ocean Dynamics*. Springer.

Oliver, M. and Vasylkevych, S. (2016). Generalized large-scale semigeostrophic approximations for the f-plane primitive equations. *J. Phys. A: Math. Theor.*, 49:184001.

O'Sullivan, D. and Dunkerton, T. J. (1995). Generation of Inertia-Gravity Waves in a Simulated Life Cycle of Baroclinic Instability. *J. Atmos. Sci.*, 52:3695–3716.

Plougonven, R. and Snyder, C. (2005). Gravity waves excited by jets: Propagation versus generation. *Geophys. Res. Lett.*, 32(18):L18802.

Plougonven, R. and Snyder, C. (2007). Inertia–gravity waves spontaneously generated by jets and fronts. Part I: Different baroclinic life cycles. *J. Atmos. Sci.*, 64(7):2502–2520.

Plougonven, R. and Zeitlin, V. (2005). Lagrangian approach to geostrophic adjustment of frontal anomalies in a stratified fluid. *Geophys. Astrophys. Fluid Dyn.*, 99(2):101–135.

Plougonven, R. and Zhang, F. (2014). Internal gravity waves from atmospheric jets and fronts. *Rev. Geophys.*, 52(1):33–76.

Ragone, F. and Badin, G. (2016). A study of surface semi-geostrophic turbulence: freely decaying dynamics. *J. Fluid Mech.*, 792:740–774.

Reznik, G. M., Zeitlin, V., and Jelloul, M. B. (2001). Nonlinear theory of geostrophic adjustment. Part 1. Rotating shallow-water model. *J. Fluid Mech.*, 445:93–120.

Rossby, C.-G. (1938). On the mutual adjustment of pressure and velocity distributions in certain simple current systems, ii. *J. Mar. Res.*, 1(3):239–263.

Salmon, R. (1985). New equations for nearly geostrophic flow. *J. Fluid Mech.*, 153:461–477.

Salmon, R. (1996). Large-scale semigeostrophic equations for use in ocean circulation models. *J. Fluid Mech.*, 318:85–105.

Saujani, S. and Shepherd, T. G. (2002). Comments on "balance and the slow quasimanifold: some explicit results". *J. Atmos. Sci.*, 59(19):2874–2877.

Snyder, C., Muraki, D. J., Plougonven, R., and Zhang, F. (2007). Inertia–gravity waves generated within a dipole vortex. *J. Atmos. Sci.*, 64(12):4417–4431.

Snyder, C., Plougonven, R., and Muraki, D. J. (2009). Mechanisms for spontaneous gravity wave generation within a dipole vortex. *J. Atmos. Sci.*, 66(11):3464–3478.

Thomas, L. N. (2012). On the effects of frontogenetic strain on symmetric instability and inertia–gravity waves. *J. Fluid Mech.*, 711:620–640.

van Haren, H. (2003). On the polarization of oscillatory currents in the Bay of Biscay. *J. Geophys. Res. Oceans*, 108(C9):3290.

Vanneste, J. (2008). Exponential smallness of inertia–gravity wave generation at small Rossby number. *J. Atmos. Sci.*, 65(5):1622–1637.

Vanneste, J. (2013). Balance and spontaneous wave generation in geophysical flows. *Ann. Rev. Fluid Mech.*, 45(1):147–172.

Vanneste, J. and Yavneh, I. (2004). Exponentially small inertia–gravity waves and the breakdown of quasigeostrophic balance. *J. Atmos. Sci.*, 61(2):211–223.

Vanneste, J. and Yavneh, I. (2007). Unbalanced instabilities of rapidly rotating stratified shear flows. *J. Fluid Mech.*, 584:373–396.

Viúdez, A. (2006). Spiral patterns of inertia–gravity waves in geophysical flows. *J. Fluid Mech.*, 562:73–82.

Viúdez, Á. (2007). The origin of the stationary frontal wave packet spontaneously generated in rotating stratified vortex dipoles. *J. Fluid Mech.*, 593:359–383.

Viúdez, A. and Dritschel, D. G. (2004). Optimal potential vorticity balance of geophysical flows. *J. Fluid Mech.*, 521:343–352.

Viúdez, A. and Dritschel, D. G. (2006). Spontaneous generation of inertia–gravity wave packets by balanced geophysical flows. *J. Fluid Mech.*, 553:107–117.

von Storch, H. and Zwiers, F. W. (1999). Statistical analysis in climate research. *Cambridge University Press*, page 484.

von Storch, J.-S., Eden, C., Fast, I., Haak, H., Hernández-Deckers, D., Maier-Reimer, E., Marotzke, J., and Stammer, D. (2012). An estimate of the Lorenz energy cycle for the world ocean based on the STORM/NCEP simulation. *J. Phys. Oceanogr.*, 42(12):2185–2205.

Wang, S. and Zhang, F. (2010). Source of gravity waves within a vortex-dipole jet revealed by a linear model. *J. Atmos. Sci.*, 67(5):1438–1455.

Wang, S., Zhang, F., and Snyder, C. (2009). Generation and propagation of inertia–gravity waves from vortex dipoles and jets. *J. Atmos. Sci.*, 66(5):1294–1314.

Xing, J. and Davies, A. M. (2004). On the influence of a surface coastal front on near-inertial wind-induced internal wave generation. *J. Geophys. Res. Oceans*, 109(C1):C01023.

Zeitlin, V. (2010). Lagrangian dynamics of fronts, vortices and waves: Understanding the (semi-) geostrophic adjustment. In *Fronts, Waves and Vortices in Geophysical Flows*, pages 109–137. Springer.

Zeitlin, V., Medvedev, S. B., and Plougonven, R. (2003). Frontal geostrophic adjustment, slow manifold and nonlinear wave phenomena in one-dimensional rotating shallow water. Part 1. Theory. *J. Fluid Mech.*, 481:269–290.

Zhang, F. (2004). Generation of mesoscale gravity waves in upper-tropospheric jet–front systems. *J. Atmos. Sci.*, 61(4):440–457.